# Dynamic Epistemic Logics of Introspection

Raul Fervari[1] and Fernando R. Velázquez-Quesada[2]

[1]FaMAF,UNC and CONICET, Argentina
[2]ILLC, UvA, The Netherlands

Workshop DaLí, Brasília, Brazil
September 2017

# Motivation

- Epistemic Logics deal with the knowledge of agents.
- Not only about propositional facts (ontic knowledge), but with the knowledge about her own and/or other agents' knowledge (high-order knowledge).
- Two important notions in high-order knowledge: positive and negative introspection.
- Positive introspection: *"if the agent knows something, then she knows that she knows it"*.
- Negative introspection: *"if the agent does not know something, then she knows that she does not know it"*.

# Positive and Negative Instrospection

- Positive introspection: *"if the agent knows something, then she knows that she knows it"*.

# Positive and Negative Instrospection

- Positive introspection: *"if the agent knows something, then she knows that she knows it"*.
- In epistemic logic, it is characterized by

$$\Box\varphi \rightarrow \Box\Box\varphi$$

# Positive and Negative Instrospection

- Positive introspection: *"if the agent knows something, then she knows that she knows it"*.
- In epistemic logic, it is characterized by

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

- At semantic level, the accessibility relation is transitive.

# Positive and Negative Instrospection

- Positive introspection: *"if the agent knows something, then she knows that she knows it"*.
- In epistemic logic, it is characterized by

$$\Box\varphi \rightarrow \Box\Box\varphi$$

- At semantic level, the accessibility relation is transitive.
- Negative introspection: *"if the agent does not know something, then she knows that she does not know it"*.

# Positive and Negative Instrospection

- Positive introspection: *"if the agent knows something, then she knows that she knows it"*.
- In epistemic logic, it is characterized by

$$\Box\varphi \rightarrow \Box\Box\varphi$$

- At semantic level, the accessibility relation is transitive.
- Negative introspection: *"if the agent does not know something, then she knows that she does not know it"*.
- In epistemic logic, it is characterized by

$$\neg\Box\varphi \rightarrow \Box\neg\Box\varphi$$

# Positive and Negative Instrospection

- Positive introspection: *"if the agent knows something, then she knows that she knows it".*
- In epistemic logic, it is characterized by

$$\Box\varphi \rightarrow \Box\Box\varphi$$

- At semantic level, the accessibility relation is transitive.
- Negative introspection: *"if the agent does not know something, then she knows that she does not know it".*
- In epistemic logic, it is characterized by

$$\neg\Box\varphi \rightarrow \Box\neg\Box\varphi$$

- At the semantic level, the accessibility relation is Euclidean.

# Not properties, but actions...

- We discussed introspection as properties in the logic and in the underlying structures (relational models).

# Not properties, but actions...

- We discussed introspection as properties in the logic and in the underlying structures (relational models).
- What about considering it as properties to be achieved?

# Not properties, but actions...

- We discussed introspection as properties in the logic and in the underlying structures (relational models).
- What about considering it as properties to be achieved?
- Properties as the eventual result of an action:

# Not properties, but actions...

- We discussed introspection as properties in the logic and in the underlying structures (relational models).
- What about considering it as properties to be achieved?
- Properties as the eventual result of an action:
  - "Syntatic" inference steps based on sets of formulas [Duc17,Ågo&Alech07,Jago09].

# Not properties, but actions...

- We discussed introspection as properties in the logic and in the underlying structures (relational models).
- What about considering it as properties to be achieved?
- Properties as the eventual result of an action:
  - "Syntatic" inference steps based on sets of formulas [Duc17,Ågo&Alech07,Jago09].
  - Inference steps in awareness models [Gross&VQ15].

# Not properties, but actions...

- We discussed introspection as properties in the logic and in the underlying structures (relational models).
- What about considering it as properties to be achieved?
- Properties as the eventual result of an action:
  - "Syntatic" inference steps based on sets of formulas [Duc17,Ågo&Alech07,Jago09].
  - Inference steps in awareness models [Gross&VQ15].
  - Dynamics of evidence, deductive inference [VBen&Pacuit11,VQ13].

