Belief, Knowledge and the Topology of Evidence

Alexandru Baltag

ILLC, University of Amsterdam

June 14, 2016

Joint work with Nick Bezhanishvili, Aybüke Özgün and Sonja Smets.

Based on developments arising from work of Johan van Benthem and Eric Pacuit.

Models for Knowledge and Belief

- Relational Models
  - Kripke Models
  - Plausibility Models
- Neighborhood Models
  - Grove spheres
  - Topological Models
  - Subset Spaces

- an agent's rational belief is based on the available evidence.
- evidence is represented both semantically and syntactically.
- belief and knowledge are not primitive, they are built from evidence pieces.

Evidence Models (van Benthem and Pacuit)

A **(uniform)** **evidence model** is a tuple $\mathcal{M} = (X, E_0, V)$, where

- X is a non-empty set of states;
- $\emptyset \neq E_0 \subseteq \mathcal{P}(X)$ s.t. $\emptyset \notin E_0$ and $X \in E_0$;
- $V : Prop \to \mathcal{P}(X)$, where *Prop* is a countable set of propositional variables.

E_0 is called the set of *basic evidence sets* or *pieces of evidence*

Consistent (finite) combination of evidence pieces

Given an evidence model $\mathcal{M} = (X, E_0, V)$, we define

- A **body of evidence** is a family $F \subseteq E_0$ of evidence pieces s.t. every finitely many of them are mutually consistent:

$$(\forall F' \subseteq_{fin} F)(F' \neq \emptyset \Rightarrow \bigcap F' \neq \emptyset)$$

- $\mathcal{F}$:= the family of all bodies of evidence over $\mathcal{M}$
- $\mathcal{F}^{fin}$:= the family of all finite bodies of evidence over $\mathcal{M}$

(Combined) Evidence

Given an evidence model $\mathcal{M} = (X, E_0, V)$, we define

- A **(combined) evidence** is any non-empty intersection of finitely many pieces of evidence.
- E is the **family of all (combined) evidence**:

$$E := \{\bigcap F \mid \; F \in \mathcal{F}^{fin}\}$$

$e \in E_0$: a basic piece of *direct* evidence.

$e \in E$: *indirect* evidence obtained by combining finitely many pieces of direct evidence.

An evidence e is **factive** (or “correct") at world x if $x \in e$.

Observation: E is a topological base on X.

Topological Evidence Models (topo-e-models)

The **evidential topology** τ_E is the topology generated by E: i.e., the smallest topology $\tau \supseteq E_0$.

An **topo-e-model** is a tuple $\mathcal{M} = (X, E_0, \tau, V)$, where

- $\mathcal{M} = (X, E_0, V)$ is an evidence model,
- $\tau = \tau_E$ is the *evidential topology*

The **evidential plausibility order** $\sqsubseteq_E$ is the specialization preorder wrt τ_E:

$$\begin{aligned} x \sqsubseteq_E y \ \text{ iff } \ & \forall U \in \tau_E (x \in U \Rightarrow y \in U) \\ \text{iff } \ & \forall e \in E_0 \, (x \in e \Rightarrow y \in e) \\ \text{iff } \ & \forall e \in E \, (x \in e \Rightarrow y \in e). \end{aligned}$$

We denote the **strict order** by

$$x \sqsubset_E y \ \text{ iff } \ x \sqsubseteq_E y \wedge y \not\sqsubseteq_E x.$$

F supports the proposition P

A body of evidence F **supports** P iff $\bigcap F \subseteq P$.

- **strength order $\subseteq$ on $\mathcal{F}$:**

$$F \subseteq F' \ \text{ means thar: } \ F' \text{ is at least as strong as } F$$

$$Max_{\subseteq}(\mathcal{F}) \ := \ \{F \in \mathcal{F} \mid \forall F' \in \mathcal{F}(F \subseteq F' \Rightarrow F = F')\}$$

Observation: $Max_{\subseteq}(\mathcal{F}) \neq \emptyset$ (Zorn's Lemma)

Evidential Support and Strength Order

A (combined) evidence e **supports** P (or e is "*evidence for*" P) iff $e \subseteq P$.

