Epistemic Logics for Introspection
Part I

Paul Égré

Institut Jean-Nicod, CNRS

FEW 09, Pittsburgh CMU- June 19, 2009

Aim of this tutorial

- Introduction to epistemic logic

Aim of this tutorial

- Introduction to epistemic logic
- Iterations of knowledge and introspection

Aim of this tutorial

- Introduction to epistemic logic
- Iterations of knowledge and introspection
- Focus on structures of imprecise/inexact knowledge

Aim of this tutorial

- Introduction to epistemic logic
- Iterations of knowledge and introspection
- Focus on structures of imprecise/inexact knowledge

Central problem: compatibility of imprecision and introspection (Williamson).

Some useful references

Textbooks:

- Fagin, Halpern, Moses, Vardi 1995. Reasoning about Knowledge, MIT Press.
- Blackburn, de Rijke, Venema 2001. Modal Logic. Cambridge Tracts in Theoretical Computer Science.
- van Ditmarsch, van der Hoek, Kooi 2007. Dynamic Epistemic Logic, Springer Synthese Library 237

On inexact knowledge

T. Williamson:

- T. Williamson 1992. Inexact Knowledge, *Mind*.
- T. Williamson 1994. Appendix to *Vagueness*, Routledge.
- T. Williamson 2000. *Knowledge and its Limits*, OUP.

On inexact knowledge

T. Williamson:

- T. Williamson 1992. Inexact Knowledge, *Mind*.
- T. Williamson 1994. Appendix to *Vagueness*, Routledge.
- T. Williamson 2000. *Knowledge and its Limits*, OUP.

Replies:

- Halpern 2004. Intransitivity and Vagueness, *KR 2004*.
- Dokic & Egré 2008. Margin for Error and the Transparency of Knowledge, *Synthese*.
- Bonnay & Egré 2009. Inexact Knowledge with Introspection, *Journal of Philosophical Logic*.

Outline for Day 1

- Background on Epistemic Logic
- Inexact knowledge
- Centered Semantics
- Comparison with explicit 2d-semantics

Outline for Day 2

- Token semantics
- Extensions: dynamic / common knowledge

The language of modal epistemic logic

$$\phi := p \mid \neg\phi \mid \phi \wedge \phi \mid \Box\phi \mid$$

The language of modal epistemic logic

$$\phi := p \mid \neg\phi \mid \phi \wedge \phi \mid \Box\phi \mid$$

- $\Box\phi = K\phi$: I know that ϕ

The language of modal epistemic logic

$$\phi := p \mid \neg\phi \mid \phi \wedge \phi \mid \Box\phi \mid$$

- $\Box\phi = K\phi$: I know that ϕ
- Focus on a single agent

The language of modal epistemic logic

$$\phi := p \mid \neg\phi \mid \phi \wedge \phi \mid \Box\phi \mid$$

- $\Box\phi = K\phi$: I know that ϕ
- Focus on a single agent
- Equally we could talk of belief instead of knowledge

Semantics

1. $M = \langle W, R, V \rangle$

   W = epistemic states
   R = epistemic uncertainty
   V = distribution of information

Semantics

1. $M = \langle W, R, V \rangle$

   W = epistemic states
   R = epistemic uncertainty
   V = distribution of information

2. $M, w \models \Box\phi$ iff for every $w' : wRw'$, $M, w' \models \phi$.

   "I know ϕ iff ϕ holds at every epistemic alternative".

More precisely

More precisely

1. $M, w \models p$ iff $w \in V(p)$

More precisely

1. $M, w \models p$ iff $w \in V(p)$
2. $M, w \models \neg\phi$ iff $M, w \not\models \phi$

More precisely

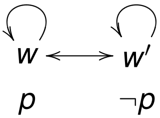
1. $M, w \models p$ iff $w \in V(p)$
2. $M, w \models \neg\phi$ iff $M, w \not\models \phi$
3. $M, w \models (\phi \wedge \psi)$ iff $M, w \models \phi$ and $M, w \models \psi$

