Epistemic Logics for Introspection
Part II

Paul Égré

Institut Jean-Nicod, CNRS

FEW 09, Pittsburgh CMU - June 20, 2009

What did we see yesterday?

- CS validates positive and negative introspection over arbitrary Kripke structures

What did we see yesterday?

- CS validates positive and negative introspection over arbitrary Kripke structures
- How does it do this?

What did we see yesterday?

- CS validates positive and negative introspection over arbitrary Kripke structures
- How does it do this? double-indexing

What did we see yesterday?

- CS validates positive and negative introspection over arbitrary Kripke structures
- How does it do this? double-indexing
- What else is it good for?

Outline

- Token semantics
- Common Knowledge and Almost Common Knowledge
- Multi-agent Token Semantics
- The Email Game

Generalizing Centered Semantics

Key ideas

- CS allows one to visit only a subpart of the initial model: worlds 1 step away. What about relaxing this constraint to worlds that are 2 steps away, 3 steps,... n steps?

Generalizing Centered Semantics

Key ideas

- CS allows one to visit only a subpart of the initial model: worlds 1 step away. What about relaxing this constraint to worlds that are 2 steps away, 3 steps,... *n* steps?
- Motivation: think of *n* as the number of nestings of knowledge operators that require checking

Generalizing Centered Semantics

Key ideas

- CS allows one to visit only a subpart of the initial model: worlds 1 step away. What about relaxing this constraint to worlds that are 2 steps away, 3 steps,... n steps?
- Motivation: think of n as the number of nestings of knowledge operators that require checking
- This number n is materialized by means of a parameter: tokens

Generalizing Centered Semantics

Key ideas

- CS allows one to visit only a subpart of the initial model: worlds 1 step away. What about relaxing this constraint to worlds that are 2 steps away, 3 steps,... *n* steps?
- Motivation: think of *n* as the number of nestings of knowledge operators that require checking
- This number *n* is materialized by means of a parameter: tokens
- Enlargement of the supervenience basis of higher-order knowledge (relative to CS)

Tokens

- Formulas are evaluated relative to a number n of tokens

Tokens

- Formulas are evaluated relative to a number *n* of tokens
- For each non-trivial move in a model (box or diamond), a token is spent, so not for reflexive moves, which come at no cost.

Tokens

- Formulas are evaluated relative to a number *n* of tokens
- For each non-trivial move in a model (box or diamond), a token is spent, so not for reflexive moves, which come at no cost.
- When all tokens have been spent, get a token back, backtrack to the previous position in the model, and continue (loop).

Token semantics

Evaluation of sentences with respect to a sequence of worlds (and a token):

(i) $M, qw \models_{\mathrm{TS}} p\ [n]$ iff $w \in V(p)$.

Token semantics

Evaluation of sentences with respect to a sequence of worlds (and a token):

(i) $M, qw \models_{\text{TS}} p\ [n]$ iff $w \in V(p)$.

(ii) $M, qw \models_{\text{TS}} \neg\phi\ [n]$ iff $M, qw \not\models_{TS} \phi\ [n]$.

Token semantics

Evaluation of sentences with respect to a sequence of worlds (and a token):

(i) $M, qw \models_{\text{TS}} p\ [n]$ iff $w \in V(p)$.

(ii) $M, qw \models_{\text{TS}} \neg\phi\ [n]$ iff $M, qw \not\models_{TS} \phi\ [n]$.

(iii) $M, qw \models_{\text{TS}} (\phi \wedge \psi)\ [n]$ iff $M, qw \models_{\text{TS}} \phi\ [n]$ and $M, qw \models_{\text{TS}} \psi\ [n]$.

Token semantics

Evaluation of sentences with respect to a sequence of worlds (and a token):

(i) $M, qw \models_{\text{TS}} p$ [n] iff $w \in V(p)$.
(ii) $M, qw \models_{\text{TS}} \neg\phi$ [n] iff $M, qw \not\models_{TS} \phi$ [n].
(iii) $M, qw \models_{\text{TS}} (\phi \wedge \psi)$ [n] iff $M, qw \models_{\text{TS}} \phi$ [n] and $M, qw \models_{\text{TS}} \psi$ [n].
(iv) $M, qw \models_{\text{TS}} \Box\psi$ [n] iff

Token semantics

Evaluation of sentences with respect to a sequence of worlds (and a token):

(i) $M, qw \models_{TS} p\ [n]$ iff $w \in V(p)$.