# Not properties, but actions...

- We discussed introspection as properties in the logic and in the underlying structures (relational models).
- What about considering it as properties to be achieved?
- Properties as the eventual result of an action:
  - "Syntatic" inference steps based on sets of formulas [Duc17,Ågo&Alech07,Jago09].
  - Inference steps in awareness models [Gross&VQ15].
  - Dynamics of evidence, deductive inference [VBen&Pacuit11,VQ13].
- Dynamic Epistemic Logic style.

# Not properties, but actions...

- We discussed introspection as properties in the logic and in the underlying structures (relational models).
- What about considering it as properties to be achieved?
- Properties as the eventual result of an action:
  - "Syntatic" inference steps based on sets of formulas [Duc17,Ågo&Alech07,Jago09].
  - Inference steps in awareness models [Gross&VQ15].
  - Dynamics of evidence, deductive inference [VBen&Pacuit11,VQ13].
- Dynamic Epistemic Logic style.
- Operations that change the model:
  - Actions for BR, preferences [vBen07,vBen&Liu07,Ghosh&VQ15].

# Not properties, but actions...

- We discussed introspection as properties in the logic and in the underlying structures (relational models).
- What about considering it as properties to be achieved?
- Properties as the eventual result of an action:
  - "Syntatic" inference steps based on sets of formulas [Duc17,Ågo&Alech07,Jago09].
  - Inference steps in awareness models [Gross&VQ15].
  - Dynamics of evidence, deductive inference [VBen&Pacuit11,VQ13].
- Dynamic Epistemic Logic style.
- Operations that change the model:
  - Actions for BR, preferences [vBen07,vBen&Liu07,Ghosh&VQ15].
  - Sabotage logic [vBen05].

# Not properties, but actions...

- We discussed introspection as properties in the logic and in the underlying structures (relational models).
- What about considering it as properties to be achieved?
- Properties as the eventual result of an action:
  - "Syntatic" inference steps based on sets of formulas [Duc17,Ågo&Alech07,Jago09].
  - Inference steps in awareness models [Gross&VQ15].
  - Dynamics of evidence, deductive inference [VBen&Pacuit11,VQ13].
- Dynamic Epistemic Logic style.
- Operations that change the model:
  - Actions for BR, preferences [vBen07,vBen&Liu07,Ghosh&VQ15].
  - Sabotage logic [vBen05].
  - Relation-changing modal logics [AFH12,14,15,Ferv14].

# Not properties, but actions...

- We discussed introspection as properties in the logic and in the underlying structures (relational models).
- What about considering it as properties to be achieved?
- Properties as the eventual result of an action:
  - "Syntatic" inference steps based on sets of formulas [Duc17,Ågo&Alech07,Jago09].
  - Inference steps in awareness models [Gross&VQ15].
  - Dynamics of evidence, deductive inference [VBen&Pacuit11,VQ13].
- Dynamic Epistemic Logic style.
- Operations that change the model:
  - Actions for BR, preferences [vBen07,vBen&Liu07,Ghosh&VQ15].
  - Sabotage logic [vBen05].
  - Relation-changing modal logics [AFH12,14,15,Ferv14].
  - Arrow updates [Kooi&Renne11].

# Not properties, but actions...

- We discussed introspection as properties in the logic and in the underlying structures (relational models).
- What about considering it as properties to be achieved?
- Properties as the eventual result of an action:
  - "Syntatic" inference steps based on sets of formulas [Duc17,Ågo&Alech07,Jago09].
  - Inference steps in awareness models [Gross&VQ15].
  - Dynamics of evidence, deductive inference [VBen&Pacuit11,VQ13].
- Dynamic Epistemic Logic style.
- Operations that change the model:
  - Actions for BR, preferences [vBen07,vBen&Liu07,Ghosh&VQ15].
  - Sabotage logic [vBen05].
  - Relation-changing modal logics [AFH12,14,15,Ferv14].
  - Arrow updates [Kooi&Renne11].
  - More....

# Our work

- We work in a single-agent framework.
- Our approach is provide operations that after being executed, $+/$- introspection is achieved.
- Two kind of operations (both in a global sense):
  - General introspection.
  - Introspection with respect to a formula.
- Two possible approaches:
  - Pessimistic: the agent loses some knowledge.
  - Optimistic: the agent gains some knowledge.

# General Positive Instrospection

- First idea: to make $\Box\varphi \rightarrow \Box\Box\varphi$ true, make the accessibility relation transitive.

