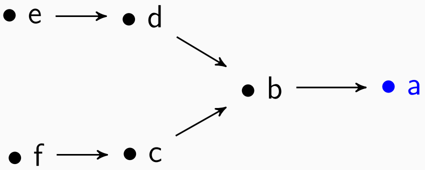
Persuasive Argumentation and Epistemic Attitudes

Carlo Proietti and Antonio Yuste-Ginel

2º DaLí Workshop, Porto, October 2019

A bloody crime

Interrogation room. 1 is the main suspect. 2 is the detective.
These are the only relevant pieces of information:



a: "1 is innocent"

b: "1 was seen close to the crime scene"

c: "1 has a twin brother living in the city"

d: "1 works in a butcher's nearby"

e: "1 was fired from the butcher's a week ago"

f: "1's twin brother was in Venice last night"

Aim of the work

- Which arguments will be disclosed by agent 1 if he tries to persuade 2 that he is innocent?

Aim of the work

- Which arguments will be disclosed by agent 1 if he tries to persuade 2 that he is innocent? **Thesis**. It strongly depends on what 1 thinks that 2 thinks about the relevant information

Aim of the work

- Which arguments will be disclosed by agent 1 if he tries to persuade 2 that he is innocent? **Thesis**. It strongly depends on what 1 thinks that 2 thinks about the relevant information
- Previous work (e.g.(Rahwan and Larson, 2009; Sakama, 2012)) have ignored this fact

Aim of the work

- Which arguments will be disclosed by agent 1 if he tries to persuade 2 that he is innocent? **Thesis**. It strongly depends on what 1 thinks that 2 thinks about the relevant information
- Previous work (e.g.(Rahwan and Larson, 2009; Sakama, 2012)) have ignored this fact
- What are the appropriate tools for capturing the epistemic component of persuasion?

Aim of the work

- Which arguments will be disclosed by agent 1 if he tries to persuade 2 that he is innocent? **Thesis**. It strongly depends on what 1 thinks that 2 thinks about the relevant information
- Previous work (e.g.(Rahwan and Larson, 2009; Sakama, 2012)) have ignored this fact
- What are the appropriate tools for capturing the epistemic component of persuasion? **Proposal**: abstract argumentation + awareness DEL

Outline

AFs and Justification Status

Persuasiveness and Epistemic Persuasiveness

New Advances and Future Work

AFs and Justification Status

Definition (Dung (1995))

An argumentation framework (AF) is a pair $(A, \rightsquigarrow)$ where:

- $A \neq \emptyset$ and finite (*arguments*)
- $\rightsquigarrow \subseteq A \times A$ (*attack relation*)

Definition (Dung (1995))

An argumentation framework (AF) is a pair $(A, \rightsquigarrow)$ where:

- $A \neq \emptyset$ and finite (*arguments*)
- $\rightsquigarrow \subseteq A \times A$ (*attack relation*)

$a \rightsquigarrow b$ is read as "a attacks b"

Definition (Dung (1995))

An argumentation framework (AF) is a pair $(A, \rightsquigarrow)$ where:

- $A \neq \emptyset$ and finite (*arguments*)
- $\rightsquigarrow \subseteq A \times A$ (*attack relation*)

$a \rightsquigarrow b$ is read as "*a* attacks *b*"

Including agents: a **TAF** is a tuple $(A, \rightsquigarrow, A_1, A_2)$ where $A_i \subseteq A$ and $i's$ **subgraph is defined** as $(A_i, \rightsquigarrow_i)$ with $\rightsquigarrow_i = \rightsquigarrow \cap (A_i \times A_i)$ for every $i \in \{1, 2\}$.

Definition (Dung (1995))

An argumentation framework (AF) is a pair $(A, \rightsquigarrow)$ where:

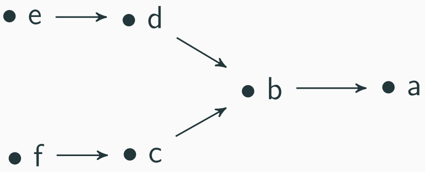
- $A \neq \emptyset$ and finite (*arguments*)
- $\rightsquigarrow \subseteq A \times A$ (*attack relation*)

$a \rightsquigarrow b$ is read as "a attacks b"

Including agents: a **TAF** is a tuple $(A, \rightsquigarrow, A_1, A_2)$ where $A_i \subseteq A$ and $i's$ **subgraph is defined** as $(A_i, \rightsquigarrow_i)$ with $\rightsquigarrow_i = \rightsquigarrow \cap (A_i \times A_i)$ for every $i \in \{1, 2\}$.

