Persuasión Abstracta y Modelos de Acción

Carlo Proietti y Antonio Yuste-Ginel[a]

Seminario Lógica y Lenguaje, Sevilla, Febrero 2020

[a]Financiado por FPU2016/041113

Kagemusha

Kagemusha

Japón medieval. Clan Takeda (1). Espías (2). Argumentos relevantes:

Kagemusha

Japón medieval. Clan Takeda (1). Espías (2). Argumentos relevantes:

- a

a: "el jefe está vivo"

Kagemusha

Japón medieval. Clan Takeda (1). Espías (2). Argumentos relevantes:

$\bullet$ b $\longrightarrow$ $\bullet$ a

a: "el jefe está vivo"

b: "ha habido un funeral en el lago"

Kagemusha

Japón medieval. Clan Takeda (1). Espías (2). Argumentos relevantes:

$$\bullet\ \mathrm{c} \longrightarrow \bullet\ \mathrm{b} \longrightarrow \bullet\ \mathrm{a}$$

a: "el jefe está vivo"

b: "ha habido un funeral en el lago"

c: "era un extraño ritual"

Kagemusha

Japón medieval. Clan Takeda (1). Espías (2). Argumentos relevantes:

$$\bullet\ \mathrm{d} \longrightarrow \bullet\ \mathrm{c} \longrightarrow \bullet\ \mathrm{b} \longrightarrow \bullet\ \mathrm{a}$$

a: "el jefe está vivo"

b: "ha habido un funeral en el lago"

c: "era un extraño ritual"

d: "¡eso es mentira!"

Kagemusha

Japón medieval. Clan Takeda (1). Espías (2). Argumentos relevantes:

$$\bullet\, d \longrightarrow \bullet\, c \longrightarrow \bullet\, b \longrightarrow \bullet\, a$$

a: "el jefe está vivo"

b: "ha habido un funeral en el lago"

c: "era un extraño ritual"

d: "¡eso es mentira!"

Preguntas

- ¿Por qué la estrategia del clan funciona?

Preguntas

- ¿Por qué la estrategia del clan funciona?
- ¿Ocurriría lo mismo si la actitud epistémica de los agentes, con respecto al estado argumentativo de su adversario, fuera distinta? Por ejemplo, si los espías supiesen que están siendo espiados.

Preguntas

- ¿Por qué la estrategia del clan funciona?
- ¿Ocurriría lo mismo si la actitud epistémica de los agentes, con respecto al estado argumentativo de su adversario, fuera distinta? Por ejemplo, si los espías supiesen que están siendo espiados.
- ¿Cuáles son las herramientas adecuadas para captar el componente epistémico de la persuasión?

Preguntas

- ¿Por qué la estrategia del clan funciona?
- ¿Ocurriría lo mismo si la actitud epistémica de los agentes, con respecto al estado argumentativo de su adversario, fuera distinta? Por ejemplo, si los espías supiesen que están siendo espiados.
- ¿Cuáles son las herramientas adecuadas para captar el componente epistémico de la persuasión? **Propuesta**: argumentación abstracta + LED con conciencia [¡de argumentos!]

Esbozo de una teoría informal de la persuasión

La persuasión entre individuos (sin emociones) tiene tres componentes fundamentales:

Esbozo de una teoría informal de la persuasión

La persuasión entre individuos (sin emociones) tiene tres componentes fundamentales:

- los objetivos comunicativos de cada participante

Esbozo de una teoría informal de la persuasión

La persuasión entre individuos (sin emociones) tiene tres componentes fundamentales:

- los objetivos comunicativos de cada participante
- las creencias (conocimiento) de cada participante

Esbozo de una teoría informal de la persuasión

La persuasión entre individuos (sin emociones) tiene tres componentes fundamentales:

- los objetivos comunicativos de cada participante
- las creencias (conocimiento) de cada participante
- las políticas de intercambio de información (tanto para el hablante como para el oyente)

Contenidos

Grafos argumentativos y estatus de justificación

Creencias y conocimiento

Políticas de intercambio de información

Persuasion epistémica

Grafos argumentativos y estatus de justificación

DEFINICIÓN (Dung (1995))

Un grafo argumentativo (GA) es un par (A, R) *donde:*

- $A \neq \emptyset$ y es finito (*argumentos*)
- $R \subseteq A \times A$ (*relación de ataque (derrota)*)

DEFINICIÓN (Dung (1995))

Un grafo argumentativo (GA) es un par (A, R) *donde:*

- $A \neq \emptyset$ y es finito (*argumentos*)
- $R \subseteq A \times A$ (*relación de ataque (derrota)*)

aRb se lee como "a ataca b" ("a derrota b")

DEFINICIÓN (Dung (1995))

Un grafo argumentativo (GA) es un par (A, R) *donde:*

- $A \neq \emptyset$ y es finito (*argumentos*)
- $R \subseteq A \times A$ (*relación de ataque (derrota)*)

aRb se lee como "a ataca b" ("a derrota b")

Incluyendo agentes: un **GAD** es una tupla (A, R, A_1, A_2) donde $A_i \subseteq A$ y **el subgrafo de** $i \in \{1, 2\}$ **se define** como (A_i, R_i) donde $R_i = R \cap (A_i \times A_i)$ para cada $i \in \{1, 2\}$ (asunción para simplificar).

DEFINICIÓN (Dung (1995))

Un grafo argumentativo (GA) es un par (A, R) *donde:*

- $A \neq \emptyset$ y es finito (*argumentos*)
- $R \subseteq A \times A$ (*relación de ataque (derrota)*)

aRb se lee como "a ataca b" ("a derrota b")

Incluyendo agentes: un **GAD** es una tupla (A, R, A_1, A_2) donde $A_i \subseteq A$ y **el subgrafo de** $i \in \{1, 2\}$ **se define** como (A_i, R_i) donde $R_i = R \cap (A_i \times A_i)$ para cada $i \in \{1, 2\}$ (asunción para simplificar).

GAD puntuados: (A, R, A_1, A_2, a) donde $a \in A_1 \cap A_2$, se usan para representar escenarios de debate acerca de a.

Ejemplo de GAD

$$\bullet\ d \longrightarrow \bullet\ c \longrightarrow \bullet\ b \longrightarrow \bullet\ a$$

a: “el jefe está vivo”

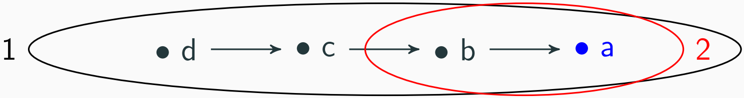
b: “ha habido un funeral en el lago”

c: “era un extraño ritual”

d: “¡eso es mentira!”

Ejemplo de GAD

Takeda (1). Espías (2)



a: "el jefe está vivo"

b: "ha habido un funeral en el lago"

c: "era un extraño ritual"

d: "¡eso es mentira!"