(ii) $M, qw \models_{TS} \neg\phi\ [n]$ iff $M, qw \not\models_{TS} \phi\ [n]$.

(iii) $M, qw \models_{TS} (\phi \wedge \psi)\ [n]$ iff $M, qw \models_{TS} \phi\ [n]$ and $M, qw \models_{TS} \psi\ [n]$.

(iv) $M, qw \models_{TS} \Box\psi\ [n]$ iff

- $n \neq 0$ and for all w' s.t. wRw', $qww' \models_{TS} \psi\ [n-k]$, with $k = 1$ if $w \neq w'$, $k = 0$ if $w = w'$.

Token semantics

Evaluation of sentences with respect to a sequence of worlds (and a token):

(i) $M, qw \models_{TS} p\ [n]$ iff $w \in V(p)$.

(ii) $M, qw \models_{TS} \neg\phi\ [n]$ iff $M, qw \not\models_{TS} \phi\ [n]$.

(iii) $M, qw \models_{TS} (\phi \wedge \psi)\ [n]$ iff $M, qw \models_{TS} \phi\ [n]$ and $M, qw \models_{TS} \psi\ [n]$.

(iv) $M, qw \models_{TS} \Box\psi\ [n]$ iff

- $n \neq 0$ and for all w' s.t. wRw', $qww' \models_{TS} \psi\ [n-k]$, with $k = 1$ if $w \neq w'$, $k = 0$ if $w = w'$.
- Or $n = 0$ and $q \models_{TS} \Box\psi\ [1]$.

Token semantics

Evaluation of sentences with respect to a sequence of worlds (and a token):

(i) $M, qw \models_{TS} p\ [n]$ iff $w \in V(p)$.

(ii) $M, qw \models_{TS} \neg\phi\ [n]$ iff $M, qw \not\models_{TS} \phi\ [n]$.

(iii) $M, qw \models_{TS} (\phi \wedge \psi)\ [n]$ iff $M, qw \models_{TS} \phi\ [n]$ and $M, qw \models_{TS} \psi\ [n]$.

(iv) $M, qw \models_{TS} \Box\psi\ [n]$ iff

- $n \neq 0$ and for all w' s.t. wRw', $qww' \models_{TS} \psi\ [n-k]$, with $k = 1$ if $w \neq w'$, $k = 0$ if $w = w'$.
- Or $n = 0$ and $q \models_{TS} \Box\psi\ [1]$.

Def: [TS(n)-semantics] $M, w \models_{TS(n)} \phi$ iff $M, w \models_{TS} \phi\ [n]$.

Token semantics

Evaluation of sentences with respect to a sequence of worlds (and a token):

(i) $M, qw \models_{TS} p\ [n]$ iff $w \in V(p)$.

(ii) $M, qw \models_{TS} \neg\phi\ [n]$ iff $M, qw \not\models_{TS} \phi\ [n]$.

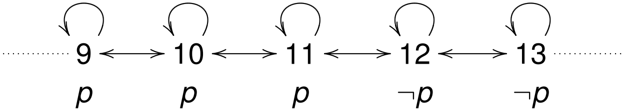
(iii) $M, qw \models_{TS} (\phi \wedge \psi)\ [n]$ iff $M, qw \models_{TS} \phi\ [n]$ and $M, qw \models_{TS} \psi\ [n]$.

(iv) $M, qw \models_{TS} \Box\psi\ [n]$ iff

- $n \neq 0$ and for all w' s.t. wRw', $qww' \models_{TS} \psi\ [n - k]$, with $k = 1$ if $w \neq w'$, $k = 0$ if $w = w'$.
- Or $n = 0$ and $q \models_{TS} \Box\psi\ [1]$.

Def: [TS(n)-semantics] $M, w \models_{TS(n)} \phi$ iff $M, w \models_{TS} \phi\ [n]$.