Pointed TAFs $(A, \rightsquigarrow, A_1, A_2, a)$ where $a \in A_1 \cap A_2$ are used to represent debate scenarios about a.

Example of TAF



a: “1 is innocent”

b: “1 was seen close to the crime scene”

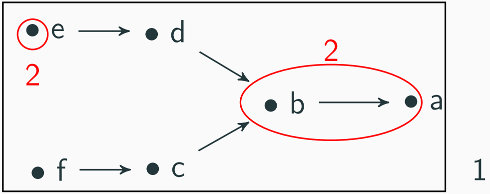
c: “1 has a twin brother living in the city”

f: “1’s twin brother was in Venice last night”

d: “1 works in a butcher’s nearby”

e: “1 was fired from the butcher’s a week ago”

Example of TAF



a: "1 is innocent"

b: "1 was seen close to the crime scene"

c: "1 has a twin brother living in the city"

f: "1's twin brother was in Venice last night"

d: "1 works in a butcher's nearby"

e: "1 was fired from the butcher's a week ago"

Justification Status

Definition (based on Wu et al. (2010))

The **justification status of a for i** is the outcome yielded by the function $\mathcal{JS}_i : A \to \wp(\{\text{in}, \text{out}, \text{undec}\})$ defined as:

$$\mathcal{JS}_i(a) := \{\mathcal{L}_i(a) \mid \mathcal{L} \quad \text{is a complete labelling of} \quad (A_i, \rightsquigarrow_i)\}$$

Justification Status

Definition (based on Wu et al. (2010))

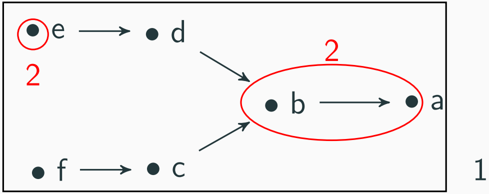
The **justification status of** a **for** i is the outcome yielded by the function $\mathcal{JS}_i : A \to \wp(\{\text{in}, \text{out}, \text{undec}\})$ defined as:

$$\mathcal{JS}_i(a) := \{\mathcal{L}_i(a) \mid \mathcal{L} \quad \text{is a complete labelling of} \quad (A_i, \rightsquigarrow_i)\}$$

JS^* is the set of possible outcomes of $\mathcal{JS}$, which naturally defines an acceptance hierarchy:

$$\text{strong acceptance } \{\text{in}\} > \{\text{in}, \text{undec}\} > \{\text{undec}\} = \{\text{in}, \text{out}, \text{undec}\} > \{\text{out}, \text{undec}\} > \{\text{out}\} \text{strong rejection}$$

Examples



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

Persuasiveness and Epistemic Persuasiveness

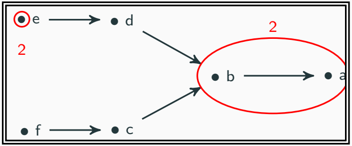
Definition (Persuasiveness of a set of arguments)