Semántica para GADs

DEFINICIÓN (Caminada and Gabbay (2009))

Semántica para GADs

DEFINICIÓN (Caminada and Gabbay (2009))

Un **etiquetado** *para* (A, R) *es* $\mathcal{L} : A \to \{\mathrm{in}, \mathrm{out}, \mathrm{undec}\}$.

Semántica para GADs

DEFINICIÓN (Caminada and Gabbay (2009))

Un **etiquetado** *para* (A, R) *es* $\mathcal{L} : A \to \{\text{in}, \text{out}, \text{undec}\}$. *Un etiquetado* $\mathcal{L}$ *es* **completo** *sii para todo* $a \in A$:

- $\mathcal{L}(a) = \text{in}$ *sii para todo* $c \in A$ *t.q.* cRa: $\mathcal{L}(c) = \text{out}$

Semántica para GADs

DEFINICIÓN (Caminada and Gabbay (2009))

Un **etiquetado** *para* (A, R) *es* $\mathcal{L} : A \rightarrow \{\mathrm{in}, \mathrm{out}, \mathrm{undec}\}$. *Un etiquetado* $\mathcal{L}$ *es* **completo** *sii para todo* $a \in A$:

- $\mathcal{L}(a) = \mathrm{in}$ *sii para todo* $c \in A$ *t.q.* cRa: $\mathcal{L}(c) = \mathrm{out}$
- $\mathcal{L}(a) = \mathrm{out}$ *sii existe un* $c \in A$ *t.q.* cRa *y* $\mathcal{L}(c) = \mathrm{in}$

Semántica para GADs

DEFINICIÓN (Caminada and Gabbay (2009))

Un **etiquetado** *para* (A, R) *es* $\mathcal{L} : A \to \{\mathrm{in}, \mathrm{out}, \mathrm{undec}\}$. *Un etiquetado* $\mathcal{L}$ *es* **completo** *sii para todo* $a \in A$:

- $\mathcal{L}(a) = \mathrm{in}$ *sii para todo* $c \in A$ *t.q.* cRa: $\mathcal{L}(c) = \mathrm{out}$
- $\mathcal{L}(a) = \mathrm{out}$ *sii existe un* $c \in A$ *t.q.* cRa *y* $\mathcal{L}(c) = \mathrm{in}$

Un etiquetado $\mathcal{L}$ completo se dice preferido sii el conjunto $\mathrm{in}(\mathcal{L}) := \{a \in A \mid \mathcal{L}(a) = \mathrm{in}\}$ es maximal c.r.a. $\subseteq$.

Semántica para GADs

DEFINICIÓN (Caminada and Gabbay (2009))

Un **etiquetado** *para* (A, R) *es* $\mathcal{L} : A \to \{\mathrm{in}, \mathrm{out}, \mathrm{undec}\}$. *Un etiquetado* $\mathcal{L}$ *es* **completo** *sii para todo* $a \in A$*:*

- $\mathcal{L}(a) = \mathrm{in}$ *sii para todo* $c \in A$ *t.q.* cRa*:* $\mathcal{L}(c) = \mathrm{out}$
- $\mathcal{L}(a) = \mathrm{out}$ *sii existe un* $c \in A$ *t.q.* cRa *y* $\mathcal{L}(c) = \mathrm{in}$

Un etiquetado $\mathcal{L}$ completo se dice preferido sii el conjunto $\mathrm{in}(\mathcal{L}) := \{a \in A \mid \mathcal{L}(a) = \mathrm{in}\}$ es maximal c.r.a. $\subseteq$.

Etiquetado preferido≈aceptación en humanos (Cramer and Guillaume, 2018).

Estatus de Justificación

DEFINICIÓN (Basada en Wu and Caminada (2010))

El **estado de justificación de** a **para** i *es el valor devuelto por la función* $\mathcal{JS}_i : A \to \wp(\{\mathrm{in}, \mathrm{out}, \mathrm{undec}\})$ *definidad como:*

$\mathcal{JS}_i(a) := \{\mathcal{L}(a) \mid \mathcal{L}$ *is un etiquetado preferido de* $(A_i, R_i)\}$

Estatus de Justificación

DEFINICIÓN (Basada en Wu and Caminada (2010))

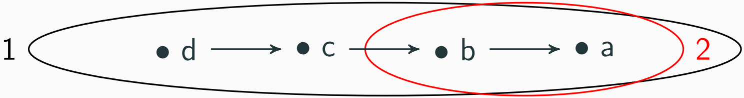
El **estado de justificación de** a **para** i *es el valor devuelto por la función* $\mathcal{JS}_i : A \to \wp(\{\mathrm{in}, \mathrm{out}, \mathrm{undec}\})$ *definidad como:*

$\mathcal{JS}_i(a) := \{\mathcal{L}(a) \mid \mathcal{L}$ *is un etiquetado preferido de* $(A_i, R_i)\}$

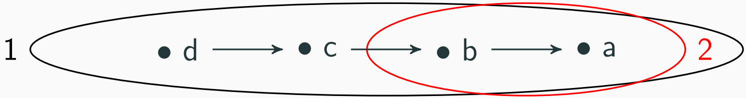
JS* es el rango de $\mathcal{JS}$. Podemos definir naturalmente una jerarquía de aceptación de un argumento:

aceptación plena $\{\mathrm{in}\} \succ \{\mathrm{in}, \mathrm{undec}\} \succ \{\mathrm{undec}\} \simeq \{\mathrm{in}, \mathrm{out}\} \simeq \{\mathrm{in}, \mathrm{out}, \mathrm{undec}\} \succ \{\mathrm{out}, \mathrm{undec}\} \succ \{\mathrm{out}\}$ rechazo pleno

Ejemplo

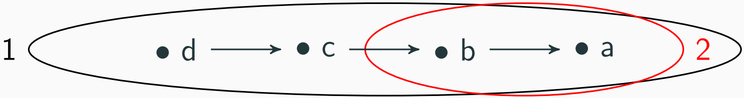


Ejemplo



$$\mathcal{JS}_1(a) = \mathcal{JS}_2(a) = \{\text{out}\}$$

Ejemplo



$$\mathcal{JS}_1(a) = \mathcal{JS}_2(a) = \{\text{out}\}$$
$$\mathcal{JS}_1(b) = \mathcal{JS}_2(b) = \{\text{in}\}$$

Codificando GADs y $\mathcal{JS}$ en lógica proposicional

- Para razonar sistemáticamente sobre estos conceptos necesitamos codificarlos adecuadamente en una lógica

Codificando GADs y $\mathcal{JS}$ en lógica proposicional

- Para razonar sistemáticamente sobre estos conceptos necesitamos codificarlos adecuadamente en una lógica
- Enfoque de la satisfacibilidad (Besnard et al., 2014)

Codificando GADs y $\mathcal{JS}$ en lógica proposicional

- Para razonar sistemáticamente sobre estos conceptos necesitamos codificarlos adecuadamente en una lógica
- Enfoque de la satisfacibilidad (Besnard et al., 2014)

¿Cómo?