# General Positive Instrospection

- First idea: to make $\Box\varphi \to \Box\Box\varphi$ true, make the accessibility relation transitive.
- Let $M = \langle W, R, V \rangle$ be a relational model, $w \in W$:

$$M, w \Vdash \langle + \rangle \varphi \quad \text{iff} \quad M^+, w \Vdash \varphi$$

with $M^+ = \langle W, R^+, V \rangle$, where $R^+$ is the transitive closure of $R$.

# General Positive Instrospection

- First idea: to make $\Box\varphi \to \Box\Box\varphi$ true, make the accessibility relation transitive.
- Let $M = \langle W, R, V \rangle$ be a relational model, $w \in W$:

$$M, w \Vdash \langle + \rangle \varphi \quad \text{iff} \quad M^+, w \Vdash \varphi$$

  with $M^+ = \langle W, R^+, V \rangle$, where $R^+$ is the transitive closure of $R$.
- Self-duality: $[+]\varphi := \neg \langle + \rangle \neg\varphi$, is equivalent to $\langle + \rangle$.

# General Positive Instrospection

- First idea: to make $\Box\varphi \to \Box\Box\varphi$ true, make the accessibility relation transitive.
- Let $M = \langle W, R, V \rangle$ be a relational model, $w \in W$:

$$M, w \Vdash \langle + \rangle \varphi \quad \text{iff} \quad M^+, w \Vdash \varphi$$

  with $M^+ = \langle W, R^+, V \rangle$, where $R^+$ is the transitive closure of $R$.
- Self-duality: $[+]\varphi := \neg \langle + \rangle \neg\varphi$, is equivalent to $\langle + \rangle$.

**Proposition**

$$\Vdash [+] (\Box\varphi \to \Box\Box\varphi).$$

# General Positive Instrospection - some properties

- Complete axiomatization (reduction axioms to $K^+$):

$$\vdash \langle + \rangle \Diamond \varphi \leftrightarrow \oplus \langle + \rangle \varphi$$

# General Positive Instrospection - some properties

- Complete axiomatization (reduction axioms to $K^+$):

$$\vdash \langle + \rangle \Diamond \varphi \leftrightarrow \oplus \langle + \rangle \varphi$$
$$\vdash \langle + \rangle \oplus \varphi \leftrightarrow \oplus \langle + \rangle \varphi$$

# General Positive Instrospection - some properties

- Complete axiomatization (reduction axioms to $K^{+}$):

$$\vdash \langle + \rangle \Diamond \varphi \leftrightarrow \oplus \langle + \rangle \varphi$$
$$\vdash \langle + \rangle \oplus \varphi \leftrightarrow \oplus \langle + \rangle \varphi$$

- However, is not the expected behaviour:

**Fact**

$$\not\vdash \Box p \rightarrow [+] (\Box p \wedge \Box \Box p)$$

# Losing knowledge

**Fact**

$$\not\models \Box p \to [+] (\Box p \wedge \Box \Box p)$$

# Losing knowledge

**Fact**

$$\not\models \Box p \rightarrow [+] (\Box p \wedge \Box \Box p)$$

# Losing knowledge

**Fact**

$$\not\models \Box p \rightarrow [+] (\Box p \wedge \Box \Box p)$$


**Observation**

We are getting introspection at cost of losing knowledge!

# Particular Positive Introspection

- A more intuitive approach: take the agent from knowing $\chi$ without knowing she knows it, to knowing $\chi$ and knowing she knows it.

# Particular Positive Introspection

- A more intuitive approach: take the agent from knowing $\chi$ without knowing she knows it, to knowing $\chi$ and knowing she knows it.
- If at $(M, w)$ the agent knows a given $\chi$ without having full positive introspection about it, then:
   1. every world $R$-reachable from $w$ in one step satisfies $\chi$,

# Particular Positive Introspection

- A more intuitive approach: take the agent from knowing $\chi$ without knowing she knows it, to knowing $\chi$ and knowing she knows it.
- If at $(M, w)$ the agent knows a given $\chi$ without having full positive introspection about it, then:
    1. every world $R$-reachable from $w$ in one step satisfies $\chi$,
    2. there is at least one world $R$-reachable from $w$ (in two or more steps) where $\chi$ fails.

