A proof-theoretic approach to formal epistemology

Sara Negri

Joint work with Edi Pavlović (University of Helsinki)

Proof Theory Virtual Seminar
3 February 2021

A problem in formal epistemology

Can knowledge be defined?

Several attempts at defining knowledge, starting already with Plato's *Meno* ($\sim$ 400 BC)

> Socrates*: Now this is an illustration of the nature of* **true opinions***: while they stay with us they are beautiful and fruitful, but they run away out of the human soul, and do not remain long, and therefore they are not of much value until they are* **fastened by the tie of the cause***; [...] But* **when they are bound, in the first place, they have the nature of knowledge***; and, in the second place, they are permanent.* **And this is why knowledge is more honourable and excellent than true opinion, because fastened by a chain.**

... and *Theaetetus* ($\sim$ 369 BC)

> Socrates*: When, therefore, any one forms the* **true opinion** *of anything without* **rational explanation**, *you may say that his mind is truly exercised, but has no* **knowledge**; *for he who cannot give and receive a reason for a thing, has no knowledge of that thing; but when he adds rational explanation, then, he is perfected in knowledge.*

Knowledge = justified true belief

Gettier (1963) counterexamples challenge the equation.
Counterexamples are of the form:
$JBel_a(P)$, P implies Q, therefore $JBel_a(Q)$.
However, P happens to be false, but Q is still true. Does *a know Q*?
I am justified in believing something that is only by chance true. Is this knowledge?

In Gettier counterexamples the relation of justification between premisses and conclusions is hampered by additional information. this takes to the
defeasibility theory of knowledge: new evidence overrides, or defeats, the subject's prima facie justification for belief

Would we still believe in A if we were given some additional evidence?

New characterization: **Knowledge as indefeasibly justified true belief**

To talk about indefeasible belief we must first explain what it means to revise a belief on the face of new information

Belief revision

One of the most popular logical approaches to belief revision is the theory AGM (Alchourròn, Gärdenfors, and Makinson 1982)

Operation of belief revision takes an epistemic state K and a proposition A and gives a new, revised, epistemic state $K * A$; this defines a non-monotonic consequence relation

$$A \mid\sim B \equiv B \in K * A$$

- $\mid\sim$ is a nonmonotonic consequence relation
- $\mid\sim$ depends of a background epistemic state, so family of consequence relations
- beliefs concern only the "facts" of the world and not higher-order beliefs (i.e., beliefs about beliefs).
- conditional at the level of the consequence relation, not at the level of the object language so cannot treat nested conditionals
- cannot treat iterated belief revision

Something more general is needed

The main goal

Fully formal epistemology.

Desiderata:

- Epistemic and doxastic modalities
- Conditionals in the object language, allowing for nesting
- Notion of justification
- Treatment of belief revision
- Valid inference, absolute and multi-agent
- Semantic and syntactic approach, completeness theorems
- Computation

In this talk we address most of these.

Assumption: familiarity with sequent calculi.

Plan of the talk

Dynamic epistemic logic

Neighbourhood semantics

Conditional doxastic logic CDL

Knowledge and simple belief

Sequent calculus G3SBK

Properties of knowledge

Proof theory and paradox control

Dynamic epistemic logic

https://plato.stanford.edu/entries/dynamic-epistemic/

"Dynamic Epistemic Logic is the study of a family of modal logics, each of which is obtained from a given logical language by adding one or more modal operators that describe model-transforming actions."

Semantic tools (the preferred approach in the literature) are a generalization of possible world semantics:

- Plausibility models;
- Grove models, Lewis' sphere semantics;

We focus on a particular dynamic epistemic logic

Neighbourhood Models for *CDL*

Neighbourhood models

- These models associate to each world a set of sets of worlds, used to interpret modalities; they were originally proposed to give an interpretation of non-normal modal logics: Montague (1968), Scott (1970), Chellas (1980)...

- Semantics of counterfactuals: Sphere models (Lewis1973);

- Semantics of belief revision: Grove (1988);

- Studied recently also by Marti (2013); Negri (2016); recent monograph: Pacuit (2017).

Neighbourhood semantics

Definition 2.1 (Neighbourhood frame)

A neighbourhood frame has the form $\langle W, I \rangle$ where W is a nonempty set and $I : W \to \mathcal{P}(\mathcal{P}(W))$ is a neighbourhood function.

We form the neighbourhood model $\mathcal{M} = \langle W, I, [\![\ ]\!] \rangle$ by adding the propositional evaluation function $[\![\ ]\!]$:

Definition 2.2 (Evaluation function $[\![\ ]\!]$)

$[\![\ ]\!] : Atm \to \mathcal{P}(W)$ is the evaluation for atomic formulas. Truth conditions for formulas extend $[\![\ ]\!]$ inductively as:
$[\![\neg A]\!] = W - [\![A]\!]$
$[\![A \,\&\, B]\!] = [\![A]\!] \cap [\![B]\!]$
$[\![A \vee B]\!] = [\![A]\!] \cup [\![B]\!]$
$[\![A \supset B]\!] = (W - [\![A]\!]) \cup [\![B]\!]$

A formulas A is valid in $\mathcal{M}$ if $[\![A]\!] = W$. We write $x \in [\![A]\!]$ as $\mathcal{M}, x \Vdash A$, and further omit $\mathcal{M}$ if no ambiguity arises.