Codificando GADs y $\mathcal{JS}$ en lógica proposicional

- Para razonar sistemáticamente sobre estos conceptos necesitamos codificarlos adecuadamente en una lógica
- Enfoque de la satisfacibilidad (Besnard et al., 2014)

¿Cómo? Dado un $\mathrm{GAD} = (A, R, A_1, A_2)$, enumeramos $\wp(A) = \{E_1, ..., E_n\}$ y asociamos a GAD el conjunto de variables $\mathcal{V}^{A}_{\mathrm{Ag}} := \mathcal{A} \cup \mathcal{O} \cup \mathcal{I}$ donde:

Codificando GADs y $\mathcal{JS}$ en lógica proposicional

- Para razonar sistemáticamente sobre estos conceptos necesitamos codificarlos adecuadamente en una lógica
- Enfoque de la satisfacibilidad (Besnard et al., 2014)

¿Cómo? Dado un GAD $= (A, R, A_1, A_2)$, enumeramos $\wp(A) = \{E_1, ..., E_n\}$ y asociamos a GAD el conjunto de variables $\mathcal{V}^A_{\mathsf{Ag}} := \mathcal{A} \cup \mathcal{O} \cup \mathcal{I}$ donde:

$$\mathcal{A} := \{a \rightsquigarrow b \mid (a, b) \in A \times A\}$$

$$\mathcal{O} := \{\mathsf{owns}_i(a) \mid i \in \mathsf{Ag}, a \in A\}$$

$$\mathcal{I} := \{\mathsf{In}^k(a) \mid a \in A, E_k \subseteq A\}$$

Codificando GADs... (continuación)

$\mathcal{L}(\mathcal{V}_{\mathrm{Ag}}^{A})$ es el lenguaje definido por

$$\varphi := p \mid \neg\varphi \mid (\varphi \wedge \varphi) \qquad p \in \mathcal{V}_{\mathrm{Ag}}^{A}$$

Codificando GADs... (continuación)

$\mathcal{L}(\mathcal{V}_{\mathsf{Ag}}^{A})$ es el lenguaje definido por

$$\varphi := p \mid \neg\varphi \mid (\varphi \wedge \varphi) \qquad p \in \mathcal{V}_{\mathsf{Ag}}^{A}$$

Podemos definir abreviaturas $\varphi_{\mathsf{GAD,c}}$ para cada concepto c definido hasta ahora, tal que $\varphi_{\mathsf{GAD,c}}$ es satisfacible sii c se cumple para GAD.

Codificando GADs... (continuación)

$\mathcal{L}(\mathcal{V}^A_{\mathrm{Ag}})$ es el lenguaje definido por

$$\varphi := p \mid \neg\varphi \mid (\varphi \wedge \varphi) \qquad p \in \mathcal{V}^A_{\mathrm{Ag}}$$

Podemos definir abreviaturas $\varphi_{\mathrm{GAD,c}}$ para cada concepto c definido hasta ahora, tal que $\varphi_{\mathrm{GAD,c}}$ es satisfacible sii c se cumple para GAD. Por ejemplo:

$$\begin{aligned}
\mathrm{Th}_{\mathrm{GAD}} := \bigwedge_{(a,b)\in R} a \rightsquigarrow b \wedge \bigwedge_{(a,b)\notin R} \neg(a \rightsquigarrow b) \quad \wedge \\
\bigwedge_{1\leq k\leq n} \Big(\bigwedge_{a\in E_k} \mathrm{In}^k(a) \wedge \bigwedge_{a\notin E_k} \neg\mathrm{In}^k(a)\Big) \quad \wedge \\
\bigwedge_{a\in A_1} \mathrm{owns}_1(a) \wedge \bigwedge_{a\notin A_1} \neg\mathrm{owns}_1(a) \quad \wedge \\
\bigwedge_{a\in A_2} \mathrm{owns}_2(a) \wedge \bigwedge_{a\notin A_2} \neg\mathrm{owns}_2(a)
\end{aligned}$$

Codificando GADs... (continuación)

$\mathcal{L}(\mathcal{V}^A_{\mathrm{Ag}})$ es el lenguaje definido por

$$\varphi := p \mid \neg\varphi \mid (\varphi \wedge \varphi) \qquad p \in \mathcal{V}^A_{\mathrm{Ag}}$$

Podemos definir abreviaturas para cada concepto definido hasta ahora, por ejemplo:

$$E_k \sqsubseteq E_l := \bigwedge_{a \in A} (\mathrm{In}^k(a) \to \mathrm{In}^l(a))$$

$$E_k \sqsubset E_l := E_k \sqsubseteq E_l \wedge \bigvee_{a \in A} (\mathrm{In}^l(a) \wedge \neg \mathrm{In}^k(a))$$

$$\mathrm{conf\_free}_i(E_k) := \bigwedge_{a \in A} \left(\mathrm{In}^k(a) \to \left(\mathrm{owns}_i(a) \wedge \neg \bigvee_{b \in A} (\mathrm{In}^k(b) \wedge b \rightsquigarrow a) \right) \right)$$

Codificando GADs... (continuación)

$\mathcal{L}(\mathcal{V}^A_{\mathsf{Ag}})$ es el lenguaje definido por

$$\varphi := p \mid \neg\varphi \mid (\varphi \wedge \varphi) \qquad p \in \mathcal{V}^A_{\mathsf{Ag}}$$

Podemos definir abreviaturas para cada concepto definido hasta ahora, por ejemplo:

y un largo etcétera

Codificando GADs... (continuación)

Proposición
Sea $\mathrm{GAD} = (A, R, A_1, A_2)$ *un GAD cualquiera:*

Codificando GADs... (continuación)

Proposición
Sea $\mathrm{GAD} = (A, R, A_1, A_2)$ *un GAD cualquiera:*

- $\mathrm{Th}_{\mathrm{GAD}} \wedge \mathrm{conf\_free}_i(E_k)$ es satisfacible sii E_k está libre de conflicto para i.

Codificando GADs... (continuación)

Proposición
Sea $\mathrm{GAD} = (A, R, A_1, A_2)$ *un GAD cualquiera:*

- $\mathrm{Th}_{\mathrm{GAD}} \wedge \mathrm{conf\_free}_i(E_k)$ es satisfacible sii E_k está libre de conflicto para i.
- $\mathrm{Th}_{\mathrm{GAD}} \wedge E_k \sqsubseteq E_l$ es satisfacible sii $E_k \subseteq E_l$

Codificando GADs... (continuación)

Proposición
Sea $\mathrm{GAD} = (A, R, A_1, A_2)$ *un GAD cualquiera:*

- $\mathrm{Th}_{\mathrm{GAD}} \wedge \mathrm{conf\_free}_i(E_k)$ es satisfacible sii E_k está libre de conflicto para i.
- $\mathrm{Th}_{\mathrm{GAD}} \wedge E_k \sqsubseteq E_l$ es satisfacible sii $E_k \subseteq E_l$
  ⋮
- $\mathrm{Th}_{\mathrm{GAD}} \wedge \mathrm{jus}_i(a) = *$ es satisfacible sii $\mathcal{JS}_i(a) = *$ (donde $* \in \mathrm{JS}^*$).

