



BEYOND “KNOWING THAT” (II)

LOGICS OF KNOWING WHETHER AND MORE

Yanjing Wang

Department of Philosophy, Peking University

NASSLLI 2018, CMU

www.wangyanjing.com

Background

Understanding the expressivity

Axiomatizations

BACKGROUND

All happy families are alike; each unhappy family is unhappy in its own way.

Leo Tolstoy, *Anna Karenina*

Whether the second half is possible depends on how many kinds of unhappiness there are (w.r.t. the number of families)...

(Un)happiness may depend on knowledge...

A loves B (p) but B doesn't love A (q);

p , $K_A q$;

p , $K_A q$, and $K_A K_B p$;

...

Sometimes, it is better not to know...

All omniscient agent are alike; each ignorant agent is ignorant in its own way.

How many different ignorant states are there?

Given a single fact p in the 2-agent setting, how many different epistemic states (maximal consistent sets in $\mathbb{S}5\text{-EL}$) are there?

If the language is countable, there are at most 2^{\aleph_0} many consistent sets, but can we really realize 2^{\aleph_0} many mutually inconsistent consistent sets? Is it trivial?

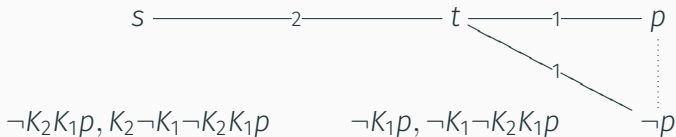
- $(p, K_1p), (p, \neg K_1p)$
- from the first: $(p, K_1p, K_2K_1p), (p, K_1p, \neg K_2K_1p)$
- from the second: $(p, \neg K_1p, K_2\neg K_1p), (p, \neg K_1p, \neg K_2\neg K_1p)$
- It seems we can continue like this arbitrarily...
- $(p, K_1p, \neg K_2K_1p, \neg K_1\neg K_2K_1p, K_2\neg K_1\neg K_2K_1p)$: **inconsistent!**
- In particular, $K_2\neg K_1\neg K_2K_1p$ and $\neg K_2K_1p$ are inconsistent.

Epistemic logic was invented to check consistency of knowledge and ignorance (Hintikka).

To show $\{K_2\neg K_1\neg K_2K_1p, \neg K_2K_1p\}$ is inconsistent:

- (1) $\neg K_1p \rightarrow \neg K_2K_1p$ (T)
- (2) $K_1\neg K_1p \rightarrow K_1\neg K_2K_1p$ (NEC, DIST)
- (3) $\neg K_1p \rightarrow K_1\neg K_2K_1p$ (5, MP)
- (4) $\neg K_1\neg K_2K_1p \rightarrow K_1p$
- (5) $K_2\neg K_1\neg K_2K_1p \rightarrow K_2K_1p$ (NEC, DIST)

We can show it semantically (recall completeness theorem: a consistent set of formulas must be satisfiable!)



THE FIRST ANSWER WAS PROVIDED BY AUMANN (1989)

Aumann showed that there are 2^{\aleph_0} many mutually inconsistent consistent sets. Hart, Heifetz, and Samet (1996) presented a much simpler construction using “knowing whether”: Define $Kw_i\phi = K_i\phi \vee K_i\neg\phi$. For any $s \in \{0, 1\}^+$ construct ϕ_s :

- $\phi_0 = \neg p$; $\phi_1 = p$

- if $|s| > 0$ is odd: $\phi_{sx} = \begin{cases} \neg Kw_1\phi_s & x = 0 \\ Kw_1\phi_s & x = 1 \end{cases}$

- if $|s| > 0$ is even: $\phi_{sx} = \begin{cases} \neg Kw_2\phi_s & x = 0 \\ Kw_2\phi_s & x = 1 \end{cases}$

E.g., $\phi_{0101} = Kw_1\neg Kw_2Kw_1\neg p$., $\phi_{1010} = \neg Kw_1Kw_2\neg Kw_1p$.

We can show

$\Phi_w = \{\phi_s \mid s \text{ is an non-empty initial segment of } w\}$ are all consistent for each $w \in \{0, 1\}^\omega$.

Let w_k be the k th position of w , e.g., $(01\dots)_1 = 0$, $(0010\dots)_3 = 1$.