The link between neighbourhood and relational semantics

Given a relational frame (W, R), one can define a neighbourhood frame by taking as neighbourhoods of a world x the supersets of worlds accessible from x

$$I^R(x) \equiv \{a \mid R(x) \subseteq a\}$$

Conversely, given a neighbourhood frame (W, I) one can define a relational frame by

$$xR^I y \equiv y \in \bigcap I(x)$$

A neighbourhood frame (W, I) is *augmented* if for all $x \in W$

- $\bigcap I(x) \in I(x)$
- $I(x)$ is closed under supersets.

Equivalently, neighbourhood frame is *augmented* iff

$$a \in I(x) \equiv \bigcap I(x) \subseteq a$$

Relational frames correspond to augmented neighbourhood frames, in the sense that given a relational frame there is an equivalent augmented neighbourhood frame, and viceversa.

Lemma 2.3

For every Kripke model $\mathcal{M} = \langle W, R, \mathcal{V} \rangle$ there is an augmented neighbourhood model $\mathcal{M}^R = \langle W, I^R, [\![\ ]\!] \rangle$ such that for any w, if $\mathcal{M}, w \Vdash A$, then $\mathcal{M}^R, w \Vdash A$.

Neighbourhood and Kripke frames

Lemma 2.4

For every augmented neighbourhood model $\mathcal{M} = \langle W, I, [\![\ ]\!] \rangle$ there is a Kripke model $\mathcal{M}^I = \langle W, R^I, \mathcal{V} \rangle$ such that for any w, if $\mathcal{M}, w \Vdash A$, then $\mathcal{M}^I, w \Vdash A$.

Proof.

Essentially runs the previous proof in reverse, using instead of proving that the neighbourhood frame is augmented. QED

Combined, these lemmas show that

Theorem 2.5 (Equivalence of Kripke and neighbourhood models)

For every Kripke model, there is an augmented neighbourhood model that validates the same formulas, and vice versa.

Conditional doxastic logic CDL

CDL uses the primitive epistemic operator of conditional belief $Bel_i(C|B)$ – "*agent i believes C, given B*".

Definition 3.1 (Formula of *CDL*)

$$A \;:=\; P \mid \bot \mid \neg A \mid A \wedge A \mid A \vee A \mid A \supset A \mid Bel_i(A|A)$$

The axiomatization of CDL Board (2004) contains the rules:

Definition 3.2 (Inference rules)

(1) If $\vdash B$, then $\vdash Bel_i(B|A)$ (epistemization rule)

(2) If $\vdash A \supset\subset B$, then $\vdash Bel_i(C|A) \supset\subset Bel_i(C|B)$ (rule of logical equivalence)

Conditional doxastic logic CDL

CDL is then axiomatized as:

Definition 3.3 (Axioms of *CDL*)

Any axiomatization of the classical propositional calculus, plus:

(3) $(Bel_i(B|A) \wedge Bel_i(B \supset C|A)) \supset Bel_i(C|A)$ (distribution axiom)

(4) $Bel_i(A|A)$ (success axiom)

(5) $Bel_i(B|A) \supset (Bel_i(C|A \wedge B) \supset\subset Bel_i(C|A))$ (minimal change principle 1)

(6) $\neg Bel_i(\neg B|A) \supset (Bel_i(C|A \wedge B) \supset\subset Bel_i(B \supset C|A))$ (minimal change principle 2)

(7) $Bel_i(B|A) \supset Bel_i(Bel_i(B|A)|C)$ (positive introspection)

(8) $\neg Bel_i(B|A) \supset Bel_i(\neg Bel_i(B|A)|C)$ (negative introspection)

(9) $A \supset \neg Bel_i(\bot|A)$ (consistency axiom)

Neighbourhood models of CDL

Definition 3.4 (Multi-agent neighbourhood models)

Let $\mathcal{A}$ be a set of agents; a *multi-agent neighbourhood model* (*NM*) has the form

$$\mathcal{M} = \langle W, \{I\}_{i \in \mathcal{A}}, [\![\ ]\!] \rangle$$

where

W is a non empty set of elements called "worlds",

$[\![\ ]\!] : Atm \to \mathcal{P}(W)$ is the evaluation for atomic formulas,

for each $i \in \mathcal{A}$, $I_i : W \to \mathcal{P}(\mathcal{P}(W))$ is the neighbourhood function, satisfying the following properties:

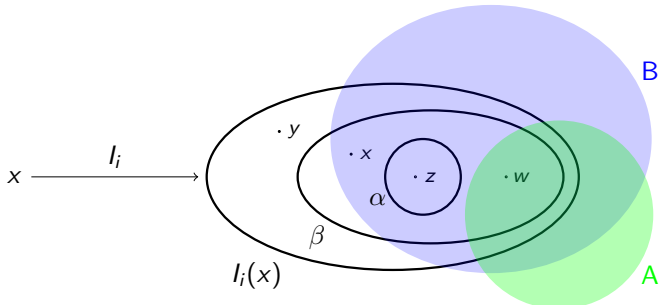
- *Non-emptiness*: $\forall \alpha \in I_i(x), \alpha \neq \emptyset$
- *Nesting*: $\forall \alpha, \beta \in I_i(x), \alpha \subseteq \beta$ or $\beta \subseteq \alpha$
- *Total reflexivity*: $\exists \alpha \in I_i(x)$ such that $x \in \alpha$
- *Local absoluteness*: If $\alpha \in I_i(x)$ and $y \in \alpha$ then $I_i(x) = I_i(y)$
- *Closure under intersection*: If $S \subseteq I_i(x)$ and $S \neq \emptyset$ then $\bigcap S \in S$ (always holds in finite models)

Conditional Belief

Truth condition

$x \Vdash Bel_i(B|A)$ *iff* $\forall \alpha \in I_i(x)(\alpha \cap [\![A]\!] = \emptyset)$ *or*

$\exists \beta \in I_i(x)(\beta \cap [\![A]\!] \neq \emptyset$ *and* $\beta \cap [\![A]\!] \subseteq [\![B]\!])$



Knowledge and simple belief

Due to Stalnaker (1998), knowledge and simple (non-conditional) belief can be defined as

Definition 4.1 (Knowledge and simple belief in *CDL*)

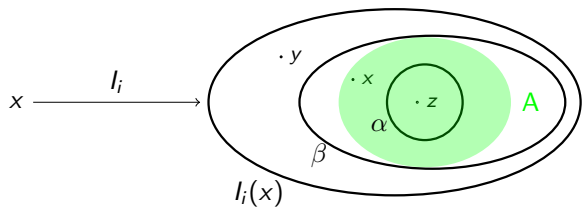
Knowledge: $K_i A \equiv Bel_i(\bot|\neg A)$
Simple belief: $Bel_i A \equiv Bel_i(A|\top)$

We unpack these definitions to obtain the truth conditions for each.

Simple belief

Truth condition

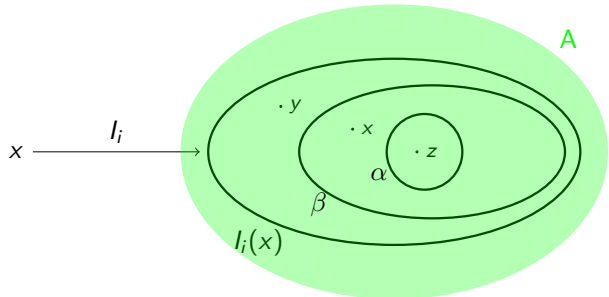
$x \Vdash Bel_i A$ *iff* $\exists \alpha \in I_i(x)\ (\alpha \subseteq [\![A]\!])$



Knowledge

Truth condition

$x \Vdash K_i A \quad \textit{iff} \quad \forall \beta \in I_i(x)\ (\beta \subseteq [\![A]\!])$



Sequent calculus G3SBK

We retain the rules of G3CDL, and extend them with rules for simple belief and knowledge, which adhere to these definitions, to obtain the sequent calculus G3SBK:

Initial sequents $x : P, \Gamma \Rightarrow \Delta, x : P$

Propositional rules: rules of **G3K** Negri (2005)

Rules for local forcing

$$\frac{x \in a, \Gamma \Rightarrow \Delta, x : A}{\Gamma \Rightarrow \Delta, a \Vdash^{\forall} A} \; R\Vdash^{\forall} (x \textit{ fresh}) \qquad \frac{x : A, x \in a, a \Vdash^{\forall} A, \Gamma \Rightarrow \Delta}{x \in a, a \Vdash^{\forall} A, \Gamma \Rightarrow \Delta} \; L\Vdash^{\forall}$$

$$\frac{x \in a, \Gamma \Rightarrow \Delta, x : A, a \Vdash^{\exists} A}{x \in a, \Gamma \Rightarrow \Delta, a \Vdash^{\exists} A} \; R\Vdash^{\exists} \qquad \frac{x \in a, x : A, \Gamma \Rightarrow \Delta}{a \Vdash^{\exists} A, \Gamma \Rightarrow \Delta} \; L\Vdash^{\exists} (x \textit{ fresh})$$