Codificando GADs... (continuación)

Proposición
Sea $\mathrm{GAD} = (A, R, A_1, A_2)$ *un GAD cualquiera:*

- $\mathrm{Th}_{\mathrm{GAD}} \wedge \mathrm{conf\_free}_i(E_k)$ es satisfacible sii E_k está libre de conflicto para i.
- $\mathrm{Th}_{\mathrm{GAD}} \wedge E_k \sqsubseteq E_l$ es satisfacible sii $E_k \subseteq E_l$

  $\vdots$
- $\mathrm{Th}_{\mathrm{GAD}} \wedge \mathrm{jus}_i(a) = * $ es satisfacible sii $\mathcal{JS}_i(a) = *$ (donde $* \in \mathrm{JS}^*$).

Codificando GADs... (continuación)

Proposición
Sea $\mathrm{GAD} = (A, R, A_1, A_2)$ *un GAD cualquiera:*

- $\mathrm{Th}_{\mathrm{GAD}} \wedge \mathrm{conf\_free}_i(E_k)$ es satisfacible sii E_k está libre de conflicto para i.
- $\mathrm{Th}_{\mathrm{GAD}} \wedge E_k \sqsubseteq E_l$ es satisfacible sii $E_k \subseteq E_l$
  ⋮
- $\mathrm{Th}_{\mathrm{GAD}} \wedge \mathrm{jus}_i(a) = *$ es satisfacible sii $\mathcal{JS}_i(a) = *$ (donde $* \in \mathrm{JS}^*$).

¡Podemos expresar todos los conceptos relevantes en LP!

Objetivos comunicacionales

Objetivos comunicacionales

Dado un $\mathrm{GAD} = (A, R, A_1, A_2, a)$ y un agente $i \in \mathrm{Ag}$, un objetivo de comunicación (persuasiva) para i es una expresión de la forma $\mathrm{jus}_j(a) = *$ donde $* \in \mathrm{JS}^*$ y $i \neq j$.

Objetivos comunicacionales

Dado un $\mathrm{GAD} = (A, R, A_1, A_2, a)$ y un agente $i \in \mathrm{Ag}$, un objetivo de comunicación (persuasiva) para i es una expresión de la forma $\mathrm{jus}_j(a) = *$ donde $* \in \mathrm{JS}^*$ y $i \neq j$.

Usamos goal_i como meta-variable para referirnos a un objetivo cualquiera para i.

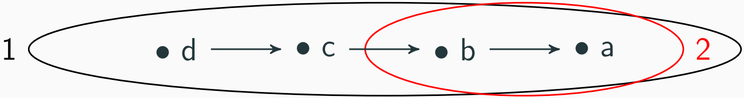
Objetivos comunicacionales

Dado un GAD $= (A, R, A_1, A_2, a)$ y un agente $i \in \mathrm{Ag}$, un objetivo de comunicación (persuasiva) para i es una expresión de la forma $\mathrm{jus}_j(a) = *$ donde $* \in \mathrm{JS}^*$ y $i \neq j$.

Usamos goal_i como meta-variable para referirnos a un objetivo cualquiera para i.

Comunicación persuasiva: cambiar el estado de justificación del oyente de forma que se satisfaga el objetivo del hablante.

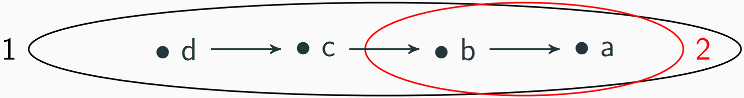
Volviendo a Kagemusha



$$\text{goal}_1 := (\text{jus}_2(a) = \{\text{in}\})$$

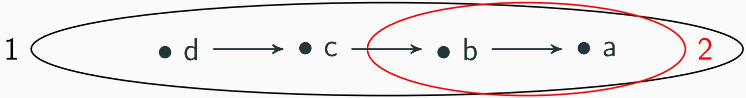
Creencias y conocimiento

Volviendo a Kagemusha



$$\mathrm{goal}_1 := (\mathrm{jus}_2(a) = \{\mathrm{in}\})$$

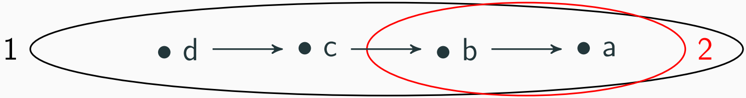
Volviendo a Kagemusha



$$\text{goal}_1 := (\text{jus}_2(a) = \{\text{in}\})$$

¿Por qué el clan decide comunicar c?

Volviendo a Kagemusha



$$\mathrm{goal}_1 := (\mathrm{jus}_2(a) = \{\mathrm{in}\})$$

¿Por qué el clan decide comunicar c? $\Rightarrow$ porque *saben* (creen, y están en lo cierto) que c va a ser persuasivo. Matiz ignorado en trabajos sobre el tema que adoptan una perspectiva de teoría de juegos, e.g. (Rahwan and Larson, 2009; Sakama, 2012).

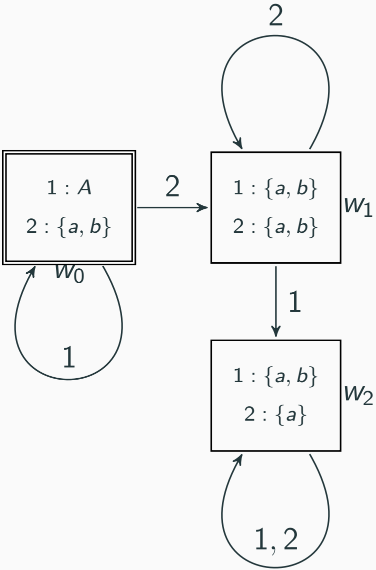


Figura 1: Kagemusha, escenario actual M

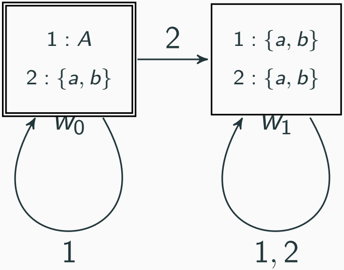


Figura 2: Kagemusha, escenario alternativo M'

EA-modelos (Insp. Schwarzentruber et al. (2012))

Un EA-modelo para $\mathcal{V}^{A}_{\mathsf{Ag}}$ es una terna $M = (W, \mathcal{R}, V)$ donde:

- $W \neq \emptyset$ (*mundos posibles*) donde $w \in W$
- $\mathcal{R} : \mathsf{Ag} \to \wp(W \times W)$ (*relaciones de accesibilidad*)
- $V : \mathcal{V}^{A}_{\mathsf{Ag}} \to \wp(W)$ (*evaluación*).