More precisely

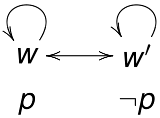
1. $M, w \models p$ iff $w \in V(p)$
2. $M, w \models \neg\phi$ iff $M, w \not\models \phi$
3. $M, w \models (\phi \wedge \psi)$ iff $M, w \models \phi$ and $M, w \models \psi$
4. $M, w \models \Box\phi$ iff $R(w) \subseteq [\phi]$

As usual: $\Diamond\phi := \neg\Box\neg\phi$: for all I know, ϕ is possible / I cannot exclude that ϕ

A very simple example

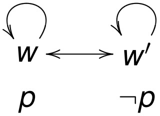


A very simple example



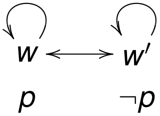
$w \models \neg\Box p$

A very simple example



$w \models \neg\Box p$
$w \models \neg\Box\neg p$

A very simple example

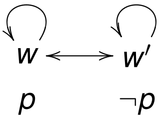


$w \models \neg\Box p$

$w \models \neg\Box\neg p$

$w \models \Box\neg\Box p$

A very simple example



$w \models \neg\Box p$
$w \models \neg\Box\neg p$
$w \models \Box\neg\Box p$
$w \models \Box\Box\neg\Box p$

Definitions

Model-validity vs Validity

- $M \models \phi$: for all $w \in M$, $M, w \models \phi$
- $\models \phi$: for all M and all $w \in M$: $M, w \models \phi$

Definitions

Model-validity vs Validity

- $M \models \phi$: for all $w \in M$, $M, w \models \phi$
- $\models \phi$: for all M and all $w \in M$: $M, w \models \phi$

Frame-validity:

- $\models_{ref} \phi$ iff ϕ is valid in all models whose accessibility relation is reflexive
- $\models_{tr} \phi$ iff ϕ is valid in all models whose accessibility relation is transitive
- $\models_{eucl} \phi$ iff ϕ is valid in all models whose accessibility relation is euclidian

Frame properties

Reflexivity	xRx
Transitivity	$xRy \wedge yRz \rightarrow xRz$
Euclideanness	$xRy \wedge xRz \rightarrow yRz$
Symmetry	$xRy \rightarrow yRx$

S5 models

T	$\Box p \rightarrow p$	factivity	reflexivity
4	$\Box p \rightarrow \Box\Box p$	positive introspection	transitivity
5	$\neg\Box p \rightarrow \Box\neg\Box p$	negative introspection	euclidianity
B	$p \rightarrow \Box\neg\Box\neg p$	"Brouwersche"	symmetry

Exact knowledge

- KT45 = KT5 = KTB4 = S5
- "S5 models" : R is an equivalence relation
- Equivalence relations determine partitional models of information: for every w, $R(w)$ is a cell of the partition induced by R when R is S5.

Inexact knowledge

Inexact knowledge

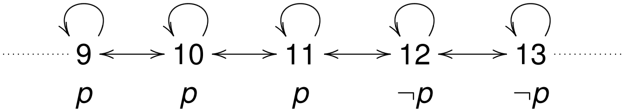
- Partitional models of information are models of exact knowledge

Inexact knowledge

- Partitional models of information are models of exact knowledge
- Situations of inexact knowledge: overlap of the information sets, or failure of transitivity / euclideanness of R

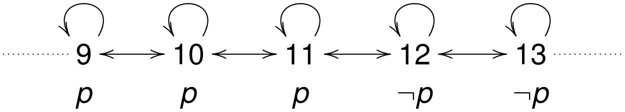
Inexact knowledge

- Partitional models of information are models of exact knowledge
- Situations of inexact knowledge: overlap of the information sets, or failure of transitivity / euclideanness of R
- Example: I don't discriminate between objects whose size differs by less than 1 cm.



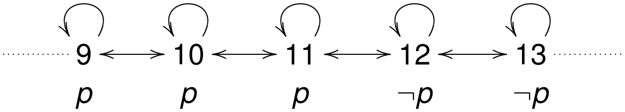
Inexact knowledge

- Partitional models of information are models of exact knowledge
- Situations of inexact knowledge: overlap of the information sets, or failure of transitivity / euclideanness of R
- Example: I don't discriminate between objects whose size differs by less than 1 cm.