# Particular Positive Introspection

- A more intuitive approach: take the agent from knowing $\chi$ without knowing she knows it, to knowing $\chi$ and knowing she knows it.
- If at $(M, w)$ the agent knows a given $\chi$ without having full positive introspection about it, then:
  1. every world $R$-reachable from $w$ in one step satisfies $\chi$,
  2. there is at least one world $R$-reachable from $w$ (in two or more steps) where $\chi$ fails.
- The operation should make the $\neg\chi$-worlds inaccessible.

# Particular Positive Introspection

- A more intuitive approach: take the agent from knowing $\chi$ without knowing she knows it, to knowing $\chi$ and knowing she knows it.
- If at $(M, w)$ the agent knows a given $\chi$ without having full positive introspection about it, then:
  1. every world $R$-reachable from $w$ in one step satisfies $\chi$,
  2. there is at least one world $R$-reachable from $w$ (in two or more steps) where $\chi$ fails.
- The operation should make the $\neg\chi$-worlds inaccessible.
- We need to remove edges.

# Particular Positive Introspection - disconnecting worlds

- Remove edges pointing to $\neg\chi$-worlds.

# Particular Positive Introspection - disconnecting worlds

- Remove edges pointing to $\neg\chi$-worlds.
- Let $M = \langle W, R, V \rangle$ be a relational model; take $U \subseteq W$.

# Particular Positive Introspection - disconnecting worlds

- Remove edges pointing to $\neg\chi$-worlds.
- Let $M = \langle W, R, V \rangle$ be a relational model; take $U \subseteq W$.
- The $U$-disconnecting operation yields the model $M_{+U} = \langle W, R', V \rangle$, with $R' := R \setminus (U \times \overline{U})$ (for $\overline{U} := W \setminus U$).

# Particular Positive Introspection - disconnecting worlds

- Remove edges pointing to $\neg\chi$-worlds.
- Let $M = \langle W, R, V \rangle$ be a relational model; take $U \subseteq W$.
- The $U$-disconnecting operation yields the model $M_{+U} = \langle W, R', V \rangle$, with $R' := R \setminus (U \times \overline{U})$ (for $\overline{U} := W \setminus U$).
- Thus, this operation removes edges from worlds on $U$ to worlds not in $U$.

# Particular Positive Introspection - disconnecting worlds

- Remove edges pointing to $\neg\chi$-worlds.
- Let $M = \langle W, R, V \rangle$ be a relational model; take $U \subseteq W$.
- The $U$-disconnecting operation yields the model $M_{+U} = \langle W, R', V \rangle$, with $R' := R \setminus (U \times \overline{U})$ (for $\overline{U} := W \setminus U$).
- Thus, this operation removes edges from worlds on $U$ to worlds not in $U$.

$$(M, w) \Vdash \langle +'\chi \rangle \varphi \quad \text{iff} \quad (M_{+[\![\chi]\!]^M}, w) \Vdash \varphi$$

# Particular Positive Introspection - disconnecting worlds

- Remove edges pointing to $\neg\chi$-worlds.
- Let $M = \langle W, R, V\rangle$ be a relational model; take $U \subseteq W$.
- The $U$-disconnecting operation yields the model $M_{+U} = \langle W, R', V\rangle$, with $R' := R \setminus (U \times \overline{U})$ (for $\overline{U} := W \setminus U$).
- Thus, this operation removes edges from worlds on $U$ to worlds not in $U$.

$$(M, w) \Vdash \langle +'\chi\rangle\varphi \quad \text{iff} \quad (M_{+[\![\chi]\!]^M}, w) \Vdash \varphi$$

- But... this operation can take place in any situation! (not only when the agent knows $\chi$)

# Particular Positive Introspection - some properties

- As we pointed out, we get the expected behaviour by including a pre-condition:

$$\langle +\chi \rangle \varphi := \Box \chi \wedge \langle +' \chi \rangle \varphi.$$

# Particular Positive Introspection - some properties

- As we pointed out, we get the expected behaviour by including a pre-condition:

$$\langle +\chi \rangle \varphi := \Box \chi \wedge \langle +' \chi \rangle \varphi.$$

- The relation of the resulting model can be equivalently defined, using PDL notation, as $R := (?\neg\chi; R) \cup (R; ?\chi)$ (easy to axiomatize with reduction axioms).