NOTE: **strength order** $\subseteq$ **goes opposite ways**:

- *on bodies of evidence $\mathcal{F}$:*

$$F \subseteq F' := F' \text{ is at least as strong as } F$$

- *on evidence E:*

$$e \supseteq e' := e' \text{ is at least as strong as } e$$

Evidence and Belief

Syntax of van Benthem and Pacuit:

$$\mathcal{L}_0 := p \mid \neg\varphi \mid \varphi \wedge \varphi \mid E_0\varphi \mid B\varphi \mid \forall\varphi$$

$E_0\varphi$:= the agent has a *basic evidence for* φ.

$B\varphi$:= the agent *believes* φ.

$\forall\varphi$:= the agent *infallibly knows* φ (i.e., φ is true in all possible worlds).

Semantics of van Benthem and Pacuit

Given an evidence model $\mathcal{M} = (X, E_0, V)$ and $x \in X$, we define a *satisfaction relation* $\mathcal{M}, x \models \varphi$ and *interpretation map* $[\![\varphi]\!]^{\mathcal{M}} := \{x \in X | \mathcal{M}, x \models \varphi\}$, by using the valuation V for atomic sentences and the usual clauses for Boolean connectives, and in rest putting:

$$
\begin{array}{lll}
\mathcal{M}, x \models \forall\varphi & \text{iff} & [\![\varphi]\!]^{\mathcal{M}} = X \\
\mathcal{M}, x \models E_0\varphi & \text{iff} & \exists e \in E_0\,(e \subseteq [\![\varphi]\!]^{\mathcal{M}}) \\
\mathcal{M}, x \models B\varphi & \text{iff} & (\forall F \in Max_{\subseteq}(\mathcal{F}))(\bigcap F \subseteq [\![\varphi]\!]^{\mathcal{M}}) \\
 & \text{iff} & Max_{\sqsubseteq_E} X \subseteq [\![\varphi]\!]^{\mathcal{M}}
\end{array}
$$

where

$$Max_{\sqsubseteq_E} X \; := \; \{y \in X | \forall z \in X (y \not\sqsubset_E z)\}$$

is the set of maximal worlds wrt $\sqsubseteq_E$ (“most plausible worlds").

Forming Beliefs based on (Fallible) Evidence

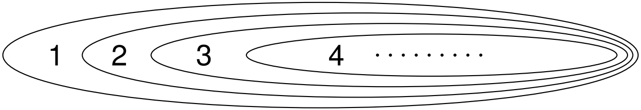
The main idea behind this semantics of belief seems to be that:

Given fallible, and possibly mutually inconsistent, pieces of evidence, the rational agent tries to form consistent beliefs, by looking at all maximally consistent "blocks" of evidence and believing whatever is entailed by all of them.

- "Having evidence for φ need not imply belief."
- "When forming beliefs, the agent should take all her available evidence for and against φ into account."
- belief is entailed by all the "strongest" evidence.
- when E_0 is finite (and in many other cases), beliefs are consistent ($\neg B\bot$)
- DRAWBACK: $B\bot$ can hold in "bad" models.

Example 1

$\mathcal{M} = (\mathbb{N}, E_0, V)$ with $E_0 = \{[n, \infty) \mid n \in \mathbb{N}\}$ and $V(p) = \emptyset$.



$Max_{\subseteq}\mathcal{F} = \{E_0\}$ and $\bigcap E_0 = \emptyset \Rightarrow B\bot$ holds in $\mathcal{M}$.

Our work

- uses the same models with a particular focus on the *evidential topology*
- different notions of evidence: basic, combined, factive, misleading etc.
- topological formalization of *argument* and *justification*
- topologically interpreted, evidence-based, *consistent* notions of **(justified) belief** and **(defeasible) knowledge**
- complete axiomatizations, finite model property
- adapt the van-Benthem-Pacuit dynamics of evidence management to this modified setting

Our Largest Evidence language

$$\mathcal{L}_1 := p \mid \neg\varphi \mid \varphi \wedge \varphi \mid \forall\varphi \mid E_0\varphi \mid E\varphi \mid \Box_0\varphi \mid \Box\varphi$$

$\forall\varphi$:= the agent infallibly knows φ.

$E_0\varphi$:= the agent has a *basic (piece of) evidence* supporting φ.

$E\varphi$:= the agent has *(combined) evidence* for φ.

$\Box_0\varphi$:= the agent has a *factive piece of evidence* for φ.