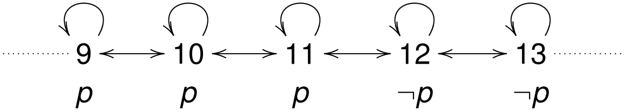
Example



- $10 \models_{TS} \Box p$ [1]

for $(10, 9), (10, 11) \models_{TS} p$ [0] and $(10, 10) \models_{TS} p$ [1]

Example



- $10 \models_{TS} \Box p$ [1]

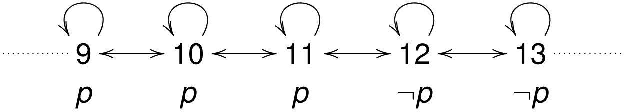
for $(10,9),(10,11) \models_{TS} p$ [0] and $(10,10) \models_{TS} p$ [1]

As in CS:

- $10 \models_{TS} \Box\Box p$ [1]

$\Leftrightarrow (10,x) \models_{TS} \Box p$ [0] for $x = 9, 11$, and $(10,x) \models_{TS} \Box p$ [1] for $x = 10$.

Example



- $10 \models_{TS} \Box p$ [1]

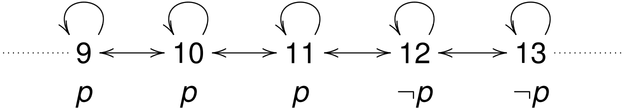
for $(10, 9), (10, 11) \models_{TS} p$ [0] and $(10, 10) \models_{TS} p$ [1]

As in CS:

- $10 \models_{TS} \Box\Box p$ [1]

$\Leftrightarrow (10, x) \models_{TS} \Box p$ [0] for $x = 9, 11$, and $(10, x) \models_{TS} \Box p$ [1] for $x = 10$.
$\Leftrightarrow 10 \models_{TS} \Box p$ [1] and $(10, 10) \models_{TS} \Box p$ [1]

Example Cont'd



However, $10 \not\models_{TS} \Box\Box p$ [2]

otherwise we would have $10, 11 \models_{\mathrm{TS}} \Box p$ [1]

and, $10, 11, 12 \models_{\mathrm{TS}} p$ [0]: not so.

A spectrum of semantics

$TS(1)$	aka Centered Semantics, validates positive and negative introspection over arbitrary structures
$TS(\omega)$	aka Kripke Semantics, no introspection principles are validated
$TS(n)$	$1 < n < \omega$, aka Token Semantics, weakened versions of the introspection principles

Main properties

- Each TS(n)-semantics has a sound and complete axiomatization

Main properties

- Each TS(n)-semantics has a sound and complete axiomatization
- The resulting logics are intermediate in strength between K45 and K

ex: $TS(2) \models \Box\Box p \rightarrow \Box\Box\Box p$
ex: $TS(3) \models \Box^3 p \rightarrow \Box\Box^3 p$
...

Main consequences for introspection

- Gradient between automatic introspection and introspection at the second order: I may fail to know that I know, but if I know that I know, then I automatically know that I know that I know.

Main consequences for introspection

- Gradient between automatic introspection and introspection at the second order: I may fail to know that I know, but if I know that I know, then I automatically know that I know that I know.
- A more fine-grained control of iterations
- Interest for the multi-agent case

Common knowledge

- Shared knowledge: everyone knows that *p*
- Common knowledge: everyone knows that *p*, everyone knows that everyone knows that *p*, everyone knows that everyone knows that everyone knows that *p*, ...

Multi-agent Epistemic logic

- $\Box_a \phi$: *a* knows/believes ϕ
- $E_{a,b}\phi \equiv \Box_a \phi \wedge \Box_b \phi$
- $C_{a,b}\phi \equiv E_{a,b}\phi \wedge E_{a,b}E_{a,b}\phi \wedge \ldots$

Multi-agent Epistemic logic

- $\Box_a \phi$: *a* knows/believes ϕ
- $E_{a,b}\phi \equiv \Box_a \phi \wedge \Box_b \phi$
- $C_{a,b}\phi \equiv E_{a,b}\phi \wedge E_{a,b}E_{a,b}\phi \wedge \ldots$

- $M, w \models \Box_a \phi$ iff for every w' in $R_a(w)$, $M, w' \models \phi$
- $M, w \models E_{a,b}\phi$ iff for every w' in $(R_a \cup R_b)(w)$, $M, w' \models \phi$
- $M, w \models C_{a,b}\phi$ iff for every w' in $(R_a \cup R_b)^*(w)$, $M, w' \models \phi$.