# Particular Positive Introspection - some properties

- As we pointed out, we get the expected behaviour by including a pre-condition:

$$\langle +\chi \rangle \varphi := \Box \chi \wedge \langle +' \chi \rangle \varphi.$$

- The relation of the resulting model can be equivalently defined, using PDL notation, as $R := (?\neg\chi; R) \cup (R; ?\chi)$ (easy to axiomatize with reduction axioms).

**Proposition**

$$\Vdash \langle +\chi \rangle \Box \varphi \;\leftrightarrow\; \Box (\chi \wedge [+' \chi] \varphi).$$

# Particular Positive Introspection - some properties

- As we pointed out, we get the expected behaviour by including a pre-condition:

$$\langle +\chi \rangle \, \varphi := \Box \, \chi \wedge \langle +' \chi \rangle \, \varphi.$$

- The relation of the resulting model can be equivalently defined, using PDL notation, as $R := (?\neg\chi; R) \cup (R; ?\chi)$ (easy to axiomatize with reduction axioms).

## Proposition

$$\Vdash \langle +\chi \rangle \, \Box \, \varphi \;\leftrightarrow\; \Box \, (\chi \wedge [+' \chi] \, \varphi).$$

*"The agent can perform a particular positive introspection step for $\chi$ after which she will know $\varphi$ iff she knows both $\chi$ and that, after the 'preconditionless' operation, $\varphi$ will be the case."*

# Particular Positive Introspection - some properties

- As we pointed out, we get the expected behaviour by including a pre-condition:

$$\langle +\chi \rangle\, \varphi := \Box\, \chi \wedge \langle +'\chi \rangle\, \varphi.$$

- The relation of the resulting model can be equivalently defined, using PDL notation, as $R := (?\neg\chi; R) \cup (R; ?\chi)$ (easy to axiomatize with reduction axioms).

**Proposition**

$$\Vdash \langle +\chi \rangle\, \Box\, \varphi \;\leftrightarrow\; \Box\, (\chi \wedge [+'\chi]\, \varphi).$$

Replacing $\varphi$ by $\Box\chi$ we get:

$$\Vdash \langle +\chi \rangle\, \Box\, \Box\, \chi \leftrightarrow \Box\, (\chi \wedge [+'\chi]\, \Box\, \chi).$$

# Introspection, Announcements, and Moorean Phenomena

- PAL requires pre-condition $\chi$ to be true, but the introspection operation requires for $\chi$ to be known.

# Introspection, Announcements, and Moorean Phenomena

- PAL requires pre-condition $\chi$ to be true, but the introspection operation requires for $\chi$ to be known.
- Moorean phenomena: in PAL, after being truthfully announced, become false (e.g. $p \wedge \Box\neg p$).

# Introspection, Announcements, and Moorean Phenomena

- PAL requires pre-condition $\chi$ to be true, but the introspection operation requires for $\chi$ to be known.
- Moorean phenomena: in PAL, after being truthfully announced, become false (e.g. $p \wedge \Box\neg p$).
- Here we have:

**Fact**

$$\not\Vdash \Box\chi \to [+'\chi]\Box\chi, \text{ as a consequence } \not\Vdash [+'\chi]\Box\chi.$$

## Introspection, Announcements, and Moorean Phenomena

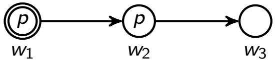
- PAL requires pre-condition $\chi$ to be true, but the introspection operation requires for $\chi$ to be known.
- Moorean phenomena: in PAL, after being truthfully announced, become false (e.g. $p \wedge \Box\neg p$).
- Here we have:

**Fact**

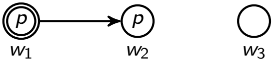
$$\not\Vdash \Box\chi \rightarrow [+'\chi]\Box\chi, \text{ as a consequence } \not\Vdash [+'\chi]\Box\chi.$$

Take $\chi := p \wedge \Diamond\Diamond\neg p$:

$M$ :




$M_{+[\![\chi]\!]^M}$ :



# Introspection, Announcements, and Moorean Phenomena

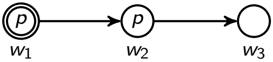
- PAL requires pre-condition $\chi$ to be true, but the introspection operation requires for $\chi$ to be known.
- Moorean phenomena: in PAL, after being truthfully announced, become false (e.g. $p \wedge \Box\neg p$).
- Here we have:

**Fact**

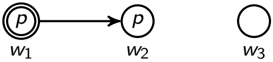
$\not\Vdash \Box\chi \rightarrow [+'\chi]\Box\chi$, as a consequence $\not\Vdash [+'\chi]\Box\chi$.