$\Box\varphi$:= the agent has *factive (combined) evidence* for φ.

Our semantics

$$\mathcal{L}_1 := p \mid \neg\varphi \mid \varphi \wedge \varphi \mid \forall\varphi \mid E_0\varphi \mid E\varphi \mid \Box_0\varphi \mid \Box\varphi$$

Given a topo-e-model $\mathcal{M} = (X, E_0, \tau, V)$ and $x \in X$,

$$\begin{array}{lll}
\mathcal{M}, x \models \forall\varphi & \text{iff} & [\![\varphi]\!] = X \\
\mathcal{M}, x \models E_0\varphi & \text{iff} & (\exists e \in E_0)(e \subseteq [\![\varphi]\!]^{\mathcal{M}}) \\
\mathcal{M}, x \models E\varphi & \text{iff} & (\exists e \in E)(e \subseteq [\![\varphi]\!]^{\mathcal{M}}) \\
\mathcal{M}, x \models \Box_0\varphi & \text{iff} & (\exists e \in E_0)(x \in e \subseteq [\![\varphi]\!]^{\mathcal{M}}) \\
\mathcal{M}, x \models \Box\varphi & \text{iff} & (\exists e \in E)(x \in e \subseteq [\![\varphi]\!]^{\mathcal{M}})
\end{array}$$

Our semantics

$$\mathcal{L}_1 := p \mid \neg\varphi \mid \varphi \wedge \varphi \mid \forall\varphi \mid E_0\varphi \mid E\varphi \mid \Box_0\varphi \mid \Box\varphi$$

Given a topo-e-model $\mathcal{M} = (X, E_0, \tau, V)$ and $x \in X$,

$$\begin{array}{lll}
\mathcal{M}, x \models \forall\varphi & \text{iff} & [\![\varphi]\!] = X \\
\mathcal{M}, x \models E_0\varphi & \text{iff} & (\exists e \in E_0)(e \subseteq [\![\varphi]\!]^{\mathcal{M}}) \\
\mathcal{M}, x \models E\varphi & \text{iff} & (\exists e \in E)(e \subseteq [\![\varphi]\!]^{\mathcal{M}}) \\
\mathcal{M}, x \models \Box_0\varphi & \text{iff} & (\exists e \in E_0)(x \in e \subseteq [\![\varphi]\!]^{\mathcal{M}}) \\
\mathcal{M}, x \models \Box\varphi & \text{iff} & (\exists e \in E)(x \in e \subseteq [\![\varphi]\!]^{\mathcal{M}})
\end{array}$$

Observations: $[\![\Box\varphi]\!]^{\mathcal{M}} = Int[\![\varphi]\!]^{\mathcal{M}}$

McKinsey-Tarski topological semantics

Argument and Justification

- An *argument* for P is a disjunction $U = \bigcup_{i \in I} e_i$ such that $e_i \subseteq P$ for all $i \in I$,

  i.e. $U \in \tau$ with $U \subseteq P$ and $IntP$ is the weakest.

  Essentially, a set of worlds is an argument (for something) iff it is open (in τ_E).
- A *justification* for P is an argument U for P that is consistent with every available evidence,

  i.e. $U \in \tau$ such that $U \subseteq P$ and $U \cap e \neq \emptyset$ for all $e \in E$,

  i.e. $U \in \tau$ such that $U \subseteq P$ and $Cl(U) = X$,

  i.e. U is a *dense open* subset of P.
- An argument (justification) U is correct at x iff $x \in U$.

Our Notion of (Justified) Belief

***P* is believed iff it is entailed by all "sufficiently strong" evidence.**

Indeed, **the following are equivalent**:

- $B\varphi$ holds;
- every finite body of evidence can be strengthened to a finite body supporting φ;
- $\forall F \in \mathcal{F}^{fin}\ \exists F' \in \mathcal{F}^{fin}(F \subseteq F' \wedge \bigcap F' \subseteq [\![\varphi]\!])$;
- every evidence e can be strengthened to some evidence e' that supports φ;
- $\forall e \in E\ \exists e' \in E(e' \subseteq e \cap [\![\varphi]\!])$;
- $\forall U \in \tau \setminus \{\emptyset\}\ \exists U' \in \tau \setminus \{\emptyset\}(U' \subseteq U \cap [\![\varphi]\!])$;
- φ includes a dense open set;
- $Int[\![\varphi]\!]$ is dense (i.e. $Cl(Int[\![\varphi]\!]) = X$);
- there is an argument for φ consistent with every evidence;
- the agent has a justification for φ.