Attaining Common Knowledge

Attaining CK sometimes can be easy, sometimes can be hard:

Attaining Common Knowledge

Attaining CK sometimes can be easy, sometimes can be hard:

- Public announcements (static, easy): “the deck has 52 cars”

Attaining Common Knowledge

Attaining CK sometimes can be easy, sometimes can be hard:

- Public announcements (static, easy): “the deck has 52 cars”
- Coordinated attack problem (dynamic, hard): 2 generals communicate sequentially; *a* send a message to *b* to say he will attack at dawn; *b* replies to *a* to confirm reception of the message; *a* replies to *b* to say he got *b*’s confirmation,...

Consecutive numbers

Kooi, van Ditmarsh, van der Hoek 2003

Two agents *a* and *b* each are given a positive natural number. Each one knows his number, not the number of the other. It is a public rule that the numbers are consecutive.

Consecutive numbers

Kooi, van Ditmarsh, van der Hoek 2003

Two agents *a* and *b* each are given a positive natural number. Each one knows his number, not the number of the other. It is a public rule that the numbers are consecutive.

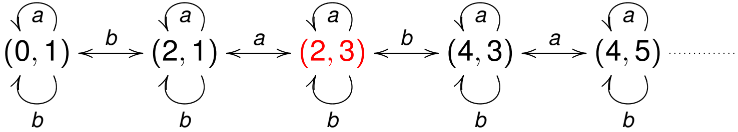
Example: *a* holds a 2 and *b* holds a 3. Is it common knowledge between them that their numbers are less than 100, 1000, 10000...?

Consecutive numbers

Kooi, van Ditmarsh, van der Hoek 2003

Two agents *a* and *b* each are given a positive natural number. Each one knows his number, not the number of the other. It is a public rule that the numbers are consecutive.

Example: *a* holds a 2 and *b* holds a 3. Is it common knowledge between them that their numbers are less than 100, 1000, 10000...?



A puzzle about common knowledge

$\phi_{\leq 10000}$:= "*a* and *b*'s numbers are less than 10000"
$\phi_{\leq n}$= "*a* and *b*'s numbers are less than *n*"

A puzzle about common knowledge

$\phi_{\leq 10000}$:= "*a* and *b*'s numbers are less than 10000"
$\phi_{\leq n}$= "*a* and *b*'s numbers are less than *n*"

$$M, (2, 3) \not\models C_{a,b}\phi_{\leq 10000}$$

A puzzle about common knowledge

$\phi_{\leq 10000}$:= "*a* and *b*'s numbers are less than 10000"
$\phi_{\leq n}$= "*a* and *b*'s numbers are less than *n*"

$$M, (2, 3) \not\models C_{a,b}\phi_{\leq 10000}$$

More generally, for every *n*,

$$M, (2, 3) \not\models C_{a,b}\phi_{\leq n}$$

A puzzle about common knowledge

$\phi_{\leq 10000}$:= "*a* and *b*'s numbers are less than 10000"
$\phi_{\leq n}$= "*a* and *b*'s numbers are less than *n*"

$$M, (2, 3) \not\models C_{a,b}\phi_{\leq 10000}$$

More generally, for every *n*,

$$M, (2, 3) \not\models C_{a,b}\phi_{\leq n}$$

Common knowledge about the size of the numbers is never attained, however large the number.

Two intuitions

Step by step reasoning:

a if *b* holds a 3
he may think I hold a 4 ($\Diamond_b 4_a$)
and think that [if I hold a 4] I think he holds a 5 ($\Diamond_b \Diamond_a 5_b$)
and think I think that [if he holds a 5] he may think I hold a 6 ($\Diamond_b \Diamond_a \Diamond_b 6_a$)

Spontaneous intuition: *a* and *b* both know that both numbers are less than 100000. Each of them believes that the other believes it, and so on / that it is common knowledge

Fixed Point vs Iterative Definition

FP: $Cp \leftrightarrow E(p \wedge Cp)$

IT: $Cp \leftrightarrow Ep \wedge EEp \wedge EEEp \wedge \ldots$

- Observation: the fixed-point definition seems better to capture the intuition that common knowledge is attainable in Consecutive Numbers.
- However: the two definitions coincide in the framework of standard Kripke semantics (see van Benthem & Sarenac 2004 for a separation)