Sequent calculus G3SBK

Rules for inclusion

$$\frac{a \subseteq a, \Gamma \Rightarrow \Delta}{\Gamma \Rightarrow \Delta}\ \mathit{Ref} \qquad \frac{c \subseteq a, c \subseteq b, b \subseteq a, \Gamma \Rightarrow \Delta}{c \subseteq b, b \subseteq a, \Gamma \Rightarrow \Delta}\ \mathit{Tr}$$

$$\frac{x \in a, a \subseteq b, x \in b, \Gamma \Rightarrow \Delta}{x \in a, a \subseteq b, \Gamma \Rightarrow \Delta}\ L\subseteq$$

Sequent calculus G3SBK

Rules for semantic conditions

$$\frac{a \subseteq b, a \in I_i(x), b \in I_i(x), \Gamma \Rightarrow \Delta \quad b \subseteq a, a \in I_i(x), b \in I_i(x), \Gamma \Rightarrow \Delta}{a \in I_i(x), b \in I_i(x), \Gamma \Rightarrow \Delta}\ S$$

$$\frac{y \in a, a \in I_i(x), \Gamma \Rightarrow \Delta}{a \in I_i(x), \Gamma \Rightarrow \Delta}\ N\ (y\ \textit{fresh}) \qquad \frac{x \in a, a \in I_i(x), \Gamma \Rightarrow \Delta}{\Gamma \Rightarrow \Delta}\ T\ (a\ \textit{fresh})$$

$$\frac{a \in I_i(x), y \in a, b \in I_i(x), b \in I_i(y), \Gamma \Rightarrow \Delta}{a \in I_i(x), y \in a, b \in I_i(x), \Gamma \Rightarrow \Delta}\ A_1$$

$$\frac{a \in I_i(x), y \in a, b \in I_i(x), b \in I_i(y), \Gamma \Rightarrow \Delta}{a \in I_i(x), y \in a, b \in I_i(y), \Gamma \Rightarrow \Delta}\ A_2$$

$$\frac{a \in I_i(x), y \in a, a \in I_i(y), \Gamma \Rightarrow \Delta}{a \in I_i(x), y \in a, \Gamma \Rightarrow \Delta}\ A_1^*$$

Sequent calculus G3SBK

Rules for knowledge and belief

$$\frac{a \in I_i(x), \Gamma \Rightarrow \Delta, a \Vdash^{\forall} A}{\Gamma \Rightarrow \Delta, x\colon K_i A}\ RK\ (a\ \textit{fresh})$$

$$\frac{a \in I_i(x), x\colon K_i A, a \Vdash^{\forall} A, \Gamma \Rightarrow \Delta}{a \in I_i(x), x\colon K_i A, \Gamma \Rightarrow \Delta}\ LK$$

$$\frac{a \in I_i(x), \Gamma \Rightarrow \Delta, x\colon Bel_i A, a \Vdash^{\forall} A}{a \in I_i(x), \Gamma \Rightarrow \Delta, x\colon Bel_i A}\ RSB$$

$$\frac{a \in I_i(x), a \Vdash^{\forall} A, \Gamma \Rightarrow \Delta}{x\colon Bel_i A, \Gamma \Rightarrow \Delta}\ LSB\ (a\ \textit{fresh})$$

Structural properties

Here we extend the proofs of structural properties from the Girlando et al. (2018) with added rules. We start with the notion of the weight of the formula:

Definition 5.1 (Weight of a labelled formula)

The weight of the labelled formula $\mathcal{F}$ is the pair $(\mathtt{w}(p(\mathcal{F})), \mathtt{w}(l(\mathcal{F})))$, where $l(\mathcal{F})$ is the label of $\mathcal{F}$, and

$$\mathtt{w}(x) = 0,\ \mathtt{w}(a) = 1,$$

and $p(\mathcal{F})$ is the part of $\mathcal{F}$ without the label and the forcing relation, and

$$\begin{aligned}
&\mathtt{w}(P) = \mathtt{w}(\top) = 1,\\
&\mathtt{w}(A \circ B) = \mathtt{w}(A) + \mathtt{w}(B) + 1,\ \circ \in \{\vee, \&, \supset\},\\
&\mathtt{w}(\neg A) = \mathtt{w}(A) + 2,\\
&\mathtt{w}(B|A) = \mathtt{w}(A) + \mathtt{w}(B) + 2\\
&\mathtt{w}(Bel_i(B|A)) = \mathtt{w}(B|A) + 1.\\
&\mathtt{w}(Bel_i A) = \mathtt{w}(A) + 4\\
&\mathtt{w}(K_i A) = \mathtt{w}(A) + 6
\end{aligned}$$

Weights of labelled formulas are ordered lexicographically.

Structural properties

Lemma 5.2 (Axiom generalization)

For any labelled formula $\mathcal{F}$, the sequent $\mathcal{F}, \Gamma \Rightarrow \Delta, \mathcal{F}$ is derivable.

Lemma 5.3 (Substitution)

If $\vdash_n \Gamma \Rightarrow \Delta$ then $\vdash_n \Gamma(y/x) \Rightarrow \Delta(y/x)$; if $\vdash_n \Gamma \Rightarrow \Delta$ then $\vdash_n \Gamma(a/b) \Rightarrow \Delta(a/b)$.