We can build a model: $\mathcal{M} = \langle \Phi_w \mid w \in \{0, 1\}^\omega, \sim_i, V \rangle$ where:

- $\Phi_w \sim_1 \Phi_v$ iff $w_k = v_k$ for all the odd k and if $w_k = v_k = 1$ for some odd k then $w_{k-1} = v_{k-1}$.
- $\Phi_w \sim_2 \Phi_v$ iff $w_k = v_k$ for all the even $k > 0$ and if $w_k = v_k = 1$ for some even $k > 1$ then $w_{k-1} = v_{k-1}$.
- $V(p) = \{\Phi_w \mid w_0 = 1\}$.

We can show that $\Phi_w \models \phi_s$ if s is an initial segment of w by induction on the length of s .

Key observation: $\models Kw_i\phi \leftrightarrow Kw_i\neg\phi$: if $w = 0001\dots$ you want to show $\neg Kw_2\neg Kw_1\neg p$ holds on all the worlds $\Phi_v \sim_1 \Phi_w$ (equiv. $\neg Kw_2Kw_1p, \neg Kw_2Kw_1\neg p$).

Fragments of **EL** [Parikh and Krasucki 92]; more generally see Klein and Pacuit (manuscript, presented at LOFT).

“KNOWING WHETHER” IS USEFUL

Natural and succinct to express:

- knowledge expression with hidden content: A “I know whether p but I won’t tell you.”
- knowledge with ignorance: B “I don’t know whether p but I know you know whether p .”

To be used (e.g., in muddy children):

- As assumptions: children know whether others are dirty.
- As precondition of actions: e.g., “step forward if you know whether you are dirty”.
- As goals of planning: e.g., eventually the children know whether they are dirty.

Using Kw can give us (exponential) succinctness.

Essentially modal logic is about necessity \Box (and possibility \Diamond) in various contexts:

- epistemic: know that
- doxastic: believe that
- deontic: ought to be that
- proof theoretical: it is provable that

$$\Delta\phi := \Box\phi \vee \Box\neg\phi \quad \nabla\phi := \neg\Delta\phi = \neg\Delta\neg\phi = \neg\Box\phi \wedge \neg\Box\neg\phi$$

Modal logic with Δ and ∇ as the primitive modality is about contingency and non-contingency in various contexts:

- alethic ∇ : contingency [Montgomery & Routley 66 and many more, see our RSL article for a survey]
- epistemic ∇ : ignorance [van der Hoek & Lomuscio 03]
- doxastic ∇ : not opinionated about
- deontic ∇ : moral indifference [von Wright 51]
- proof theoretical ∇ : undecided [Zolin 2001]

UNDERSTANDING THE EXPRESSIVITY

NON-CONTINGENCY (KNOWING WHETHER) OPERATOR

NCL is defined as follows:

$$\phi ::= \top \mid p \mid \neg\phi \mid (\phi \wedge \phi) \mid \Delta_i\phi$$

where $p \in \mathbf{P}$ and $i \in \mathbf{I}$. A Kripke model \mathcal{M} is a triple

$$\langle S, \{\rightarrow_i \mid i \in \mathbf{I}\}, V \rangle$$

where S is a non-empty set, $\rightarrow_i \subseteq S \times S$ and $V : \mathbf{P} \rightarrow 2^S$.

$\mathcal{M}, s \models \Delta_i\phi \Leftrightarrow$	for all t_1, t_2 such that $s \rightarrow_i t_1, s \rightarrow_i t_2$:
	$(\mathcal{M}, t_1 \models \phi \Leftrightarrow \mathcal{M}, t_2 \models \phi)$
\Leftrightarrow	either for all t such that $s \rightarrow_i t : \mathcal{M}, t \models \phi$
	or for all t such that $s \rightarrow_i t : \mathcal{M}, t \not\models \phi$

NCL is clearly no more expressive than **ML** since we can define a translation $t : \mathbf{NCL} \rightarrow \mathbf{ML}$ such that:

$$t(\Delta_i \phi) = \Box_i t(\phi) \vee \Box_i \neg t(\phi)$$

What about the other way around? It is also easy if we restrict ourselves to reflexive models: we can define a translation $t' : \mathbf{ML} \rightarrow \mathbf{NCL}$ such that:

$$t'(\Box_i \phi) = t'(\phi) \wedge \Delta_i t'(\phi)$$

Are they equally expressive over arbitrary models? If not, how to characterize the expressive power of **NCL** within **ML**?