EA-modelos (Insp. Schwarzentruber et al. (2012))

Un EA-modelo para $\mathcal{V}^{A}_{\mathrm{Ag}}$ es una terna $M = (W, \mathcal{R}, V)$ donde:

- $W \neq \emptyset$ (*mundos posibles*) donde $w \in W$
- $\mathcal{R} : \mathrm{Ag} \to \wp(W \times W)$ (*relaciones de accesibilidad*)
- $V : \mathcal{V}^{A}_{\mathrm{Ag}} \to \wp(W)$ (*evaluación*). Notación $\mathrm{owns}_i(w) := \{a \in A \mid w \in V(\mathrm{owns}_i(a))\}$

EA-modelos (Insp. Schwarzentruber et al. (2012))

Un EA-modelo para $\mathcal{V}^{A}_{\mathrm{Ag}}$ es una terna $M = (W, \mathcal{R}, V)$ donde:

- $W \neq \emptyset$ (*mundos posibles*) donde $w \in W$
- $\mathcal{R} : \mathrm{Ag} \to \wp(W \times W)$ (*relaciones de accesibilidad*)
- $V : \mathcal{V}^{A}_{\mathrm{Ag}} \to \wp(W)$ (*evaluación*). Notación $\mathrm{owns}_i(w) := \{a \in A \mid w \in V(\mathrm{owns}_i(a))\}$

EA-modelos

Además asumimos:

EA-modelos

Además asumimos:

1. $V(p) = W$ o $V(p) = \emptyset$ para todo $p \in \mathcal{A} \cup \mathcal{I}$
(uniformidad de ataques y subconjuntos)

EA-modelos

Además asumimos:

1. $V(p) = W$ o $V(p) = \emptyset$ para todo $p \in \mathcal{A} \cup \mathcal{I}$ (uniformidad de ataques y subconjuntos)
2. Si $w\mathcal{R}_i u$, entonces $\mathsf{owns}_i(w) \subseteq \mathsf{owns}_i(u)$ (introsp. positiva)

EA-modelos

Además asumimos:

1. $V(p) = W$ o $V(p) = \emptyset$ para todo $p \in \mathcal{A} \cup \mathcal{I}$ (uniformidad de ataques y subconjuntos)
2. Si $w\mathcal{R}_i u$, entonces $\text{owns}_i(w) \subseteq \text{owns}_i(u)$ (introsp. positiva)
3. If $w\mathcal{R}_i u$, then $\text{owns}_{\mathbf{j}}(u) \subseteq \text{owns}_i(w)$ (introsp. negativa generalizada)

Lenguaje y axiomas

$\mathcal{L}(\mathcal{V}^A_{\mathrm{Ag}}, \Box)$ se define como

$\varphi ::= p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid \Box_j\varphi \qquad p \in \mathcal{V}^A_{\mathrm{Ag}}, j \in \mathrm{Ag}$

Lenguaje y axiomas

$\mathcal{L}(\mathcal{V}^A_{\mathrm{Ag}}, \Box)$ se define como

$$\varphi ::= p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid \Box_j\varphi \qquad p \in \mathcal{V}^A_{\mathrm{Ag}}, j \in \mathrm{Ag}$$

Axioms			
All propositional tautologies	(Taut)	$\Box_i(\varphi \to \psi) \to (\Box_j\varphi \to \Box_j\psi)$	(K)
$a \rightsquigarrow b \to \Box_{\mathrm{Ag}} a \rightsquigarrow b$	(PKA)	$\neg a \rightsquigarrow b \to \Box_{\mathrm{Ag}} \neg a \rightsquigarrow b$	(NKA)
$\mathrm{In}^k(a) \to \Box_{\mathrm{Ag}} \mathrm{In}^k(a)$	(PKS)	$\neg \mathrm{In}^k(a) \to \Box_{\mathrm{Ag}} \neg \mathrm{In}^k(a)$	(NKS)
$\mathrm{owns}_i(a) \to \Box_i \mathrm{owns}_i(a)$	(PI)	$\neg \mathrm{owns}_i(a) \to \Box_i \neg \mathrm{owns}_j(a)$	(GNI)
Rules			
From $\varphi \to \psi$ and φ, infer ψ	MP	From φ infer $\Box_j\varphi$	$\mathrm{NEC}^{\Box}$

Cuadro 1: Axiomas para el fragmento estático

Theorem
La lógica del cuadro anterior es correcta y fuertemente completa con respecto a la clase de EA modelos. Sus extensiones para creencia (resp. conocimiento) lo son con respecto a las restricciones usuales en las relaciones de accesibilidad.

Theorem
La lógica del cuadro anterior es correcta y fuertemente completa con respecto a la clase de EA modelos. Sus extensiones para creencia (resp. conocimiento) lo son con respecto a las restricciones usuales en las relaciones de accesibilidad.

Comentario: El modelo canónico **no** es un EA-modelo, pero sus submodelos generados **sí** lo son.