Lemma 5.4 (Weakening)

Weakening is height-preserving admissible.

Lemma 5.5 (Invertibility)

All the rules of G3SBK are height-preserving invertible.

Lemma 5.6 (Contraction)

The rules of left and right contraction are height-preserving admissible.

Structural properties

Theorem 5.7 (Cut)

Cut is admissible.

Proof is by primary induction on the weight of the formula and secondary induction on the sum of the heights of the premises of cut. We illustrate for the case where the cut formula is principal in both premises and of the form $x : K_i A$.

Structural properties

Proof.

$$\dfrac{\dfrac{b \in I_i(x), \Gamma \Rightarrow \Delta, b \Vdash^{\forall} A}{\Gamma \Rightarrow \Delta, x\colon K_i A}\, RK \qquad \dfrac{a \in I_i(x), x\colon K_i A, a \Vdash^{\forall} A, \Gamma' \Rightarrow \Delta'}{a \in I_i(x), x\colon K_i A, \Gamma' \Rightarrow \Delta'}\, LK}{a \in I_i(x), \Gamma', \Gamma \Rightarrow \Delta, \Delta'}\, Cut$$

This is transformed into:

$$\dfrac{\dfrac{b \in I_i(x), \Gamma \Rightarrow \Delta, b \Vdash^{\forall} A}{a \in I_i(x), \Gamma \Rightarrow \Delta, a \Vdash^{\forall} A}\, Lm\ 5.3 \qquad \dfrac{\dfrac{b \in I_i(x), \Gamma \Rightarrow \Delta, b \Vdash^{\forall} A}{\Gamma \Rightarrow \Delta, x\colon K_i A}\, RK \qquad a \in I_i(x), x\colon K_i A, a \Vdash^{\forall} A, \Gamma' \Rightarrow \Delta'}{a \in I_i(x), a \Vdash^{\forall} A, \Gamma', \Gamma \Rightarrow \Delta, \Delta'}\, Cut_1}{\dfrac{a \in I_i(x), a \in I_i(x), \Gamma', \Gamma, \Gamma \Rightarrow \Delta, \Delta, \Delta'}{a \in I_i(x), \Gamma', \Gamma \Rightarrow \Delta, \Delta'}\, Lm\ 5.6}\, Cut_2$$

The application of the Cut rule labeled Cut_1 is of lower height, and that labeled Cut_2 is of lower weight (recall again the lexicographical ordering). QED

Properties of knowledge

We can show that:

Theorem 6.1 (K_i is S5)

K_i is (at least) an S5 operator. Specifically, the following hold of it:

(i) $K_i A \supset A$
(ii) $K_i A \supset K_i K_i A$
(iii) $\neg K_i A \supset K_i \neg K_i A$

In fact, we can be more fine-grained and relate semantic conditions to properties of K_i.

Factivity

Factivity *(i)* $K_i A \supset A$ of knowledge follows from **total reflexivity**:

(i)

$$\dfrac{\dfrac{\dfrac{x \in a, a \in I_i(x), x : K_i A, a \Vdash^{\forall} A, x : A \Rightarrow x : A}{x \in a, a \in I_i(x), x : K_i A, a \Vdash^{\forall} A \Rightarrow x : A}\, L\Vdash^{\forall}}{x \in a, a \in I_i(x), x : K_i A \Rightarrow x : A}\, LK}{x : K_i A \Rightarrow x : A}\, T$$

Positive introspection

Positive introspection *(ii)* $K_iA \supset K_iK_iA$ for knowledge follows from one direction of **local absoluteness**:

(ii)

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{z : A, b \Vdash^{\forall} A, b \in I_i(x), z \in b, b \in I_i(y), y \in a, a \in I_i(x), x : K_iA \Rightarrow z : A}{b \Vdash^{\forall} A, b \in I_i(x), z \in b, b \in I_i(y), y \in a, a \in I_i(x), x : K_iA \Rightarrow z : A}\, L\Vdash^{\forall}}{b \in I_i(x), z \in b, b \in I_i(y), y \in a, a \in I_i(x), x : K_iA \Rightarrow z : A}\, LK}{z \in b, b \in I_i(y), y \in a, a \in I_i(x), x : K_iA \Rightarrow z : A}\, A_2}{b \in I_i(y), y \in a, a \in I_i(x), x : K_iA \Rightarrow b \Vdash^{\forall} A}\, R\Vdash^{\forall}}{y \in a, a \in I_i(x), x : K_iA \Rightarrow y : K_iA}\, RK}{a \in I_i(x), x : K_iA \Rightarrow a \Vdash^{\forall} K_iA}\, R\Vdash^{\forall}}{x : K_iA \Rightarrow x : K_iK_iA}\, RK$$

Negative introspection

Negative introspection *(iii)* $\neg K_i A \supset K_i \neg K_i A$ for knowledge follows from the other direction of **local absoluteness**:

(iii)

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{a \in I_i(z), z : K_i A, a \Vdash^{\forall} A, y : A, z \in b, y \in a, b \in I_i(x), a \in I_i(x) \Rightarrow y : A}{a \in I_i(z), z : K_i A, a \Vdash^{\forall} A, z \in b, y \in a, b \in I_i(x), a \in I_i(x) \Rightarrow y : A}\, L\Vdash^{\forall}}{a \in I_i(z), z : K_i A, z \in b, y \in a, b \in I_i(x), a \in I_i(x) \Rightarrow y : A}\, LK}{z : K_i A, z \in b, y \in a, b \in I_i(x), a \in I_i(x) \Rightarrow y : A}\, A_1}{z \in b, y \in a, b \in I_i(x), a \in I_i(x) \Rightarrow z : \neg K_i A, y : A}\, R\neg}{y \in a, b \in I_i(x), a \in I_i(x) \Rightarrow b \Vdash^{\forall} \neg K_i A, y : A}\, R\Vdash^{\forall}}{b \in I_i(x), a \in I_i(x) \Rightarrow b \Vdash^{\forall} \neg K_i A, a \Vdash^{\forall} A}\, R\Vdash^{\forall}}{a \in I_i(x) \Rightarrow x : K_i \neg K_i A, a \Vdash^{\forall} A}\, RK}{\Rightarrow x : K_i \neg K_i A, x : K_i A}\, RK}{x : \neg K_i A \Rightarrow x : K_i \neg K_i A}\, L\neg$$

Formal formal epistemology

- Safe belief can be characterized as the doxastic attitude which is stable under revision with arbitrary *true* information
- Knowledge has the stronger property of stability under revision with *arbitrary* (including deceitful) information.

(1) Stability of safe belief

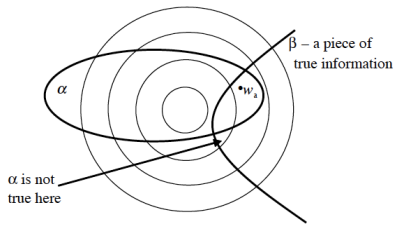
$$\dfrac{\dfrac{\dfrac{\vdots}{\ldots x \in a, x : P \Rightarrow a \Vdash^{\exists} P} \quad \dfrac{\vdots}{\ldots x \in a, a \Vdash^{\forall} A, x : P, b \Vdash^{\exists} P \Rightarrow a \Vdash^{\forall} P \supset A}}{a \in I_i(x), b \in I_i(x), x \in a, a \Vdash^{\forall} A, x : P, b \Vdash^{\exists} P \Rightarrow x \Vdash_i A|P}\ RC}{a \in I_i(x), x \in a, a \Vdash^{\forall} A, x : P \Rightarrow x : Bel(A|P)}\ RB}{x : Bel_i^{\mathsf{Sf}} A, x : P \Rightarrow x : Bel(A|P)}\ LSF$$

(2) stability of knowledge

$$\dfrac{\dfrac{\dfrac{\dfrac{\vdots}{\ldots\ a\Vdash^{\exists} P, x : KA \Rightarrow a\Vdash^{\exists} P} \quad \dfrac{\vdots}{a \in I_i(x), a\Vdash^{\forall} A, a\Vdash^{\exists} P, x : KA \Rightarrow a\Vdash^{\forall} P \supset A}\ RC}{a \in I_i(x), a\Vdash^{\forall} A, a\Vdash^{\exists} P, x : KA \Rightarrow x \Vdash_i A|P}\ LK}{a \in I_i(x), a\Vdash^{\exists} P, x : KA \Rightarrow x \Vdash_i A|P}\ RB}{x : KA \Rightarrow x : Bel(A|P)}$$

Stability under arbitrary revision gives knowledge

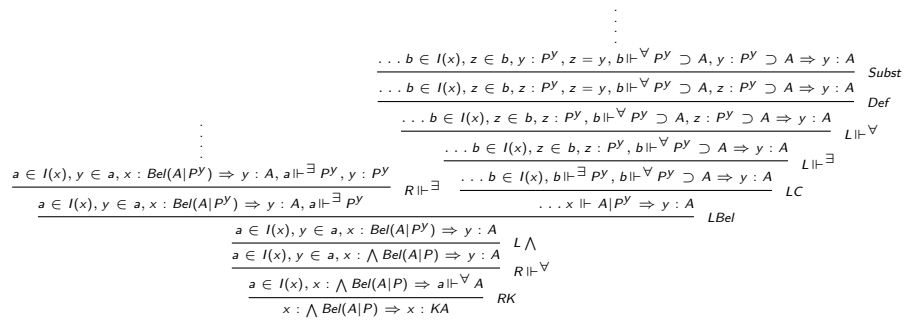
Graphic proof:



Rott (2004)