Almost common knowledge

- (Rubinstein 1989) “by ‘almost common knowledge’, I refer to the case when the numbers [of iterations] are ‘very large’”: ie sufficiently large but finite amount of shared knowledge (NB. probably what people would intuitively understand by CK)

Almost common knowledge

- (Rubinstein 1989) "by 'almost common knowledge', I refer to the case when the numbers [of iterations] are 'very large'": ie sufficiently large but finite amount of shared knowledge (NB. probably what people would intuitively understand by CK)
- In the game of consecutive numbers, the agents have almost common knowledge that the numbers are less than, say, 1000, or even 100

Proposal

- Account for situations of this kind by generalizing tokens to several agents
- Show that (almost) common knowledge can then be reduced to a finite level of shared knowledge

Multi-agent Token Semantics (2 agents)

- Main idea: use as many token registers as there are agents

$$M, qw \vDash_{\text{MTS}} \phi\ [m_a, m_b]$$

- The semantics, informally: same as the one-agent case, but when $m_i = 0$ and $\Box_i$ is to be evaluated:

  (i) backtrack to the closest antecedent world v reached by an i-move

  (ii) pick up and reassign all tokens that were spent along the way, including for other agents.

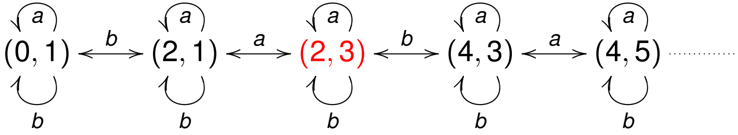
  (iii) continue.

Common Knowledge Trivialized

Theorem (trivialization)

$\vDash_{\mathrm{MTS}} (E_{a,b})^{\leq n+n}\phi \leftrightarrow C_{a,b}\ \phi\ [n,n]$

Example: $M,(2,3) \vDash_{\mathrm{MTS}} C_{a,b}\ \phi_{\leq 5}\ [1,1]$



Interpretation

How legitimate is it to equate common knowledge with some finite amount of shared knowledge?

In principle, the use of TS is neutral between two interpretations:

- Illusion of common knowledge as a side-effect of bounded rationality (agents are lazy in their computations)
  or
- Common knowledge actually reached on a finite amount of shared knowledge.

Problem: how can we tease apart the two interpretations?

An objection against TS

If it is CK that *a*'s number is less than $k + 1$, then it is CK that it is less than k [from an. rev.]

An objection against TS

If it is CK that a's number is less than $k+1$, then it is CK that it is less than k [from an. rev.]

1. $C\,(n(a) \leq k+1)$ (assumption)

An objection against TS

If it is CK that *a*'s number is less than $k + 1$, then it is CK that it is less than k [from an. rev.]

1. $C\ (n(a) \leq k + 1)$ (assumption)

2. $n(a) = k + 1 \rightarrow \neg\Box_a\Box_b\ (n(a) \leq k + 1)$ (structure of the game)

An objection against TS

If it is CK that *a*'s number is less than $k + 1$, then it is CK that it is less than k [from an. rev.]

1. $C\ (n(a) \leq k + 1)$ (assumption)
2. $n(a) = k + 1 \rightarrow \neg\Box_a\Box_b\ (n(a) \leq k + 1)$ (structure of the game)
3. $C\ (n(a) = k + 1) \rightarrow \neg\Box_a\Box_b\ (n(a) \leq k + 1)$ (CK of the structure of the game)

An objection against TS

If it is CK that *a*'s number is less than $k+1$, then it is CK that it is less than k [from an. rev.]

1. $C\,(n(a) \leq k+1)$ (assumption)
2. $n(a) = k+1 \rightarrow \neg\Box_a\Box_b\,(n(a) \leq k+1)$ (structure of the game)
3. $C\,(n(a) = k+1) \rightarrow \neg\Box_a\Box_b\,(n(a) \leq k+1)$ (CK of the structure of the game)
4. $C\Box_a\Box_b\,(n(a) \leq k+1)$ (from the def. of C)

An objection against TS

If it is CK that *a*'s number is less than $k+1$, then it is CK that it is less than k [from an. rev.]