Políticas de intercambio de información

Volviendo a Kagemusha

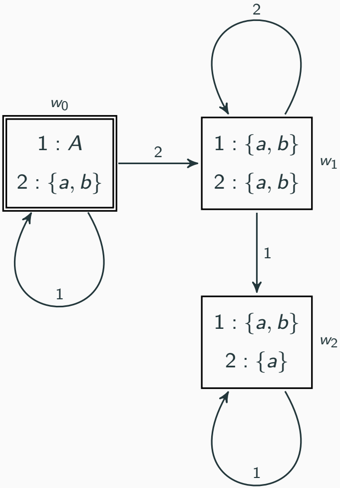


Figura 3: Kagemusha: la estrategia del clan funciona

Volviendo a Kagemusha

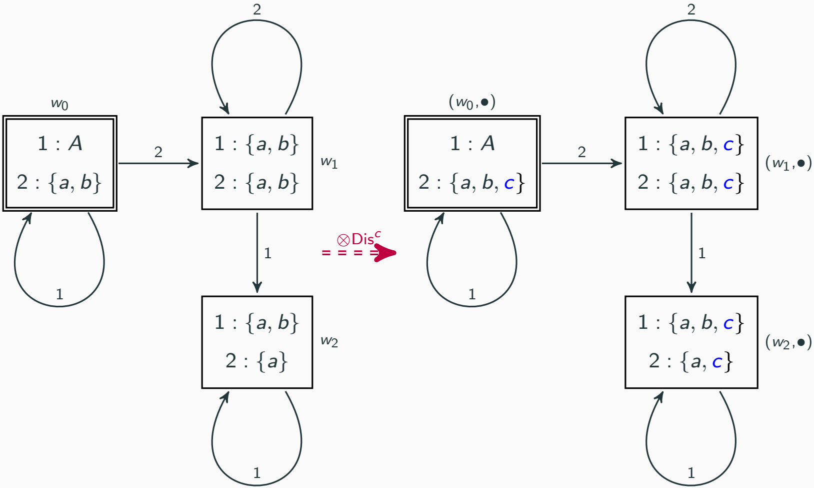


Figura 3: Kagemusha: la estrategia del clan funciona

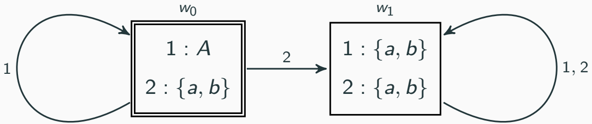


Figura 4: La estrategia del clan fracasa

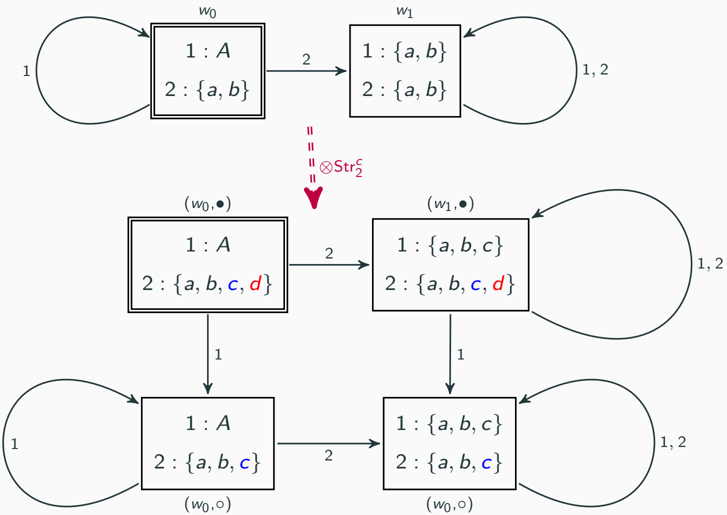


Figura 4: La estrategia del clan fracasa

Sustituciones (Insp. (Van Benthem et al., 2006))

Una *sustitución* para $\mathcal{V}^{A}_{\mathsf{Ag}}$ es $\sigma : \mathcal{V}^{A}_{\mathsf{Ag}} \to \mathcal{V}^{A}_{\mathsf{Ag}} \cup \{\bot, \top\}$ t.q.:

- $\sigma(p) = p$ para todo $p \in \mathcal{A} \cup \mathcal{I}$
- o bien $\sigma(\mathsf{owns}_i(a)) = \mathsf{owns}_i(a)$ o bien $\sigma(\mathsf{owns}_i(a)) = \top$ o bien $\sigma(\mathsf{owns}_i(a)) = \bot$ para cada $\mathsf{owns}_i(a) \in \mathcal{O}$

Sustituciones (Insp. (Van Benthem et al., 2006))

Una *sustitución* para $\mathcal{V}^A_{\mathsf{Ag}}$ es $\sigma : \mathcal{V}^A_{\mathsf{Ag}} \to \mathcal{V}^A_{\mathsf{Ag}} \cup \{\bot, \top\}$ t.q.:

- $\sigma(p) = p$ para todo $p \in \mathcal{A} \cup \mathcal{I}$
- o bien $\sigma(\mathsf{owns}_i(a)) = \mathsf{owns}_i(a)$ o bien $\sigma(\mathsf{owns}_i(a)) = \top$ o bien $\sigma(\mathsf{owns}_i(a)) = \bot$ para cada $\mathsf{owns}_i(a) \in \mathcal{O}$

Representables como $\{p_1 \mapsto *_1, ..., p_n \mapsto *_n\}$ donde $p_m \in \mathcal{O}$,$*_m \in \{\top, \bot\}$ para cada $0 \leq m \leq n$, y $p_m \neq p_k$ para cada $m \neq k$.

Modelos de acción (Insp. (Baltag and Moss, 2004))

Un modelo de acción para $\mathcal{V}^{A}_{\mathrm{Ag}}$ es una tupla $E = (S, \mathcal{T}, \mathrm{pre}, \mathrm{pos})$ donde:

Modelos de acción (Insp. (Baltag and Moss, 2004))

Un modelo de acción para $\mathcal{V}^{A}_{\mathsf{Ag}}$ es una tupla $E = (S, \mathcal{T}, \mathsf{pre}, \mathsf{pos})$ donde:

- $S \neq \emptyset$ y finito (eventos);

Modelos de acción (Insp. (Baltag and Moss, 2004))

Un modelo de acción para $\mathcal{V}_{\mathrm{Ag}}^{A}$ es una tupla
$E = (S, \mathcal{T}, \mathrm{pre}, \mathrm{pos})$ donde:

- $S \neq \emptyset$ y finito (eventos);
- $\mathcal{T} : \mathrm{Ag} \to \wp(S \times S)$ (accesibilidad entre eventos "lo que cada agente percibe de cada evento");

Modelos de acción (Insp. (Baltag and Moss, 2004))

Un modelo de acción para $\mathcal{V}_{\mathsf{Ag}}^{A}$ es una tupla $E = (S, \mathcal{T}, \mathsf{pre}, \mathsf{pos})$ donde:

- $S \neq \emptyset$ y finito (eventos);
- $\mathcal{T} : \mathsf{Ag} \to \wp(S \times S)$ (accesibilidad entre eventos "lo que cada agente percibe de cada evento");
- $\mathsf{pre} : S \to \mathcal{L}(\mathcal{V}_{\mathsf{Ag}}^{A}, \Box)$ (precondiciones);

Modelos de acción (Insp. (Baltag and Moss, 2004))

Un modelo de acción para $\mathcal{V}^{A}_{\mathrm{Ag}}$ es una tupla $E = (S, \mathcal{T}, \mathrm{pre}, \mathrm{pos})$ donde:

- $S \neq \emptyset$ y finito (eventos);
- $\mathcal{T} : \mathrm{Ag} \to \wp(S \times S)$ (accesibilidad entre eventos "lo que cada agente percibe de cada evento");
- $\mathrm{pre} : S \to \mathcal{L}(\mathcal{V}^{A}_{\mathrm{Ag}}, \Box)$ (precondiciones);
- $\mathrm{pos} : S \to \mathrm{SUB}$ (postcondiciones)

Modelos de acción (Insp. (Baltag and Moss, 2004))

Un modelo de acción para $\mathcal{V}^{A}_{\mathsf{Ag}}$ es una tupla
$E = (S, \mathcal{T}, \mathsf{pre}, \mathsf{pos})$ donde:

- $S \neq \emptyset$ y finito (eventos);
- $\mathcal{T} : \mathsf{Ag} \to \wp(S \times S)$ (accesibilidad entre eventos "lo que cada agente percibe de cada evento");
- $\mathsf{pre} : S \to \mathcal{L}(\mathcal{V}^{A}_{\mathsf{Ag}}, \Box)$ (precondiciones);
- $\mathsf{pos} : S \to \mathsf{SUB}$ (postcondiciones)

Modelo de acción puntuado (E, s) donde $s \in E[S]$.