The picture "shows" that if we deny knowledge of A (here α), then it is possible to find a proposition B (β) that destroys conditional belief in A. Hidden assumption for the graphic proof to work: for all b such that $b \| \vdash^{\exists} B$ we have $y \in b$ (y is the point indicated by the arrow)

Stability under arbitrary revision gives knowledge (cont.) Sequent calculus proof: P^y denotes a proposition which is true exactly at y and $\bigwedge$ is an infinitary conjunction that ranges over all the propositions



Stability under true revision gives safe belief

$$\vdots$$

$$\cfrac{\cfrac{\dots b \in I(x), y \in b, y : P^x, b \Vdash^{\forall} P^x \supset A \Rightarrow x : A}{\dots b \in I(x), b \Vdash^{\exists} P^x, b \Vdash^{\forall} P^x \supset A \Rightarrow x : A}\, L\Vdash^{\exists}}{\{x\} \in I(x), x \Vdash A|P^x, x : Bel(A|P^x) \Rightarrow x : Bel_i^{\mathsf{Sf}} A, x : A}$$

$$\cfrac{\{x\} \in I(x), x : Bel(A|P^x) \Rightarrow x : Bel_i^{\mathsf{Sf}} A, x : A, \{x\} \Vdash^{\exists} P \qquad \{x\} \in I(x), x \Vdash A|P^x, x : Bel(A|P^x) \Rightarrow x : Bel_i^{\mathsf{Sf}} A, x : A}{\{x\} \in I(x), x : Bel(A|P^x) \Rightarrow x : Bel_i^{\mathsf{Sf}} A, x : A}\, LB$$

$$\cfrac{\{x\} \in I(x), x : Bel(A|P^x) \Rightarrow x : Bel_i^{\mathsf{Sf}} A, x : A}{\{x\} \in I(x), x : \bigwedge_{x \Vdash P} Bel(A|P) \Rightarrow x : Bel_i^{\mathsf{Sf}} A, x : A}\, L\wedge$$

$$\cfrac{\cfrac{\{x\} \in I(x), x : \bigwedge_{x \Vdash P} Bel(A|P) \Rightarrow x : Bel_i^{\mathsf{Sf}} A, x : A}{\{x\} \in I(x), x : \bigwedge_{x \Vdash P} Bel(A|P) \Rightarrow x : Bel_i^{\mathsf{Sf}} A}\, RSF}{x : \bigwedge_{x \Vdash P} Bel(A|P) \Rightarrow x : Bel_i^{\mathsf{Sf}} A}$$

- Enough to require stability under revision with propositions true exactly in the point of evaluation to get safe belief
- An extra assumption, $\{x\} \in I(x)$, is needed. Equivalent to validity of $A \supset Bel(A)$ (in general, belief in true statements is not a requirement).
- Indefeasible knowledge is a normal, S5 modality.

Argument against a perfect believer

The paradox of the perfect believer (Baltag and Smets (2008)) is a derivation of an implication from *belief of knowledge* to *knowledge* using (apparently) reasonable assumption on the classical epistemic/doxastic operators:

Infallibility, $\neg Bel_i \bot$,
Knowledge implies belief, $K_i A \supset Bel_i A$ and
Introspection about belief, $Bel_i A \supset K_i Bel_i A$.

We take what is needed to have the same assumptions used in the puzzle.

Argument against a perfect believer

Infallibility, $\neg Bel_i\, \bot$, follows from N (non-emptiness):

$$\dfrac{\dfrac{\dfrac{\dfrac{}{y\colon \bot, y \in a, a \in I(x), a \Vdash^{\forall} \bot \Rightarrow}\, L\bot}{y \in a, a \in I(x), a \Vdash^{\forall} \bot \Rightarrow}\, L\Vdash^{\forall}}{a \in I(x), a \Vdash^{\forall} \bot \Rightarrow}\, N}{x\colon Bel_i \bot \Rightarrow}\, LSB$$

Knowledge implies belief $K_i A \supset Bel_i A$ is valid thanks to T (*total reflexivity*):

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{y \in a, x \in a, a \in I(x), x\colon K_i A, a \Vdash^{\forall} A, y\colon A \Rightarrow x\colon Bel_i A, y\colon A}{y \in a, x \in a, a \in I(x), x\colon K_i A, a \Vdash^{\forall} A \Rightarrow x\colon Bel_i A, y\colon A}\, L\Vdash^{\forall}}{x \in a, a \in I(x), x\colon K_i A, a \Vdash^{\forall} A \Rightarrow x\colon Bel_i A, a \Vdash^{\forall} A}\, R\Vdash^{\forall}}{x \in a, a \in I(x), x\colon K_i A, a \Vdash^{\forall} A \Rightarrow x\colon Bel_i A}\, RSB}{x \in a, a \in I(x), x\colon K_i A \Rightarrow x\colon Bel_i A}\, LK}{x\colon K_i A \Rightarrow x\colon Bel_i A}\, T$$