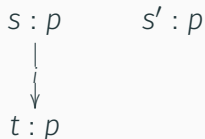
Definition (Bisimulation)

Let $\mathcal{M} = \langle S, \{\rightarrow_i \mid i \in I\}, V \rangle$, $\mathcal{N} = \langle S', \{\rightarrow'_i \mid i \in I\}, V' \rangle$ be two models. A binary relation Z over $S \times S'$ is a *bisimulation* between \mathcal{M} and \mathcal{N} , if Z is non-empty and whenever sZs' :

- (Invariance) s and s' satisfy the same propositional variables;
- (Zig) if $s \rightarrow_i t$, then there is a t' such that $s' \rightarrow'_i t'$ and tZt' ;
- (Zag) if $s' \rightarrow'_i t'$, then there is a t such that $s \rightarrow_i t$ and tZt' .

\mathcal{M}, s is *bisimilar* to \mathcal{N}, t ($\mathcal{M}, s \Leftrightarrow \mathcal{N}, t$) if there is a bisimulation between \mathcal{M} and \mathcal{N} linking s with t .

Bisimilarity is clearly an invariance relation for **NCL** but it is too strong:



These two finite models satisfy the same **NCL** formulas but they are not bisimilar.

But in most of the cases when there are two and more successors bisimulation seems fine.

A CRUCIAL OBSERVATION

To tell the subtle difference we need to connect Δ and \Box :

\Box_j is *almost definable* by Δ_j (call the schema AD):

Proposition

For any ϕ, ψ , $\models \neg\Delta_j\psi \rightarrow (\Box_j\phi \leftrightarrow (\Delta_j\phi \wedge \Delta_j(\psi \rightarrow \phi)))$.

This inspires us to:

- come up with a structural equivalence notion of Δ_j -*bisimulation* and characterize the expressive power;
- come up with the right definition of canonical relations in the latter completeness proofs;
- find the right axioms for special frame properties.

If there are two states which can be told apart by **NCL** formulas then the bisimulation conditions should work. However, to make it into a purely *structural* is quite non-trivial. Idea: define it within a *single* model.

Definition (Δ -Bisimulation)

Let $\mathcal{M} = \langle S, R, V \rangle$ be a model. A binary relation Z over S is a Δ -bisimulation on \mathcal{M} , if Z is non-empty and whenever sZs' :

- (Invariance) s and s' satisfy the same basic p ;
- (Zig) if there are two different successors t_1, t_2 of s such that $(t_1, t_2) \notin Z$ and $s \rightarrow_i t$, then $\exists t'$ s.t. $s' \rightarrow_i t'$ and tZt' ;
- (Zag) if there are two different successors t'_1, t'_2 of s' such that $(t'_1, t'_2) \notin Z$ and $s' \rightarrow_i t'$, then $\exists t$ s.t. $s \rightarrow_i t$ and tZt' .

Δ -BISIMILARITY

\mathcal{M}, s and \mathcal{N}, t are Δ -bisimilar ($\mathcal{M}, s \Leftrightarrow_{\Delta} \mathcal{N}, t$) if there is a Δ -bisimulation on the *disjoint union* of \mathcal{M} and \mathcal{N} linking s and t .

Don't forget to show that Δ -bisimilarity is indeed an *equivalence relation*! Transitivity is hard.

Theorem (Fan, Wang, van Ditmarsch AiML14)

For image-finite (or NCL saturated models) \mathcal{M}, s and \mathcal{N}, t :
 $\mathcal{M}, s \Leftrightarrow_{\Delta} \mathcal{N}, t \iff \mathcal{M}, s \equiv_{\text{NCL}} \mathcal{N}, t$ (satisfying the same NCL formulas).

Proof ideas: (\implies) suppose $\mathcal{M}, s \not\equiv_{\text{NCL}} \mathcal{N}, t$ show that $\mathcal{M}, s \not\equiv_{\Delta} \mathcal{N}, t$ by Zig and IH. (\impliedby) take \equiv_{NCL} as Z and use the AD schema to express $\diamond \bigwedge \Gamma$ for Zig.