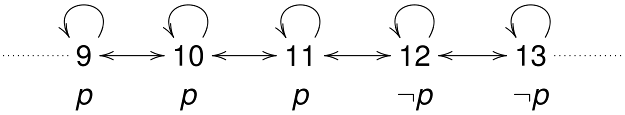
- R : epistemic uncertainty as perceptual indiscriminability

First-order knowledge



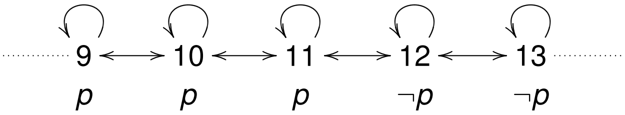
- $10 \models \Box p$
- $11, 12 \models \neg\Box p \wedge \neg\Box\neg p$
- $13 \models \Box\neg p$

Higher-order knowledge



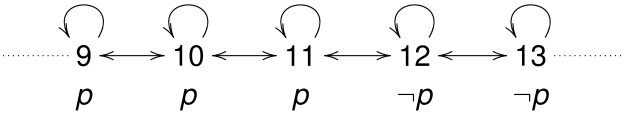
- $10 \models \neg\Box\Box p$

Higher-order knowledge



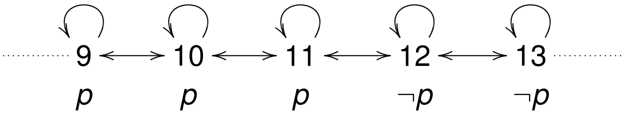
- $10 \models \neg\Box\Box p$
- $9 \models \Box\Box p \wedge \neg\Box\Box\Box p$

Higher-order knowledge



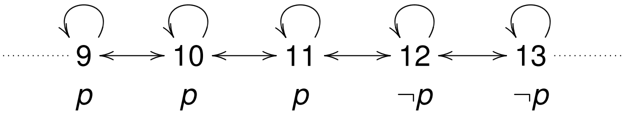
- $10 \models \neg\Box\Box p$
- $9 \models \Box\Box p \wedge \neg\Box\Box\Box p$
- $8 \models \Box\Box\Box p \wedge \neg\Box\Box\Box\Box p$

Higher-order knowledge



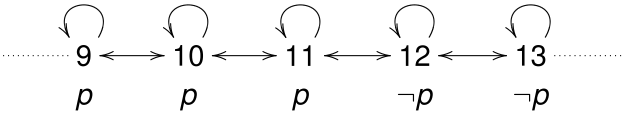
- $10 \models \neg\Box\Box p$
- $9 \models \Box\Box p \wedge \neg\Box\Box\Box p$
- $8 \models \Box\Box\Box p \wedge \neg\Box\Box\Box\Box p$
- ...

Higher-order knowledge



- $10 \models \neg\Box\Box p$
- $9 \models \Box\Box p \wedge \neg\Box\Box\Box p$
- $8 \models \Box\Box\Box p \wedge \neg\Box\Box\Box\Box p$
- ...
- $0 \models \Box^{10} p \wedge \neg\Box\Box^{10} p$

Higher-order knowledge



- $10 \models \neg\Box\Box p$
- $9 \models \Box\Box p \wedge \neg\Box\Box\Box p$
- $8 \models \Box\Box\Box p \wedge \neg\Box\Box\Box\Box p$
- ...
- $0 \models \Box^{10} p \wedge \neg\Box\Box^{10} p$

Williamson 1992: “iteration of knowledge operators is a process of gradual erosion”

Margin for error semantics

Williamson 1992, 1994, “Logic of Clarity”

Margin for error semantics

Williamson 1992, 1994, "Logic of Clarity"

- Margin models: $M = \langle W, d, \alpha, V \rangle$

  d = metric over W
  $\alpha \in \mathbb{R}^+$ = margin for error

Margin for error semantics

Williamson 1992, 1994, "Logic of Clarity"

- Margin models: $M = \langle W, d, \alpha, V \rangle$

  d = metric over W
  $\alpha \in \mathbb{R}^+$ = margin for error
- $M, w \models_{FM} \Box\phi$ iff for all v s. t. $d(v, w) \leq \alpha$, $M, v \models_{FM} \phi$.