1. $C\ (n(a) \leq k+1)$ (assumption)

2. $n(a) = k+1 \rightarrow \neg\Box_a\Box_b\ (n(a) \leq k+1)$ (structure of the game)

3. $C\ (n(a) = k+1) \rightarrow \neg\Box_a\Box_b\ (n(a) \leq k+1)$ (CK of the structure of the game)

4. $C\Box_a\Box_b\ (n(a) \leq k+1)$ (from the def. of C)

5. $C\ (n(a) \neq k+1)$

An objection against TS

If it is CK that *a*'s number is less than $k + 1$, then it is CK that it is less than k [from an. rev.]

1. $C\ (n(a) \leq k + 1)$ (assumption)
2. $n(a) = k + 1 \rightarrow \neg\Box_a\Box_b\ (n(a) \leq k + 1)$ (structure of the game)
3. $C\ (n(a) = k + 1) \rightarrow \neg\Box_a\Box_b\ (n(a) \leq k + 1)$ (CK of the structure of the game)
4. $C\Box_a\Box_b\ (n(a) \leq k + 1)$ (from the def. of C)
5. $C\ (n(a) \neq k + 1)$

However: *TS* does not validate the inference from 2 to 3: a property can be true everywhere in a game without being CK.

Interpretation

- Fact: In TS, a property can be true everywhere in a game without being CK

Interpretation

- Fact: In TS, a property can be true everywhere in a game without being CK

1. $n(a) = k \rightarrow \Diamond_a \Diamond_b (n(a) > k)$ (structure of the game)
2. $C (n(a) = k \rightarrow \Diamond_a \Diamond_b (n(a) > k))$ (CK of the structure of the game)

Interpretation

- Fact: In TS, a property can be true everywhere in a game without being CK

1. $n(a) = k \rightarrow \Diamond_a \Diamond_b (n(a) > k)$ (structure of the game)
2. $C(n(a) = k \rightarrow \Diamond_a \Diamond_b (n(a) > k))$ (CK of the structure of the game)

- The concept of CK described using TS is most likely a common illusion of common knowledge, rather than real common knowledge.
- This does not mean that such a notion is not operational for practical decisions.

The Electronic Mail Game

Rubinstein 1989

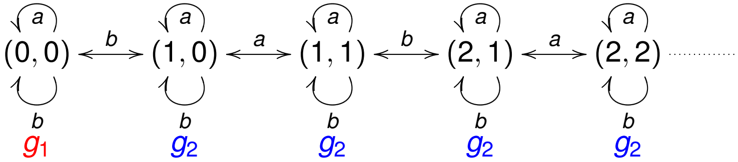
Bayesian game: Agents a and b have the choice between two actions A and B. The game is either g_1 or g_2, depending on the state of nature, which only a can observe. a sends an email to b only if the game is g_2; b's machine sends an automatic response in that case, and likewise for a. Both machines have the same probability of transmission failure ε. Each agent sees on his screen the number of messages he sent at the end of the communication process, but not the other's number.

g_1	A	B
A	10,10	0, -5
B	-5,0	0,0

g_2	A	B
A	0,0	0, -5
B	-5,0	10,10

g_1	A	B
A	10,10	0, -5
B	-5,0	0,0

g_2	A	B
A	0,0	0, -5
B	-5,0	10,10



Theorem (Rubinstein)

The email game has a unique Nash Equilibrium, in which both players always choose A.

Main ingredients of the proof:

- Induction, with base case the fact that action A is strictly dominant for *a* in the state (0,0) (when the game is g_1)
- Bayesian hypotheses in order to compute *b*'s best action in that case and in the following.

Diagnosis

Rubinstein: “the source of the discrepancy lies in the fact that mathematical induction is not part of the reasoning process of human beings”.

Diagnosis

Rubinstein: “the source of the discrepancy lies in the fact that mathematical induction is not part of the reasoning process of human beings”.

- the induction proof rests crucially on the fact that the state (0,0) is a relevant epistemic alternative for at least one player
- However, it is relevant only when the numbers are sufficiently small. When the numbers are high, agents simply fail to compute knowledge iterations that would lead them too far from their respective context.