Theorem (Fan, Wang, van Ditmarsch AiML14)

NCL is the Δ -bisimilarity invariant fragment of ML (and FOL).

Proof ideas:

- Since **NCL** are **ML** are compact, we can show **NCL** is the \equiv_{NCL} -invariant fragment of **ML**.
- Replace \equiv_{NCL} by \leftrightarrow_{Δ} : show that each \leftrightarrow_{Δ} invariant formula is also \equiv_{NCL} invariant, by using ultrafilter extension and saturation based on the Hennessy-Milner-like theorem.

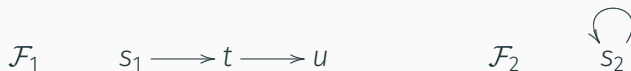
For characterizarion w.r.t. **FOL** we need to use van Benthem theorem and the fact that \leftrightarrow_{Δ} is coarser than \leftrightarrow .

Is the difference in expressivity just a negligible subtlety?

Theorem

*The frame properties of seriality, reflexivity, transitivity, symmetry, and Euclidicity are **not** definable in **NCL**.*

Proof idea:



We can show $\mathcal{F}_1 \models \phi \iff \mathcal{F}_2 \models \phi$ by its contrapositive and \leftrightarrow_{Δ} .
Then towards contradiction...

AXIOMATIZATIONS

Apparent difficulties:

- **NCL** formulas cannot capture the frame properties.
- **NCL** is not normal: $\Delta_i(\phi \rightarrow \psi) \wedge \Delta_i\phi \rightarrow \Delta_i\psi$ is not valid.
- With extra axioms like $\Delta_i\phi \leftrightarrow \Delta_i\neg\phi$.

Consider the following axiom schemas and rules as system **SNCL**:

TAUT	all instances of tautologies
KwCon	$\Delta_i(\phi) \wedge \Delta_i(\psi) \rightarrow \Delta_i(\phi \wedge \psi)$
KwDis	$\Delta_i\phi \rightarrow \Delta_i(\phi \rightarrow \psi) \vee \Delta_i(\neg\phi \rightarrow \chi)$
KwNeg	$\Delta_i\phi \leftrightarrow \Delta_i\neg\phi$
KwTop	$\Delta_i\top$
MP	From ϕ and $\phi \rightarrow \psi$ infer ψ
REKw	From $\phi \leftrightarrow \psi$ infer $\Delta_i\phi \leftrightarrow \Delta_i\psi$

Theorem

SNCL is sound and strongly complete w.r.t. NCL over the class of arbitrary frames (and serial frames).

The proof is based on the following canonical model construction, inspired by the “almost definability” schema AD:

$$\neg\Delta_i\psi \rightarrow (\Box_i\phi \leftrightarrow (\Delta_i\phi \wedge \Delta_i(\psi \rightarrow \phi)))$$

Definition (Canonical model)

Define $\mathcal{M}^c = \langle S^c, R^c, V^c \rangle$ as follows:

- $S^c = \{s \mid s \text{ is a maximal consistent set of } \mathcal{SNCL}\}$
- For all $s, t \in S^c$, $sR_i^c t$ iff **there exists** χ such that:
 - $\neg\Delta_i\chi \in s$, and
 - for all ϕ , $\Delta_i\phi \wedge \Delta_i(\chi \rightarrow \phi) \in s$ implies $\phi \in t$.
- $V^c(p) = \{s \in S^c \mid p \in s\}$.

See the similarity with the standard canonical definition:

For all $s, t \in S^c$, $sR_i^c t$ iff for all ϕ : $\Box_i\phi \in s$ implies $\phi \in t$.

Lemma

For all $\phi \in \mathbf{NCL}$: $\mathcal{M}^c, s \vDash \phi \iff \phi \in s$.

Proof idea (for the case of $\Delta_i\psi$): \implies : suppose $\Delta_i\psi \notin s$ then $\neg\Delta_i\psi \in s$. To show $\mathcal{M}^c, s \not\vDash \Delta_i\phi$ we need to construct **two** i -successors of s that disagree about ψ . Having Lindenbaum lemma in mind, we need to show:

1. $\{\phi \mid \Delta_i\phi \wedge \Delta_i(\psi \rightarrow \phi) \in s\} \cup \{\psi\}$ is consistent.
2. $\{\phi \mid \Delta_i\phi \wedge \Delta_i(\neg\psi \rightarrow \phi) \in s\} \cup \{\neg\psi\}$ is consistent.