(Justified) Belief

$B\varphi$ holds iff $Cl(Int[\![\varphi]\!]) = X$

iff $Int(Cl[\![\neg\varphi]\!]) = \emptyset$

iff $[\![\neg\varphi]\!]$ is *nowhere dense*

iff φ is true in "*almost all*" epistemically possible states

- Our B coincides with the one of van Benthem-Pacuit when E_0 is finite.
- But our belief is always consistent: $B\perp$ never holds, since $Cl(Int(\emptyset)) = \emptyset$.
- The logic of belief is $KD45$.

(Defeasible) Knowledge

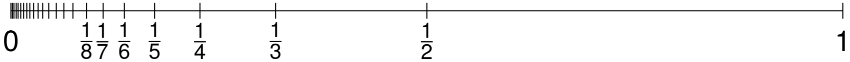
Given a topo-e-model $\mathcal{M} = (X, E_0, \tau, V)$,

$$
\begin{aligned}
K\varphi \text{ holds at } x \text{ iff } [\![\varphi]\!] \text{ includes a dense open neighborhood of } x \quad (1) \\
\text{iff } \exists U \in \tau (x \in U \subseteq [\![\varphi]\!] \wedge Cl(U) = X) \\
\text{iff } x \in Int[\![\varphi]\!] \text{ and } Cl(Int[\![\varphi]\!]) = X \\
\text{iff } \Box\varphi \wedge B\varphi \text{ holds at } x \quad (2) \\
\text{iff the agent has a } \textit{correct} \text{ justification for } \varphi \text{ at } x
\end{aligned}
$$

- Knowledge is *correctly justified* belief.
- The logic of knowledge is $S4.2$.

Example 2

$\mathcal{M} = ([0,1], E_0, \tau, V)$ with $E_0 = \{(a,b) \cap [0,1] \mid a, b \in \mathbb{R}, a < b\}$



$P = [0,1] \setminus \{\frac{1}{n} : n \in \mathbb{N}\}$ and $\neg P = \{\frac{1}{n} : n \in \mathbb{N}\}$

e.g. $U = \bigcup_{n \geq 1} (\frac{1}{n+1}, \frac{1}{n}) \subseteq P$ is dense and open.

- BP holds (everywhere)
- KP holds at every state in P, <u>except at 0</u>:

$$0 \notin IntP$$

Our knowledge is *defeasible*-1

- *irrevocable knowledge*: cannot be defeated any evidence gathered later
- *in-defeasible knowledge*: cannot be defeated any factive evidence gathered later

Defeasibility Theory of Knowledge (Lehrer, Klein etc): an agent “in-defeasibly knows" P iff:

1. P is true
2. she believes that P is true
3. her belief in P cannot be defeated by new *factive* information. *stable belief*

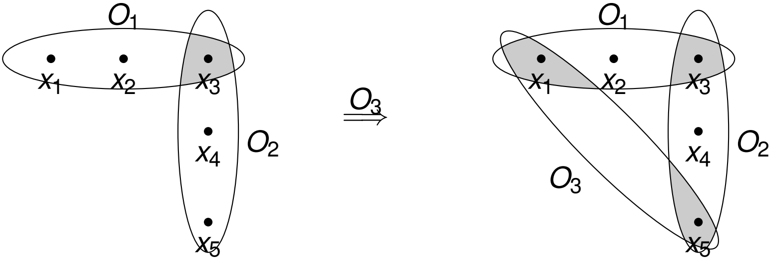
Our knowledge is *defeasible*-2

- *irrevocable knowledge*: cannot be defeated any evidence gathered later
- *in-defeasible knowledge*: cannot be defeated any factive evidence gathered later