  "I know ϕ iff ϕ holds throughout the margin of error"

Theorem (Williamson 1992)

$\models_{FM} \phi$ *iff* $\vdash_{KTB} \phi$

Theorem (Williamson 1992)

$\models_{FM} \phi$ *iff* $\vdash_{KTB} \phi$

Corollary

Neither 4 nor 5 is FM-valid.

Luminosity

- Luminosity Paradox: suppose $\Box p \rightarrow \Box\Box p$ were to hold everywhere in the model. Then: $0 \models \Box p \Rightarrow i \models p$ for every $i \geq 0$: "every pen will fit in the box" (!)

Luminosity

- Luminosity Paradox: suppose $\Box p \rightarrow \Box\Box p$ were to hold everywhere in the model. Then: $0 \models \Box p \Rightarrow i \models p$ for every $i \geq 0$: "every pen will fit in the box" (!)
- Call a sentence ϕ luminous iff $\phi \rightarrow \Box\phi$ is valid.

Luminosity

- Luminosity Paradox: suppose $\Box p \rightarrow \Box\Box p$ were to hold everywhere in the model. Then: $0 \models \Box p \Rightarrow i \models p$ for every $i \geq 0$: "every pen will fit in the box" (!)
- Call a sentence ϕ luminous iff $\phi \rightarrow \Box\phi$ is valid.

Theorem (Williamson 1992)

$\models_{FM} \phi \rightarrow \Box\phi$ *iff* $\models_{FM} \phi$ *or* $\models_{FM} \neg\phi$

Luminosity

- Luminosity Paradox: suppose $\Box p \rightarrow \Box\Box p$ were to hold everywhere in the model. Then: $0 \models \Box p \Rightarrow i \models p$ for every $i \geq 0$: "every pen will fit in the box" (!)
- Call a sentence ϕ luminous iff $\phi \rightarrow \Box\phi$ is valid.

Theorem (Williamson 1992)

$\models_{FM} \phi \rightarrow \Box\phi$ *iff* $\models_{FM} \phi$ *or* $\models_{FM} \neg\phi$

- Whenever knowledge obeys a margin for error, the only luminous properties are the trivial properties (holding everywhere or nowhere)

Anti-luminosity

Application to mental states:

- A state of mind e is luminous iff its occurrence entails the knowledge that one is in e
- A state of mind is non-trivial iff it lasts for some time, not all the time

Anti-luminosity

Application to mental states:

- A state of mind *e* is luminous iff its occurrence entails the knowledge that one is in *e*
- A state of mind is non-trivial iff it lasts for some time, not all the time

Anti-Luminosity: no non-trivial mental state is luminous, not even states of knowledge (Williamson 2000)

Supervenience issue

Things may be viewed the other way around:

- How can I know that I know without knowing that I know that I know? or know that I know that I know without knowing that I know that I know that I know?

Supervenience issue

Things may be viewed the other way around:

- How can I know that I know without knowing that I know that I know? or know that I know that I know without knowing that I know that I know that I know?
- Supervenience issue: should each further level of knowledge necessarily supervene on more than what the previous level supervenes on?

Supervenience issue

Things may be viewed the other way around:

- How can I know that I know without knowing that I know that I know? or know that I know that I know without knowing that I know that I know that I know?
- Supervenience issue: should each further level of knowledge necessarily supervene on more than what the previous level supervenes on?

Answer: not necessarily so, possibly second-order knowledge supervenes only on no-more than first-order knowledge.

Centered Semantics

Bonnay & Egré 2006, 2008

- A "cartesian" logic of knowledge, satisfying strong introspection properties
- A contextualist, two-dimensional semantics, in which alternatives relevant to evaluate higher-order knowledge are the same as those relevant for the evaluation of first-order knowledge

Centered semantics

Given a Kripke structure $M = \langle W, R, V\rangle$ like the one pictured:

1. $M, (w, w') \models_{CS} p$ iff $w' \in V(p)$
2. $M, (w, w') \models_{CS} \neg p$ iff $M, (w, w') \not\models_{CS} p$
3. $M, (w, w') \models_{CS} (\phi \wedge \psi)$ iff $M, (w, w') \models_{CS} \phi$ and $M, (w, w') \models_{CS} \psi$.
4. $M, (w, w') \models_{CS} \Box\phi$ iff for every w'' such that wRw'', $M, (w, w'') \models_{CS} \phi$