Argument against a perfect believer

Introspection about belief $Bel_i A \supset K_i Bel_i A$ is valid thanks to A rules (*local absoluteness*):

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{z \in a, a \in I(y), y \in b, b \in I(x), a \in I(x), a \Vdash^{\forall} A, z\colon A \Rightarrow y\colon Bel_i A, z\colon A}{z \in a, a \in I(y), y \in b, b \in I(x), a \in I(x), a \Vdash^{\forall} A \Rightarrow y\colon Bel_i A, z\colon A}\, L\Vdash^{\forall}}{a \in I(y), y \in b, b \in I(x), a \in I(x), a \Vdash^{\forall} A \Rightarrow y\colon Bel_i A, a \Vdash^{\forall} A}\, R\Vdash^{\forall}}{a \in I(y), y \in b, b \in I(x), a \in I(x), a \Vdash^{\forall} A \Rightarrow y\colon Bel_i A}\, RSB}{y \in b, b \in I(x), a \in I(x), a \Vdash^{\forall} A \Rightarrow y\colon Bel_i A}\, A_1}{b \in I(x), a \in I(x), a \Vdash^{\forall} A \Rightarrow b \Vdash^{\forall} Bel_i A}\, R\Vdash^{\forall}}{a \in I(x), a \Vdash^{\forall} A \Rightarrow x\colon K_i Bel_i A}\, RK}{x\colon Bel_i A \Rightarrow x\colon K_i Bel_i A}\, LSB$$

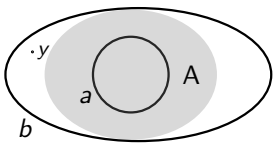
Argument against a perfect believer

The derivation of the paradox proceeds as follows:

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dots, y : A \Rightarrow y : A}{\dots, y \in b, b \in I_i(y), b \Vdash^{\forall} A \Rightarrow y : A}\, L\Vdash^{\forall}}{\dots, b \in I_i(y), y : K_i A \Rightarrow y : A}\, LK_i}{\dots, y : K_i A \Rightarrow y : A}\, A_1}{y \in a, \dots a \Vdash^{\forall} K_i A \Rightarrow y : A}\, L\Vdash^{\forall}}{b \subseteq a, \dots, y \in b, a \Vdash^{\forall} K_i A \Rightarrow y : A}\, L\subseteq \qquad \dfrac{\dfrac{\dfrac{\dfrac{\dots, y : A \Rightarrow y : A}{b \in I_i(z), z \in a, \dots, y \in b, b \Vdash^{\forall} A \Rightarrow y : A}\, L\Vdash^{\forall}}{b \in I_i(z), z \in a, \dots, y \in b, z : K_i A \Rightarrow y : A}\, LK_i}{z \in a, \dots, y \in b, z : K_i A \Rightarrow y : A}\, A_1}{a \subseteq b, a \in I_i(x), b \in I_i(x), y \in b, a \Vdash^{\forall} K_i A \Rightarrow y : A}\, N}{a \in I_i(x), b \in I_i(x), y \in b, a \Vdash^{\forall} K_i A \Rightarrow y : A}\, S$$

$$\dfrac{\dfrac{\dfrac{a \in I_i(x), b \in I_i(x), y \in b, a \Vdash^{\forall} K_i A \Rightarrow y : A}{a \in I_i(x), b \in I(x)_i, a \Vdash^{\forall} K_i A \Rightarrow a \Vdash^{\forall} A}\, R\Vdash^{\forall}}{a \in I_i(x), a \Vdash^{\forall} K_i A \Rightarrow x : K_i A}\, RK}{x : Bel_i K_i A \Rightarrow x : K_i A}\, LSB$$

Obviously, without N proof search stops on the right and we obtain the countermodel from the failed proof search:



Thank you/Grazie/Hvala!

References

A. Baltag and S. Smets. *New Perspectives on Games and Interaction*, volume 4 of *Texts in Logic and Games*, pages 9 – 31. Amsterdam University Press, 2008. ISBN 9789089640574.

Oliver Board. Dynamic interactive epistemology. *Games and Economic Behavior*, 49(1):49 – 80, 2004. ISSN 0899-8256. doi: https://doi.org/10.1016/j.geb.2003.10.006.

M. Girlando, S. Negri, N. Olivetti, and V. Risch. Conditional beliefs: from neighbourhood semantics to sequent calculus. *The Review of Symbolic Logic*, 11(4):736 – 779, 2018. doi: 10.1017/S1755020318000023.

Sara Negri. Proof analysis in modal logic. *Journal of Philosophical Logic*, 34:507 – 544, 2005. doi: https://doi.org/10.1007/s10992-005-2267-3.

Robert Stalnaker. Belief revision in games: Forward and backward induction. *Mathematical Social Sciences*, 36(1):31 – 56, 1998. ISSN 0165-4896. doi: https://doi.org/10.1016/S0165-4896(98)00007-9. URL `http://www.sciencedirect.com/science/article/pii/S0165489698000079`.