Modelos de acción (Insp. (Baltag and Moss, 2004))

Un modelo de acción para $\mathcal{V}_{\mathsf{Ag}}^{A}$ es una tupla $E = (S, \mathcal{T}, \mathsf{pre}, \mathsf{pos})$ donde:

- $S \neq \emptyset$ y finito (eventos);
- $\mathcal{T} : \mathsf{Ag} \to \wp(S \times S)$ (accesibilidad entre eventos "lo que cada agente percibe de cada evento");
- $\mathsf{pre} : S \to \mathcal{L}(\mathcal{V}_{\mathsf{Ag}}^{A}, \Box)$ (precondiciones);
- $\mathsf{pos} : S \to \mathsf{SUB}$ (postcondiciones)

Modelo de acción puntuado (E, s) donde $s \in E[S]$.

Todos los modelos de acción act

Ejecución de un MA

Dados $M = (W, \mathcal{R}, V)$ y $E = (S, \mathcal{T}, \mathrm{pre}, \mathrm{pos})$ definimos $M \otimes E := (W', \mathcal{R}', V')$ donde:

Ejecución de un MA

Dados $M = (W, \mathcal{R}, V)$ y $E = (S, \mathcal{T}, \text{pre}, \text{pos})$ definimos $M \otimes E := (W', \mathcal{R}', V')$ donde:

- $W' := \{(w, s) \mid M, w \vDash \text{pre}(s)\}$

Ejecución de un MA

Dados $M = (W, \mathcal{R}, V)$ y $E = (S, \mathcal{T}, \text{pre}, \text{pos})$ definimos $M \otimes E := (W', \mathcal{R}', V')$ donde:

- $W' := \{(w, s) \mid M, w \vDash \text{pre}(s)\}$
- $(w, s)\mathcal{R}'_i(w', s')$ sii $w\mathcal{R}_i w'$ y $s\mathcal{T}_i s'$

Ejecución de un MA

Dados $M = (W, \mathcal{R}, V)$ y $E = (S, \mathcal{T}, \mathsf{pre}, \mathsf{pos})$ definimos $M \otimes E := (W', \mathcal{R}', V')$ donde:

- $W' := \{(w, s) \mid M, w \vDash \mathsf{pre}(s)\}$
- $(w, s)\mathcal{R}'_i(w', s')$ sii $w\mathcal{R}_i w'$ y $s\mathcal{T}_i s'$
- $V'(p) := \{(w, s) \in W' \mid M, w \vDash \mathsf{pos}(s)(p)\}$

Lenguaje y Semántica Dinámicos

$\mathcal{L}^{\text{act}}(\mathcal{V}^{A}_{\mathrm{Ag}}, \Box)$ definido por

$$\varphi ::= p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid \Box_j \varphi \mid [E, s]\varphi$$

$$p \in \mathcal{V}^{A}_{\mathrm{Ag}}, j \in \mathrm{Ag}, E \in \text{act}, s \in E[S]$$

Lenguaje y Semántica Dinámicos

$\mathcal{L}^{\mathsf{act}}(\mathcal{V}^{A}_{\mathrm{Ag}}, \Box)$ definido por

$$\varphi ::= p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid \Box_j \varphi \mid [E, s]\varphi$$

$$p \in \mathcal{V}^{A}_{\mathrm{Ag}}, j \in \mathrm{Ag}, E \in \mathsf{act}, s \in E[S]$$

$$M, w \models [E, s]\varphi \quad \text{sii} \quad M, w \models \mathrm{pre}(s) \quad \text{implica} \qquad M \otimes E, (w, s) \models \varphi$$

Anuncio público y comunicación pública

Pub^a

$$\begin{tikzcd} \bullet \arrow[loop above, "i,j"] \end{tikzcd}$$

$$\mathrm{pre}(\bullet) := \top$$

$$\mathrm{pos}(\bullet) := \{\mathrm{owns}_j(a) \mapsto \top\}_{j \in \mathrm{Ag}}$$

Anuncio público y comunicación pública

Pub^a

i, j

$\bullet$

$$\mathrm{pre}(\bullet) := \top$$

$$\mathrm{pos}(\bullet) := \{\mathrm{owns}_j(a) \mapsto \top\}_{j \in \mathrm{Ag}}$$

$$[\mathrm{Dis}_i^a]\varphi := (\neg \mathrm{owns}_i(a) \to \varphi) \wedge (\mathrm{owns}_i(a) \to [\mathrm{Pub}^a, \bullet]\varphi)$$

Aprendizaje privado y actualización escéptica

Pri_i^a

$$\mathrm{pre}(\bullet) = \mathrm{pre}(\circ) = \top$$

$$\mathrm{pos}(\bullet) = \{\mathrm{owns}_i(a) \mapsto \top\}$$

$$\begin{tikzcd}
\bullet \arrow[loop above, "i"] \arrow[r, "j"] & \circ \arrow[loop above, "{i,j}"]
\end{tikzcd}$$

$$\mathrm{pos}(\circ) = \epsilon$$

Aprendizaje privado y actualización escéptica

Pri_i^a

$\mathrm{pre}(\bullet) = \mathrm{pre}(\circ) = \top$

$\mathrm{pos}(\bullet) = \{\mathrm{owns}_i(a) \mapsto \top\}$

$$\begin{tikzcd} \bullet \arrow[loop above, "i"] \arrow[rr, "j"] & & \circ \arrow[loop above, "i,j"] \end{tikzcd}$$

$\mathrm{pos}(\circ) = \epsilon$

$$[\mathsf{Scp}_j^a]\varphi := [\mathsf{Pub}^{a!}, \bullet]\Big(\bigwedge_{b \in A} (b \rightsquigarrow a \rightarrow [\mathsf{Pri}_j^b, \bullet]\varphi)\Big)$$

Aprendizaje privado y actualización escéptica

Pri_i^a

$$\mathrm{pre}(\bullet) = \mathrm{pre}(\circ) = \top$$

$$\mathrm{pos}(\bullet) = \{\mathrm{owns}_i(a) \mapsto \top\}$$

$$\begin{tikzcd}
\bullet \arrow[loop above, "i"] \arrow[rr, "j"] & & \circ \arrow[loop above, "{i,j}"]
\end{tikzcd}$$

$$\mathrm{pos}(\circ) = \epsilon$$

$$[\mathsf{Scp}_j^a]\varphi := [\mathsf{Pub}^{a!}, \bullet]\Big(\bigwedge_{b \in A} \big(b \rightsquigarrow a \to [\mathsf{Pri}_j^b, \bullet]\varphi\big)\Big)$$

$$[\mathsf{Str}_j^a]\varphi := (\neg\psi \to [\mathsf{Pub}^{a!}, \bullet]\varphi) \land (\psi \to [\mathsf{Scp}_j^a]\varphi))$$

donde $\psi := \Box_j \Box_i(\neg \mathrm{goal}_i \land [\mathsf{Pub}^a, \bullet]\mathrm{goal}_i)$

Persuasion epistémica

Persuasión y persuasión epistémica I

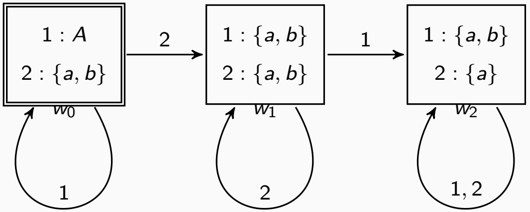
Sea $\mathrm{GAD} = (A, R, A_1, A_2)$, sea goal_1 un objetivo comunicativo y sea (M, w) un EA-modelo para GAD.