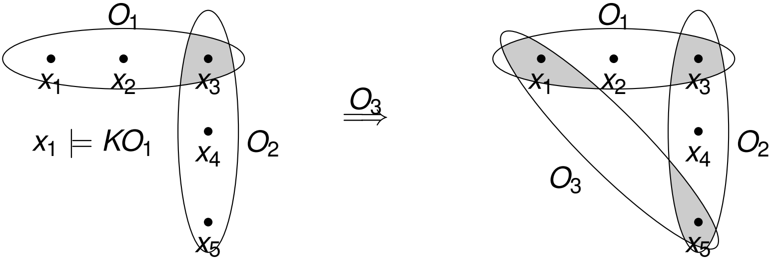
an agent knows P :

1. P is true
2. she believes that P is true
3. her belief in P cannot be defeated by new *factive* information. *stable belief*
4. its justification is undefeated by new *factive* information. *stable justification*

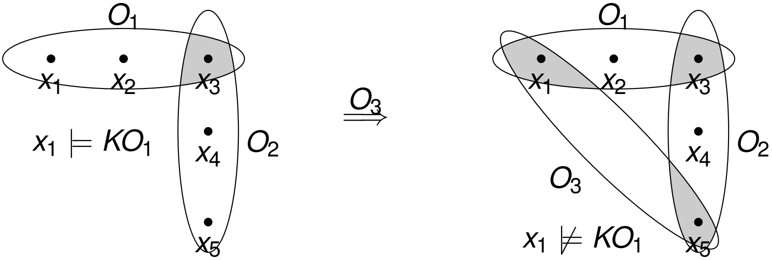
Our knowledge is *defeasible*



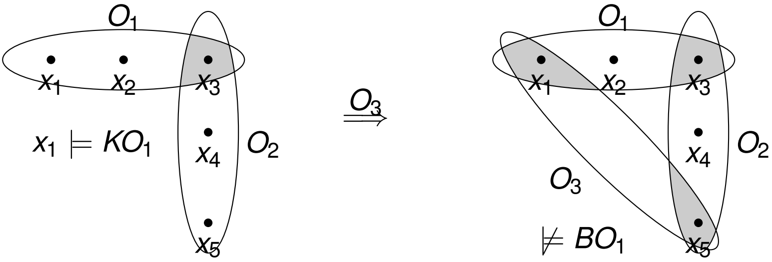
Our knowledge is *defeasible*



Our knowledge is *defeasible*



Our knowledge is *defeasible*



BUT O_3 is a *misleading defeater*: it produces NEW FALSE (combined) evidence $O_3 \cap O_2$.

Non-misleading defeaters

K is defeasible for factive evidence, but *in-defeasible* for "non-misleading" evidence.

Given a topo-e-model $\mathcal{M} = (X, E_0, \tau, V)$ and $x \in X$,

$Q \subseteq X$ is *misleading* iff its addition to E_0 produces some false new evidence:

$Q \subseteq X$ is *misleading* iff $x \notin Q \cap e \notin E \cup \{\emptyset\}$ for some $e \in E$.

Topologically, a misleading evidence adds an open set to the evidential topology that does not include the actual state.

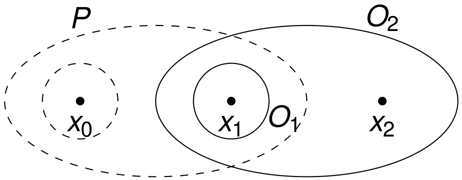
Weak Stability?

Our knowledge is “weakly stable":

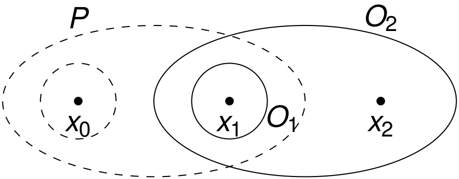
1. *P* is true
2. she believes that *P* is true
3. her belief in *P* cannot be defeated by new *non-misleading* evidence. *weakly stable true belief*

But weakly stable true belief is NOT enough for knowledge!