Centered semantics

Given a Kripke structure $M = \langle W, R, V \rangle$ like the one pictured:

1. $M, (w, w') \models_{CS} p$ iff $w' \in V(p)$
2. $M, (w, w') \models_{CS} \neg p$ iff $M, (w, w') \not\models_{CS} p$
3. $M, (w, w') \models_{CS} (\phi \wedge \psi)$ iff $M, (w, w') \models_{CS} \phi$ and $M, (w, w') \models_{CS} \psi$.
4. $M, (w, w') \models_{CS} \Box\phi$ iff for every w'' such that wRw'', $M, (w, w'') \models_{CS} \phi$

Def: $M, w \models_{CS} \phi$ iff $M, (w, w) \models_{CS} \phi$

Centered semantics

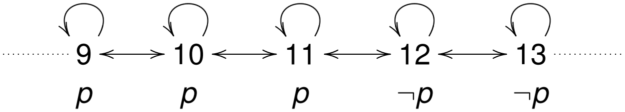
Given a Kripke structure $M = \langle W, R, V \rangle$ like the one pictured:

1. $M, (w, w') \models_{CS} p$ iff $w' \in V(p)$
2. $M, (w, w') \models_{CS} \neg p$ iff $M, (w, w') \not\models_{CS} p$
3. $M, (w, w') \models_{CS} (\phi \wedge \psi)$ iff $M, (w, w') \models_{CS} \phi$ and $M, (w, w') \models_{CS} \psi$.
4. $M, (w, w') \models_{CS} \Box\phi$ iff for every w'' such that wRw'', $M, (w, w'') \models_{CS} \phi$

Def: $M, w \models_{CS} \phi$ iff $M, (w, w) \models_{CS} \phi$

- “Perceptual” statements are evaluated with respect to the second index, and “Reflective” statements are evaluated w.r.t. the first index only.

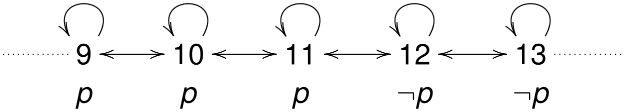
Example



- Identical predictions for first-order knowledge:

$10 \models_{CS} \Box p$

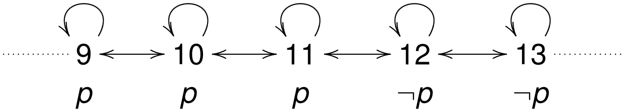
Example



- Identical predictions for first-order knowledge:

$10 \models_{cs} \Box p$
for $(10, 9), (10, 10), (10, 11) \models_{cs} p$

Example



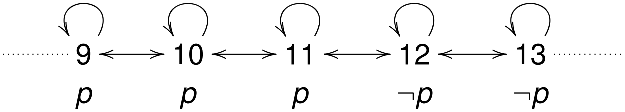
- Identical predictions for first-order knowledge:

$10 \models_{CS} \Box p$
for $(10, 9), (10, 10), (10, 11) \models_{CS} p$

- But different predictions at higher-orders:

$10 \models_{CS} \Box\Box p$

Example



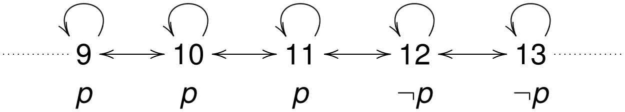
- Identical predictions for first-order knowledge:

$10 \models_{CS} \Box p$
for $(10, 9), (10, 10), (10, 11) \models_{CS} p$

- But different predictions at higher-orders:

$10 \models_{CS} \Box\Box p$
$\Leftrightarrow (10, 10) \models_{CS} \Box\Box p$

Example



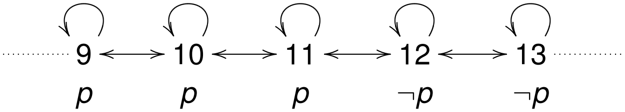
- Identical predictions for first-order knowledge:

$10 \models_{CS} \Box p$
for $(10, 9), (10, 10), (10, 11) \models_{CS} p$

- But different predictions at higher-orders:

$10 \models_{CS} \Box\Box p$
$\Leftrightarrow (10, 10) \models_{CS} \Box\Box p$
$\Leftrightarrow (10, 9)$, $(10, 10)$ and $(10, 11) \models_{CS} \Box p$

Example



- Identical predictions for first-order knowledge:

$10 \models_{CS} \Box p$
for $(10, 9), (10, 10), (10, 11) \models_{CS} p$

- But different predictions at higher-orders:

$10 \models_{CS} \Box\Box p$
$\Leftrightarrow (10, 10) \models_{CS} \Box\Box p$
$\Leftrightarrow (10, 9), (10, 10)$ and $(10, 11) \models_{CS} \Box p$
$\Leftrightarrow (10, 9), (10, 10), (10, 11) \models_{CS} p$: $\surd$

Main properties

Theorem

Proposition 1: $\models_{CS} \phi$ *iff* $\vdash_{K45} \phi$

$\Rightarrow$ CS as a logic of introspective belief

Main properties

Theorem

Proposition 1: $\models_{CS} \phi$ *iff* $\vdash_{K45} \phi$

$\Rightarrow$ CS as a logic of introspective belief

Definition: (CMS semantics) $M, (w, w') \models_{CMS} \Box\phi$ iff for every v such that $d(w, v) \leq \alpha$, $M, (w, v) \models_{CMS} \phi$

Main properties

Theorem

Proposition 1: $\models_{CS} \phi$ *iff* $\vdash_{K45} \phi$

$\Rightarrow$ CS as a logic of introspective belief

Definition: (CMS semantics) $M, (w, w') \models_{CMS} \Box\phi$ iff for every v such that $d(w, v) \leq \alpha$, $M, (w, v) \models_{CMS} \phi$

Theorem

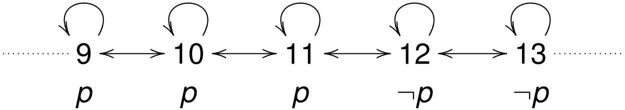
Proposition 2: $\models_{CMS} \phi$ *iff* $\vdash_{S5} \phi$

$\Rightarrow$ CMS as a logic of introspective knowledge

$\Rightarrow$ K45 and S5 are not logics of exact knowledge per se, since we can now work with non-transitive and non-euclidian models.

Back to luminosity

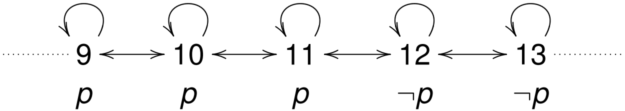
Luminosity-without-triviality: $\models_{cs} \phi \rightarrow \Box\phi \not\Rightarrow \models_{cs} \phi$ or $\models_{cs} \neg\phi$



$\Box p$ is luminous in the model, yet not trivial.

Back to luminosity

Luminosity-without-triviality: $\models_{cs} \phi \to \Box\phi \not\Rightarrow \models_{cs} \phi$ or $\models_{cs} \neg\phi$



$\Box p$ is luminous in the model, yet not trivial.

Consequence: we can agree with Williamson that not every mental state is luminous, or even that most of our mental states are not luminous, and still disagree about knowledge (seen as a mental state).

Comparisons

CS can be related to:

- Standard 2d-semantics with actuality operators (enriching the language)
- Halpern's 2d semantics (transforming the models)

Actuality operators
Indexical knowledge

"I know ϕ iff ϕ holds at all my actual epistemic alternatives. (cf. Kamp 1971 for the analog in temporal case)

- $M, (w, w') \vDash_{\text{K2S}} A\phi$ iff $M, (w, w) \models \phi$
- $M, (w, w') \vDash_{\text{K2S}} K\phi$ iff for every w'' such that $w'Rw''$, $M, (w, w'') \vDash_{\text{K2S}} \phi$