Towards a solution

- Suppose the real state of the world is (17,16), namely *a*'s last message failed. p_1 = the game is G_1, and p_2 = the game is G_2.
- Suppose that each agent has 2 tokens

$$(17, 16) \vDash_{\mathrm{MTS}} C_{a,b}p_2 \; [2, 2]$$

Work in Progress

- Can (B,B) be derived as an interesting outcome (equilibrium?) of the game, if one makes use of the revised concept of common knowledge?
- Idea: consider the first state (m, n) from $(0, 0)$ such that it becomes CK (in MTS) that the game played is g_2. Can we prove that below (m, n), (A,A) is the equilibrium, and that from (m, n) onward, (B,B) becomes the equilibrium?

Common knowledge of (bounded) rationality

Do agents have common knowledge about their number of tokens?

Common knowledge of (bounded) rationality

Do agents have common knowledge about their number of tokens?

- "Common knowledge of rationality" is taken to mean that agents have CK that they are fully rational: in our framework, this means (background) CK that the number of tokens is ω.

Common knowledge of (bounded) rationality

Do agents have common knowledge about their number of tokens?

- "Common knowledge of rationality" is taken to mean that agents have CK that they are fully rational: in our framework, this means (background) CK that the number of tokens is ω.
- "Almost common knowledge" is vague: CK that the number of tokens is n for finite n is therefore implausible.

Common knowledge of (bounded) rationality

Do agents have common knowledge about their number of tokens?

- “Common knowledge of rationality” is taken to mean that agents have CK that they are fully rational: in our framework, this means (background) CK that the number of tokens is ω.
- “Almost common knowledge” is vague: CK that the number of tokens is n for finite n is therefore implausible.
- Maybe subjects simply make use of the underspecified and highly context-dependent principle that the computation of iterations cannot reasonably exceed a certain bound.

The Vagueness problem

- Arbitrariness of the number of tokens assigned to the agents: below 4 or 5 messages exchanged, agents are likely to consider (0,0) as a relevant alternative, while above 50 messages exchanged, (0,0) certainly is no longer considered relevant.
- Experimental data by Camerer & al. 2003: when the Email Game is repeated a number of times, agents gradually learn to play A after experiencing a loss on unsuccessful play of B.

Summary and conclusion

- TS: logics for introspection, bridging K and K45
- MTS: Literal implementation of the idea of bounded rationality

Perspectives

- Further applications of TS: higher-order vagueness (Egré & Bonnay forthcoming)

Perspectives

- Further applications of TS: higher-order vagueness (Egré & Bonnay forthcoming)
- Still work to do!
- Applications in game theory to work out
- Work in progress on dynamic centered semantics and learning

MTS

Evaluation relative to sequences (w, k) of ordered pairs $k = 0$ if no token is spent, $k = i$ if i spends one token.

i) $M, q(w, k) \vDash_{\mathrm{MTS}} \phi\ [m_a, m_b]$ iff $w \in V(p)$.

(ii) $M, q(w, k) \vDash_{\mathrm{MTS}} \neg\phi\ [m_a, m_b]$ iff $M, q(w, k) \nvDash_{MTS} \phi\ [m_a, m_b]$.

(iii) $M, q(w, k) \vDash_{\mathrm{MTS}} (\phi \wedge \psi)\ [m_a, m_b]$ iff $M, q(w, k) \vDash_{\mathrm{MTS}} \phi$ and $M, q(w, k) \vDash_{\mathrm{MTS}} \psi\ [m_a, m_b]$.

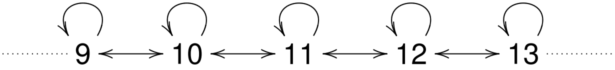
(iv) $M, q(w, k) \vDash_{\mathrm{MTS}} \Box_a\psi\ [m_a, m_b]$ iff
- $m_a \neq 0$ and for all w' such that wR_aw', $M, q(w, k)(w', l) \vDash_{\mathrm{MTS}} \psi\ [m_a - s, m_b]$ where $(l, s) = (1, i)$ for non reflexive moves, $s = l = 0$ otherwise.
- Or $m_a = 0$ and $M, q' \vDash_{\mathrm{MTS}} \Box_a\psi\ [m_a + r_a, m_b + r_b]$ with r_i=number of tokens picked up along the path to reach q' where q' is the longest initial segment of $q(w, k)$ such that (v, i) belongs to $q(w, k)$ but not q