1 relies on the validity of the following long formula (proved by using **KwCon** and **KwDis**.) and 2 needs **KwNeg** in addition.

$$\Delta_i\left(\bigwedge_{j=1}^k \phi_j \rightarrow \neg\psi\right) \wedge \bigwedge_{j=1}^k \Delta_i\phi_j \wedge \bigwedge_{j=1}^k \Delta_i(\psi \rightarrow \phi_j) \rightarrow \Delta_i\psi$$

NCL OVER OTHER FRAME CLASSES

Notation	Axiom Schemas	Systems	Frames
KwT	$\Delta_i\phi \wedge \Delta_i(\phi \rightarrow \psi) \wedge \phi \rightarrow \Delta_i\psi$	SNCLT = SNCL + KwT	<i>reflexive</i>
Kw4	$\Delta_i\phi \rightarrow \Delta_i(\Delta_i\phi \vee \psi)$	SNCL4 = SNCL + Kw4	<i>transitive</i>
Kw5	$\neg\Delta_i\phi \rightarrow \Delta_i(\neg\Delta_i\phi \vee \psi)$	SNCL5 = SNCL + Kw5	<i>euclidean</i>
wKw4	$\Delta_i\phi \rightarrow \Delta_i\Delta_i\phi$	SNCLS4 = SNCL + KwT + wKw4	<i>ref.&trans.</i>
wKw5	$\neg\Delta_i\phi \rightarrow \Delta_i\neg\Delta_i\phi$	SNCLS5 = SNCL + KwT + wKw5	<i>equivalence</i>
KwB	$\phi \rightarrow \Delta_i((\Delta_i\phi \wedge \Delta_i(\phi \rightarrow \psi) \wedge \neg\Delta_i\psi) \rightarrow \chi)$	SNCLB = SNCL + KwB	<i>symmetric</i>

Note: SNCL + wKw4 and SNCL + wKw5 are **not** complete over the classes of transitive and euclidean frames respectively.

- We find axioms inspired by AD.
- We manipulate the canonical model.

EXAMPLE: OVER REFLEXIVE FRAMES

How do we get the axioms? The naive translation of axiom T does not work: $\Delta_i\phi \wedge \phi \rightarrow \phi$ is simply a tautology. Instead, we use the AD schema to translate \Box_i . We start with (a version) of the T axiom $\Box_i\neg\phi \rightarrow \neg\phi$ and add a precondition $\neg\Delta_i\psi$:

$$\neg\Delta_i\psi \rightarrow (\Box_i\neg\phi \rightarrow \neg\phi) \quad (1)$$

$$\Leftrightarrow \neg\Delta_i\neg\psi \wedge \Box_i\neg\phi \rightarrow \neg\phi \quad (2)$$

$$\Leftrightarrow \neg\Delta_i\neg\psi \wedge \Delta_i\neg\phi \wedge \Delta_i(\neg\psi \rightarrow \neg\phi) \rightarrow \neg\phi \quad (3)$$

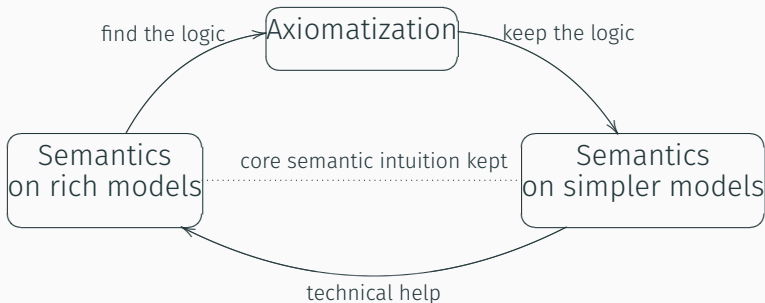
$$\Leftrightarrow \Delta_i\phi \wedge \Delta_i(\phi \rightarrow \psi) \wedge \phi \rightarrow \Delta_i\psi \quad (4)$$

KwT is valid on all the reflexive frames. The Canonical model is not reflexive but we can add the reflexive arrows safely (preserving true lemma). The symmetry case is much more complicated.