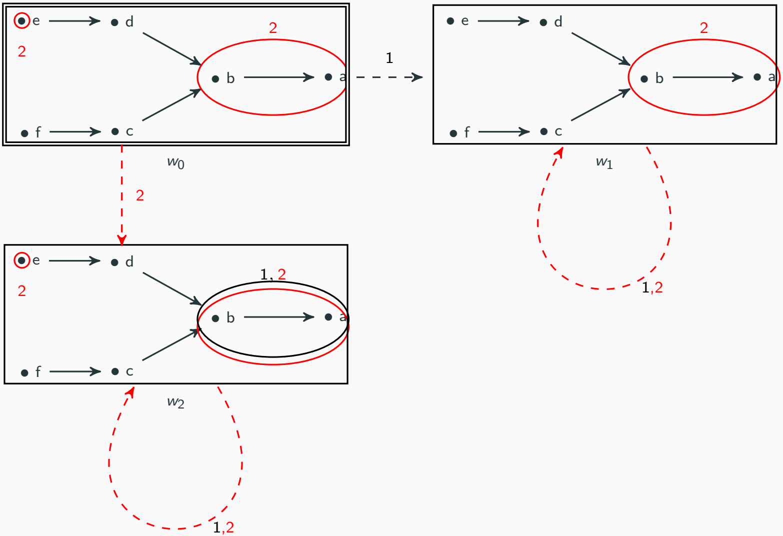
Let $\mathcal{G} = (A, \rightsquigarrow, A_1, A_2, a)$ and $B \subseteq A_1$ the *resulting pointed TAF* is $\mathcal{G}^B := (A, \rightsquigarrow, A_1, A_2^B, a)$ where $A_2^B = A_2 \cup B$. Let goal $\in$ JS*, B is said to be **persuasive** iff $\mathcal{JS}_2(a) =$ goal w.r.t. $\mathcal{G}^B$

Definition (Persuasiveness of a set of arguments)

Let $\mathcal{G} = (A, \rightsquigarrow, A_1, A_2, a)$ and $B \subseteq A_1$ the *resulting pointed TAF* is $\mathcal{G}^B := (A, \rightsquigarrow, A_1, A_2^B, a)$ where $A_2^B = A_2 \cup B$. Let goal $\in$ JS*, B is said to be **persuasive** iff $\mathcal{JS}_2(a) = \text{goal}$ w.r.t. $\mathcal{G}^B$

Persuasion is understood as a change in the hearer's justification status that matches the speaker's intentions.


e
d
2
2
b
a
f
c


e
d
2
2
b
a
1
f
c
w_0
2
e
d
b
a
f
c
w_1
1,2
e
d
2
1, 2
b
a
f
c
w_2
1,2

Epistemic argumentative models

Definition (Schwarzentruber et al. (2012))

An pointed model for $(A, \rightsquigarrow, A_1, A_2)$ is $(M, w) = ((W, \mathcal{R}, \mathcal{D}), w)$ where:

- $W \neq \emptyset$ (*possible worlds*) with $w \in W$
- $\mathcal{R} : \mathrm{Ag} \rightarrow \wp(W \times W)$ (*accessibility relations*)
- $\mathcal{D} : (\mathrm{Ag} \times W) \rightarrow \wp(A)$ (*awareness function*) s.t. $\mathcal{D}_1(w) = A_1$ and $\mathcal{D}_2(w) = A_2$.

Epistemic argumentative models

Definition (Schwarzentruber et al. (2012))

An pointed model for $(A, \rightsquigarrow, A_1, A_2)$ is $(M, w) = ((W, \mathcal{R}, \mathcal{D}), w)$ where:

- $W \neq \emptyset$ (*possible worlds*) with $w \in W$
- $\mathcal{R} : \mathrm{Ag} \rightarrow \wp(W \times W)$ (*accessibility relations*)
- $\mathcal{D} : (\mathrm{Ag} \times W) \rightarrow \wp(A)$ (*awareness function*) s.t. $\mathcal{D}_1(w) = A_1$ and $\mathcal{D}_2(w) = A_2$.

1. If $w\mathcal{R}_i u$, then $\mathcal{D}_i(w) \subseteq \mathcal{D}_i(u)$ (Positive Introspection)
2. If $w\mathcal{R}_i u$, then $\mathcal{D}_{\mathbf{j}}(u) \subseteq \mathcal{D}_i(w)$ (General Negative Introspection)

Definition (Communication Model)

A *communication pointed model* $(M, w)^{+b} := ((W, \mathcal{R}, \mathcal{D}^{+b}), w)$ where $\mathcal{D}^{+b} : (\text{Ag} \times W) \to \wp(A)$ is defined by cases for each $i \in \text{Ag}$ and each $v \in W$ as follows:

$$\begin{array}{ll} \mathcal{D}_i(v) \cup \{b\} & \text{if } b \in \mathcal{D}_1(w) \\ \mathcal{D}_i(v) & \text{otherwise} \end{array}$$