Model-necessitation and CS

The rule of necessitation: if ϕ, then $\Box\phi$

- is standardly valid over frames and over models, namely $M \models \phi$ implies $M \models \Box\phi$ for Kripke semantics.
- is not model-valid relative to CS, although frame-valid

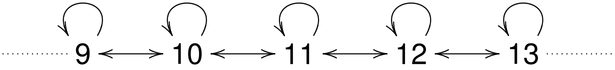
Model validity and CS



- $M \models_{\mathrm{CS}} \Box\neg(i+1) \rightarrow \neg i$ (for $i \in \mathcal{N}$)
- but $M \not\models_{CS} \Box(\Box\neg(i+1) \rightarrow \neg i)$

Example:

Model validity and CS

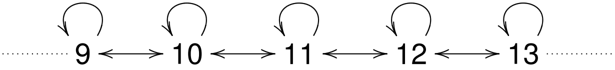


- $M \vDash_{CS} \Box\neg(i+1) \rightarrow \neg i$ (for $i \in \mathcal{N}$)
- but $M \nvDash_{CS} \Box(\Box\neg(i+1) \rightarrow \neg i)$

Example:

- $10 \vDash_{CS} \Box\neg 12 \rightarrow \neg 11$

Model validity and CS

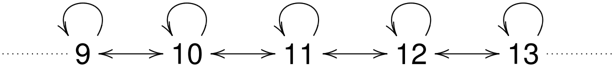


- $M \vDash_{\mathrm{CS}} \Box\neg(i+1) \rightarrow \neg i$ (for $i \in \mathcal{N}$)
- but $M \nvDash_{CS} \Box(\Box\neg(i+1) \rightarrow \neg i)$

Example:

- $10 \vDash_{\mathrm{CS}} \Box\neg 12 \rightarrow \neg 11$
- but $10 \nvDash_{CS} \Box(\Box\neg 12 \rightarrow \neg 11)$

Model validity and CS

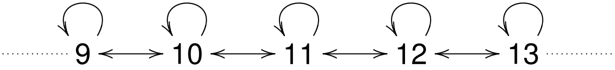


- $M \vDash_{CS} \Box\neg(i+1) \rightarrow \neg i$ (for $i \in \mathcal{N}$)
- but $M \nvDash_{CS} \Box(\Box\neg(i+1) \rightarrow \neg i)$

Example:

- $10 \vDash_{CS} \Box\neg 12 \rightarrow \neg 11$
- but $10 \nvDash_{CS} \Box(\Box\neg 12 \rightarrow \neg 11)$
- because $\Rightarrow 10, 11 \vDash_{CS} \Box\neg 12 \rightarrow \neg 11$

Model validity and CS



- $M \models_{CS} \Box\neg(i+1) \rightarrow \neg i$ (for $i \in \mathcal{N}$)
- but $M \not\models_{CS} \Box(\Box\neg(i+1) \rightarrow \neg i)$

Example:

- $10 \models_{CS} \Box\neg 12 \rightarrow \neg 11$
- but $10 \not\models_{CS} \Box(\Box\neg 12 \rightarrow \neg 11)$
- because $\Rightarrow 10, 11 \models_{CS} \Box\neg 12 \rightarrow \neg 11$
- yet $10, 11 \models_{CS} \Box\neg 12$, but $10, 11 \not\models_{CS} \neg 11$.

Special thanks

Special thanks go to my coauthor Denis Bonnay to whom I am most grateful for an exciting collaboration in the area of epistemic logic since 2005. These two tutorials reflects our joint work and ongoing collaboration in this domain. Thank you Denis!

Further special thanks to my coauthor and colleague Jérôme Dokic with whom this project started. Thank you Jérôme!

Further acknowledgments

- Thanks to Patrick Blackburn, Johan van Benthem, Mikaël Cozic, Olivier Roy, Timothy Williamson for much inspiration, for comments and for stimulating exchanges on these issues over the years.
- Thanks to the Agence Nationale de la Recherche (ANR project "Cognitive Origins of Vagueness", Grant ANR-07-JCJC-0070-01) for support and funding.