Take $\chi := p \wedge \Diamond\Diamond\neg p$:

$M$ :




$M_{+[\![\chi]\!]^M}$ :




**Proposition**

If $\Vdash \chi \rightarrow [+'\chi]\chi$, then after the operation the agent will have positive introspection about $\chi$, i.e., $\Vdash \langle +\chi\rangle\Box\Box\chi \leftrightarrow \Box\chi$.

# General Negative Instrospection

- First idea: to make $\neg \Box \varphi \to \Box \neg \Box \varphi$ true, make the accessibility relation Euclidean.

# General Negative Instrospection

- First idea: to make $\neg\Box\varphi \rightarrow \Box\neg\Box\varphi$ true, make the accessibility relation Euclidean.
- Let $M = \langle W, R, V\rangle$ be a relational model, $w \in W$; $M^{-} = \langle W, R^{E}, V\rangle$ in which $R^{E}$ is the Euclidean closure of $R$, that is,

$$R^{E} := R \cup (\reflectbox{$R$} \circ (R \cup \reflectbox{$R$})^{*} \circ R).$$

# General Negative Instrospection

- First idea: to make $\neg\Box\varphi \rightarrow \Box\neg\Box\varphi$ true, make the accessibility relation Euclidean.
- Let $M = \langle W, R, V\rangle$ be a relational model, $w \in W$; $M^{-} = \langle W, R^{E}, V\rangle$ in which $R^{E}$ is the Euclidean closure of $R$, that is,

$$R^{E} := R \cup (Я \circ (R \cup Я)^{*} \circ R).$$

- The associated modality:

$$(M, w) \Vdash \langle - \rangle\,\varphi \quad \text{iff} \quad (M^{-}, w) \Vdash \varphi.$$

# General Negative Instrospection

- First idea: to make $\neg\Box\varphi \rightarrow \Box\neg\Box\varphi$ true, make the accessibility relation Euclidean.
- Let $M = \langle W, R, V\rangle$ be a relational model, $w \in W$; $M^{-} = \langle W, R^{E}, V\rangle$ in which $R^{E}$ is the Euclidean closure of $R$, that is,

$$R^{E} := R \cup (\reflectbox{R} \circ (R \cup \reflectbox{R})^{*} \circ R).$$

- The associated modality:

$$(M, w) \Vdash \langle - \rangle\varphi \quad \text{iff} \quad (M^{-}, w) \Vdash \varphi.$$

## Proposition

$$\Vdash [-](\neg\Box\varphi \rightarrow \Box\neg\Box\varphi).$$

# General Negative Instrospection

- First idea: to make $\neg \Box \varphi \to \Box \neg \Box \varphi$ true, make the accessibility relation Euclidean.
- Let $M = \langle W, R, V \rangle$ be a relational model, $w \in W$; $M^{-} = \langle W, R^E, V \rangle$ in which $R^E$ is the Euclidean closure of $R$, that is,

$$R^E := R \cup (\reflectbox{$R$} \circ (R \cup \reflectbox{$R$})^* \circ R).$$

- The associated modality:

$$(M, w) \Vdash \langle - \rangle \varphi \quad \text{iff} \quad (M^{-}, w) \Vdash \varphi.$$

**Proposition**

$$\Vdash [-] (\neg \Box \varphi \to \Box \neg \Box \varphi).$$

Even more:

**Proposition**

If $\varphi$ is propositional, then $\Vdash \neg \Box \varphi \to [-] (\neg \Box \varphi \land \Box \neg \Box \varphi)$.

# General Negative Instrospection - some properties

- As expected, it also has Moorean behaviour for arbitrary formulas:

# General Negative Instrospection - some properties

- As expected, it also has Moorean behaviour for arbitrary formulas:

**Proposition**

$$\not\Vdash \neg \Box \varphi \rightarrow [-] \Box \neg \Box \varphi.$$

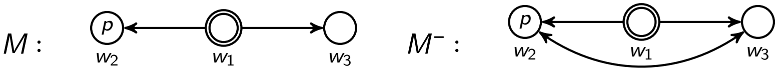
# General Negative Instrospection - some properties

- As expected, it also has Moorean behaviour for arbitrary formulas:

**Proposition**

$$\not\Vdash \neg \Box \varphi \rightarrow [-] \Box \neg \Box \varphi.$$

Consider $\varphi := \neg \Box p$:



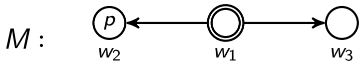
# General Negative Instrospection - some properties

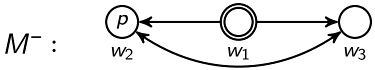
- As expected, it also has Moorean behaviour for arbitrary formulas:

**Proposition**

$$\not\Vdash \neg \Box \varphi \rightarrow [-] \Box \neg \Box \varphi.$$

Consider $\varphi := \neg \Box p$:




- Axiomatizable in PDL (with test and converse).

# Particular Negative Introspection

- For the propositional case, the general operation already behaves as expected.

# Particular Negative Introspection

- For the propositional case, the general operation already behaves as expected.
- However, for uniformity we also define the particular version.

# Particular Negative Introspection

- For the propositional case, the general operation already behaves as expected.
- However, for uniformity we also define the particular version.
- Let $M = \langle W, R, V \rangle$ and $U \subseteq W$. The $U$-connecting operation gives the model $M_{-U} = \langle W, R', V \rangle$ with

$$R' := R \cup (\reflectbox{R} \circ (R \cup \reflectbox{R})^{*} \circ R \circ \mathsf{Id}_U^M),$$

where $\mathsf{Id}_U^M := \{(u, u) \mid u \in U\}$.

# Particular Negative Introspection

- For the propositional case, the general operation already behaves as expected.
- However, for uniformity we also define the particular version.
- Let $M = \langle W, R, V \rangle$ and $U \subseteq W$. The $U$-connecting operation gives the model $M_{-U} = \langle W, R', V \rangle$ with

$$R' := R \cup (\reflectbox{R} \circ (R \cup \reflectbox{R})^* \circ R \circ \mathsf{Id}^M_U),$$

  where $\mathsf{Id}^M_U := \{(u, u) \mid u \in U\}$.
- We define a pre-conditionless operation:

$$(M, w) \Vdash \langle -'\chi \rangle \varphi \quad \text{iff} \quad (M_{-[\![\neg\chi]\!]^M}, w) \Vdash \varphi.$$

# Particular Negative Introspection

- For the propositional case, the general operation already behaves as expected.
- However, for uniformity we also define the particular version.
- Let $M = \langle W, R, V \rangle$ and $U \subseteq W$. The *U*-connecting operation gives the model $M_{-U} = \langle W, R', V \rangle$ with

$$R' := R \cup (\reflectbox{R} \circ (R \cup \reflectbox{R})^* \circ R \circ \mathsf{Id}_U^M),$$

  where $\mathsf{Id}_U^M := \{(u, u) \mid u \in U\}$.
- We define a pre-conditionless operation:

$$(M, w) \Vdash \langle -'\chi \rangle \varphi \quad \text{iff} \quad (M_{-[\![\neg\chi]\!]^M}, w) \Vdash \varphi.$$

- And the corresponding modality:

$$\langle -\chi \rangle \varphi := \neg \Box \chi \wedge \langle -'\chi \rangle \varphi.$$

# Conclusions

- We define operations for achieving positive and negative introspection, both in general and also for a particular formula.

# Conclusions

- We define operations for achieving positive and negative introspection, both in general and also for a particular formula.
- We studied their properties, and obtained complete axiomatizations (via reductions axioms into PDL).

# Conclusions

- We define operations for achieving positive and negative introspection, both in general and also for a particular formula.
- We studied their properties, and obtained complete axiomatizations (via reductions axioms into PDL).
- For positive introspection: one operation adds edges, the other removes edges.

# Conclusions

- We define operations for achieving positive and negative introspection, both in general and also for a particular formula.
- We studied their properties, and obtained complete axiomatizations (via reductions axioms into PDL).
- For positive introspection: one operation adds edges, the other removes edges.
- For negative introspection both operations adds edges.

# Future Work

- Getting introspection in one level (more local): from $\Box p \wedge \neg \Box \Box p$ to $\Box p \wedge \Box \Box p \wedge \neg \Box \Box \Box p$
- We would like to explore this kind of operations in a multi-agent setting.
  - Public, private versions of these operations.
  - Reaching common knowledge.
- Connections with cognitive operations.