Translation from $\mathcal{L}(K)$ to $\mathcal{L}(A, K)$: $p^* = p$, $(\phi \wedge \psi)^* = (\phi^* \wedge \psi^*)$, $(\neg\phi)^* = \neg\phi^*$, $(K\phi)^* = AK\phi^*$

Theorem

$M, (w, w') \vDash_{\text{CS}} \phi$ iff $M, (w, w') \vDash_{\text{K2S}} \phi^*$

Halpern's logic

Also a two-dimensional framework, but for a logic with two modalities:

"Intransitivity in reports of perceptions does not necessarily imply intransitivity in actual perceptions" (Halpern 2004)

- $R\phi$: "I report that ϕ" ($\Box\phi$)
- $D\phi$: "according to me, ϕ is definitely the case"

Main idea: the composition of two equivalence relations need not be transitive.

Halpern's semantics

A simplified Halpern model: $M = \langle W, \sim_s, \sim_o, V \rangle$, with $W \subseteq S \times O$

$\sim_s$, $\sim_o$ equivalence relations

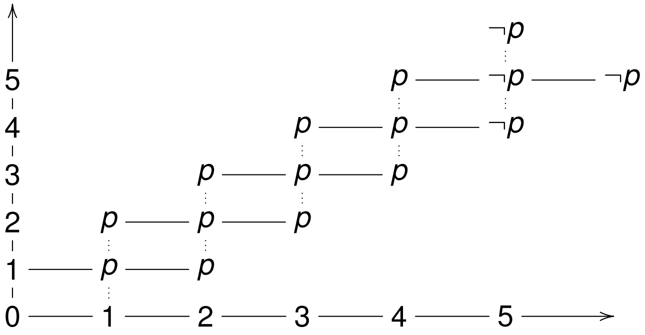
- $M, (w, w') \models R\phi$ iff for every (t, t') such that $(w, w') \sim_s (t, t')$, $M, (t, t') \models \phi$.
- $M, (w, w') \models D\phi$ iff for every (t, t') such that $(w, w') \sim_o (t, t')$, $M, (t, t') \models \phi$.

ex: $M, (2, 3) \models Rp$: when the actual value is 3 and when I measure 2, I report that p"

$W = \{(n, m) \in \mathbb{N} \times \mathbb{N}; |n - m| \leq 1\}$

$(n, m) \sim_s (n', m')$ iff $m = m'$ $(n, m) \sim_o (n', m')$ iff $n = n'$

$(2, 3) \models DRp$, but $(2, 3) \not\models DRDRp$



Layering

Transformation: $M = \langle W, R, V \rangle \rightsquigarrow L(M) = \langle W', R', V' \rangle$

- $W' = \{(w, w') \in W \times W; w'Rw \vee w' = w\}$
- $(w, w')R'(u, u')$ iff $w' = u'$ and $w'Ru$
- $(w, w') \in V'(p)$ iff $w \in V(p)$.

Theorem

For all $(w, w') \in L(M)$: $M, (w, w') \vDash_{\mathrm{CS}} \phi$ *iff* $L(M), (w', w) \models \phi$

Corollary

$M, w \vDash_{\mathrm{CS}} \phi$ *iff* $L(M), (w, w) \models \phi$

NB. Given any R, R' is necessarily transitive and euclidian.

Interpretation

Layering shows how to recover a transitive relation of epistemic uncertainty from a non-transitive relation.

Same relativization of higher-order knowledge to actual epistemic alternatives

Summary for today

What did we see?

- Basic epistemic logic
- Margin semantics
- Centered semantics
- Correspondence with other two-dimensional frameworks

Main lesson from today

- Positive and negative introspection can be forced to be valid on non-transitive/non-euclidean structures
- Centered semantics does not handle first-order knowledge and higher-order knowledge on a par: FO-knowledge is constrained by a margin of error, but not so for HO-knowledge.

What are we going to see tomorrow

Closer confrontation between Williamson's argument and the present framework:

What are we going to see tomorrow

Closer confrontation between Williamson's argument and the present framework:

- Token semantics: generalization of Centered semantics

What are we going to see tomorrow

Closer confrontation between Williamson's argument and the present framework:

- Token semantics: generalization of Centered semantics
- Finer features of Centered Semantics
- Applications to common knowledge / Dynamic version of CS