$X \subseteq A$ es abiertamente persuasivo sii $M, w \vDash [\mathrm{Dis}_1^X]\mathrm{goal}_1$.
X es (abiertamente) epistémicamente persuasivo sii $M, w \vDash \Box_1[\mathrm{Dis}_1^X]\mathrm{goal}_1$.

Persuasión y persuasión epistémica I

Sea GAD $= (A, R, A_1, A_2)$, sea goal_1 un objetivo comunicativo y sea (M, w) un EA-modelo para GAD.

$X \subseteq A$ es abiertamente persuasivo sii $M, w \vDash [\mathrm{Dis}_1^X]\mathrm{goal}_1$.
X es (abiertamente) epistémicamente persuasivo sii
$M, w \vDash \Box_1[\mathrm{Dis}_1^X]\mathrm{goal}_1$.



Persuasión y persuasión epistémica I

Sea $\mathrm{GAD} = (A, R, A_1, A_2)$, sea goal_1 un objetivo comunicativo y sea (M, w) un EA-modelo para GAD.

$X \subseteq A$ es abiertamente persuasivo sii $M, w \models [\mathrm{Dis}_1^X]\mathrm{goal}_1$.
X es (abiertamente) epistémicamente persuasivo sii $M, w \models \Box_1[\mathrm{Dis}_1^X]\mathrm{goal}_1$.

$$
\begin{tikzcd}
\boxed{\begin{array}{c} 1 : A \\ 2 : \{a, b\} \end{array}}_{w_0} \arrow[r, "2"] \arrow[loop below, "1"] & \boxed{\begin{array}{c} 1 : \{a, b\} \\ 2 : \{a, b\} \end{array}}_{w_1} \arrow[r, "1"] \arrow[loop below, "2"] & \boxed{\begin{array}{c} 1 : \{a, b\} \\ 2 : \{a\} \end{array}}_{w_2} \arrow[loop below, "{1, 2}"]
\end{tikzcd}
$$

$\{c\}$ es persuasivo y persuasivo desde la perspectiva de 1 en M, w_0

Persuasión y persuasión epistémica II

Sea $\mathrm{GAD} = (A, R, A_1, A_2)$, sea goal_1 un objetivo comunicativo y sea (M, w) un EA-modelo para GAD.

$x \in A_1$ es (estratégicamente) persuasivo sii
$M, w \vDash [\mathrm{Str}_2^x]\mathrm{goal}_1$.
x es (estratégicamente) epistémicamente persuasivo sii
$M, w \vDash [\mathrm{Str}_2^x]\mathrm{goal}_1$.

Persuasión y persuasión epistémica II

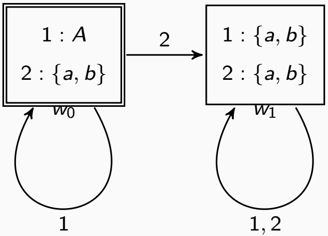
Sea $\mathrm{GAD} = (A, R, A_1, A_2)$, sea goal_1 un objetivo comunicativo y sea (M, w) un EA-modelo para GAD.

$x \in A_1$ es (estratégicamente) persuasivo sii

$M, w \models [\mathrm{Str}_2^x]\mathrm{goal}_1$.

x es (estratégicamente) epistémicamente persuasivo sii

$M, w \models [\mathrm{Str}_2^x]\mathrm{goal}_1$.



c no es estratégicamente persuasivo ni estratégicamente persuasivo desde la perspectiva de 1

Trabajo en marcha y retos pendientes

Trabajo en marcha y retos pendientes

- Axiomatización de las extensiones dinámicas.

Trabajo en marcha y retos pendientes

- Axiomatización de las extensiones dinámicas.
- Caracterizar la clase de *sceptical updates*

Trabajo en marcha y retos pendientes

- Axiomatización de las extensiones dinámicas.
- Caracterizar la clase de *sceptical updates*
- Estudiar la efectividad de las *sceptical updates*

Trabajo en marcha y retos pendientes

- Axiomatización de las extensiones dinámicas.
- Caracterizar la clase de *sceptical updates*
- Estudiar la efectividad de las *sceptical updates*
- Caracterizar la clase de actitudes doxásticas mixtas.

¡Gracias por vuestra atención!

Referencias

Alexandru Baltag and Lawrence S Moss. Logics for epistemic programs. *Synthese*, 139(2):165–224, 2004.

Philippe Besnard, Sylvie Doutre, and Andreas Herzig. Encoding argument graphs in logic. In A. Laurent, O. Strauss, B. Bouchon-Meunier, and R.R. Yager, editors, *International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*, pages 345–354. Springer, 2014.

References ii

Martin WA Caminada and Dov M Gabbay. A logical account of formal argumentation. *Studia Logica*, 93(2-3):109–145, 2009.

Marcos Cramer and Mathieu Guillaume. Empirical cognitive study on abstract argumentation semantics. In S. Modgil, K. Budzynska, and J. Lawrence, editors, *Computational Models of Argument. Proceedings of COMMA 2018*, volume 305, pages 413 – 424. IOS Press, 2018. doi: 10.3233/978-1-61499-906-5-413.

References iii

Phan Minh Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77 (2):321–357, 1995.

Iyad Rahwan and Kate Larson. Argumentation and game theory. In *Argumentation in Artificial Intelligence*, pages 321–339. Springer, 2009.

Chiaki Sakama. Dishonest arguments in debate games. *Computational Models of Argument. Proceedings of COMMA 2012*, pages 177–184, 2012.

References iv

François Schwarzentruber, Srdjan Vesic, and Tjitze Rienstra. Building an epistemic logic for argumentation. In *Logics in Artificial Intelligence*, pages 359–371. Springer, 2012.

Johan Van Benthem, Jan Van Eijck, and Barteld Kooi. Logics of communication and change. *Information and computation*, 204(11):1620–1662, 2006.

Yining Wu and Martin Caminada. A labelling-based justification status of arguments. *Studies in Logic*, 3(4): 12–29, 2010.