Counterexample

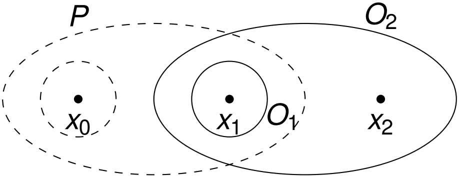


Counterexample



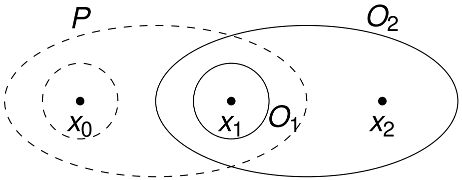
- *BP* holds, since $ClIntP = Cl\{x_1\} = X$

Counterexample



- BP holds, since $ClIntP = Cl\{x_1\} = X$
- BP is stable (under addition of any non-misleading information)

Counterexample



- BP holds, since $ClIntP = Cl\{x_1\} = X$
- BP is stable (under addition of any non-misleading information)
- BUT $x_0 \not\models KP$, since $x_0 \notin IntP = \{x_1\}$

Our knowledge is *weakly in-defeasible*

An agent knows P (in our sense of K) iff:

1. P is true
2. she believes that P is true
3. her belief in P cannot be defeated by new *non-misleading* evidence. *weak stable belief*
4. (the belief in) <u>its justification</u> cannot be defeated by new *non-misleading* evidence. *weak stable jutification*

$x \models KP$ iff $\exists U \in \tau \setminus \{\emptyset\}$ s.t. $U \subseteq P$ and $U \cap Q \neq \emptyset$ for all non-misleading Q

Technical Results

$$\mathcal{L} := p \mid \neg\varphi \mid \varphi \wedge \varphi \mid \forall\varphi \mid E_0\varphi \mid E\varphi \mid \Box_0\varphi \mid \Box\varphi \mid B\varphi \mid K\varphi$$

The following equivalences are valid in all topo-e-models:

$$E_0\varphi \leftrightarrow \exists\Box_0\varphi \qquad B\varphi \leftrightarrow \forall\Diamond\Box\varphi$$
$$E\varphi \leftrightarrow \exists\Box\varphi \qquad K\varphi \leftrightarrow \Box\varphi \wedge B\varphi$$

where

$$\exists\varphi := \neg\forall\neg\varphi.$$

So *our largest language above is actually co-expressive with a smaller one*:

$$\mathcal{L} := p \mid \neg\varphi \mid \varphi \wedge \varphi \mid \forall\varphi \mid \Box_0\varphi \mid \Box\varphi$$

Axiomatization

the S5 axioms and rules for $\forall$

the S4 axioms and rules for $\Box$

$\Box_0\varphi \to \Box_0\Box_0\varphi$

$\forall\varphi \to \Box_0\varphi$

$\Box_0\varphi \to \Box\varphi$

$(\Box_0\varphi \wedge \forall\psi) \to \Box_0(\varphi \wedge \forall\psi)$

from $\varphi \to \psi$, infer $\Box_0\varphi \to \Box_0\psi$

Theorem

The logic of evidence has the finite model property, is decidable, and is completely axiomatized by the above system.

Fragments

The system $KD45$ is complete for the B fragment.

The system $S4.2$ is complete for the K fragment.

The KB fragment is completely axiomatized by Stalnaker's axioms for doxastic-epistemic logic:

1. the $S4$ axioms and rules for Knowledge K
2. Consistency of Belief: $B\phi \to \neg B \neg \phi$;
3. Knowledge implies Belief: $K\phi \to B\phi$;
4. Strong Positive and Negative Introspection for Belief: $B\phi \to KB\phi$; $\neg B\phi \to K\neg B\phi$;
5. the "Strong Belief" axiom: $B\phi \to BK\phi$.

The case E_0 finite

Let us call an evidence model "**feasible**" if the family E_0 of available evidence is **finite** (even if there are infinitely many possible worlds in X, and even if all or some of the evidence pieces $e \in E_0$ comprise infinitely many worlds).

"Real" agents are bounded: they can only gather finitely many (independent) pieces of evidence at any given moment.

As we saw, **for feasible evidence models, our notion of belief coincides with the van-Benthem-Pacuit belief**.

Moreover, **all the above proof systems are sound and complete (for the respective languages) wrt the van-Benthem-Pacuit semantics restricted to feasible evidence models.**

Summary of Our work

- uses the same models with a particular focus on the *evidential topology*
- different notions of evidence: basic, combined, factive, misleading...
- topological formalizations of *argument* and *justification*
- topologically interpreted, evidence-based *consistent* notions of (justified) belief and (defeasible) knowledge
- *complete axiomatizations, finite model property*
- adapts the van Benthem-Pacuit dynamics of evidence management to our modified setting
  (-but this part was not included in this presentation! see full paper for details).

Thank you!