SIMPLIFY THE SEMANTICS WHILE KEEPING THE LOGIC

Common difficulties: weak language vs. rich semantics

To restore the balance between the language and model:



RECALL THE AXIOMATIZATION

- TAUT all instances of tautologies
- KwCon $\Delta_i(\phi) \wedge \Delta_i(\psi) \rightarrow \Delta_i(\phi \wedge \psi)$
- KwDis $\Delta_i\phi \rightarrow \Delta_i(\phi \rightarrow \psi) \vee \Delta_i(\neg\phi \rightarrow \chi)$
- KwNeg $\Delta_i\phi \leftrightarrow \Delta_i\neg\phi$
- KwTop $\Delta_i\top$
- MP From ϕ and $\phi \rightarrow \psi$ infer ψ
- REKw From $\phi \leftrightarrow \psi$ infer $\Delta_i\phi \leftrightarrow \Delta_i\psi$

A neighborhood model is $\mathcal{M} = \langle W, N, V \rangle$ where $N : W \rightarrow 2^{2^W}$ assigns each world a set of subsets of W .

The semantics is given by:

$$\mathcal{M}, w \Vdash \Box\phi \iff \llbracket \phi \rrbracket^{\mathcal{M}} \in N(w)$$

We say the neighbourhood model is a **NCL**-model if for all $w \in W$

- $W \in N(w)$
- $N(w)$ is closed under complementation
- $N(w)$ is closed under intersection
- $N(w)$ is closed under supersets or co-supersets:
 $X, Y, Z \subseteq W, X \in N(w)$ implies $X \cup Y \in N(w)$ or
 $(W \setminus X) \cup Z \in N(w)$.

We have exactly the same valid formulas over **NCL**-models.

We can add public announcements and event updates into the language **PALNC**:

$$\phi ::= \top \mid p \mid \neg\phi \mid (\phi \wedge \phi) \mid \Delta_i\phi \mid [\phi]\phi$$

With the usual reduction axiom and the following one we can easily axiomatize **PALNC** over various classes of frames:

$$[\phi]\Delta_i\psi \leftrightarrow (\phi \rightarrow (\Delta_i[\phi]\psi \vee \Delta_i[\phi]\neg\psi))$$

Similar story holds if we introduce the event modality.

$$\Delta\phi := \Box\phi \vee \Box\neg\phi$$

We can define $\odot\phi$ as:

- false belief: $\neg\phi \wedge \Box\phi$
- essence: $\phi \rightarrow \Box\phi$, accident: $\phi \wedge \neg\Box\phi$ (Pan & Yang, Fan)
- strong non-contingency: $(\phi \rightarrow \Box\phi) \wedge (\neg\phi \rightarrow \Box\neg\phi)$ (Fan)
- strong belief disagreement: $\Box_1\phi \wedge \Box_2\neg\phi$ (Chen & Pan SL18)
- knowing whether ϕ or ψ : $\Box\phi \vee \Box\psi$ (Aloni, Égré, Jäger 09)
- propositional dependency/subservience: knowing whether ψ implies knowing whether ϕ (Goranko & Kuusisto RSL18, Fan)
- in the neighbourhood setting (Fan, van Ditmarsch ICLA15)

What is the general theory of such propositional bundles?

- distributed knowing whether (Dw) [Fan, Su, LORI17]
- commonly knowing whether (Cw) [Su thesis 2018]

1. $C\phi \vee C\neg\phi$
2. $CEw\phi$ where $Ew\phi$ says everyone knows whether ϕ
3. $\bigwedge_{k \in \omega} Ew^k\phi$
4. $\bigwedge_{i \in G} CKw_i\phi \vee C\neg Kw_i\phi$
5. $\bigwedge_{s \in G^*} Kw_s\phi$ where Kw_s is the abbreviation of $Kw_{i_1} \dots Kw_{i_n}$

Implication: $1 \rightarrow 2 \rightarrow 3, 4, 5$, but 3, 4, 5 are independent, over arbitrary models.

Cw^5 is not expressible by $C + K$. It is a highly “strange” operator.

There is a large body of research on non-contingency since 1960s involving authors such as Montgomery, Routley, Humberstone, Segerberg, Creswell, Kuhn, Steinsvold and so on.

There are also works on ignorance logic (knowing whether) in epistemic logic.

Surprisingly, the two communities were ignorant about each other's work on such logics! See our RSL article.