Examples

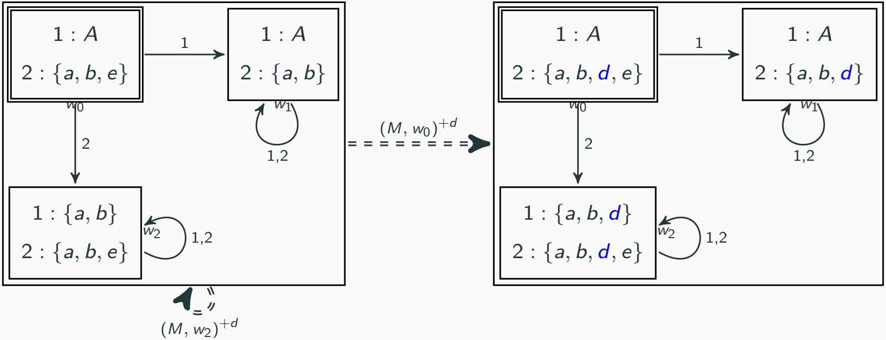


Figure 1: Communication Model

Epistemic Persuasiveness

Definition (Epistemic-based persuasive arguments)

Let (M, w) be a pointed model for $(A, \rightsquigarrow, \mathcal{D}_1(w), \mathcal{D}_2(w), a)$, let goal $\in$ JS*, we say that $\mathbf{B} \subseteq \mathcal{D}_1(w)$ is **persuasive from 1's perspective** iff $\mathcal{JS}_2(a) =$ goal w.r.t. $(\mathcal{D}_2^{+B}(w'), \rightsquigarrow\restriction \mathcal{D}_2^{+B}(w'))$ for all $w' \in W$ s.t. $w\mathcal{R}_1 w'$.

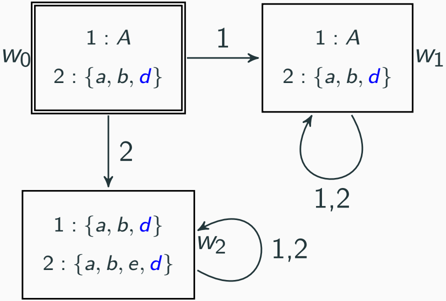
Epistemic Persuasiveness

Definition (Epistemic-based persuasive arguments)

Let (M, w) be a pointed model for $(A, \rightsquigarrow, \mathcal{D}_1(w), \mathcal{D}_2(w), a)$, let goal $\in$ JS*, we say that $\mathbf{B} \subseteq \mathcal{D}_1(w)$ is **persuasive from 1's perspective** iff $\mathcal{JS}_2(a) =$ goal w.r.t. $(\mathcal{D}_2^{+B}(w'), \rightsquigarrow \upharpoonright \mathcal{D}_2^{+B}(w'))$ for all $w' \in W$ s.t. $w \mathcal{R}_1 w'$.

A set of arguments is epistemically persuasive iff it is thought to be persuasive by the speaker.

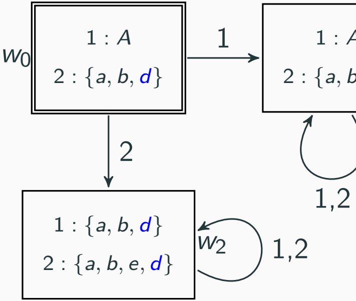
Plain Persuasion and Epistemic Persuasion



Assume goal $= \{\text{in}\}$

Figure 2: $(M, w_0)^{+d}$

Plain Persuasion and Epistemic Persuasion

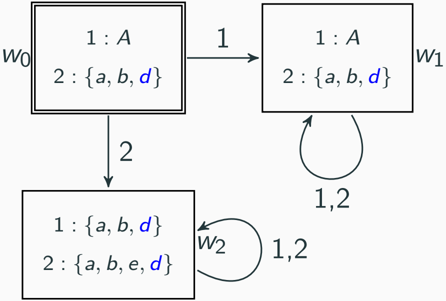


Assume goal = {in}

Figure 2: $(M, w_0)^{+d}$

- $\{d\}$ is persuasive from 1's perspective in (M, w_0)

Plain Persuasion and Epistemic Persuasion

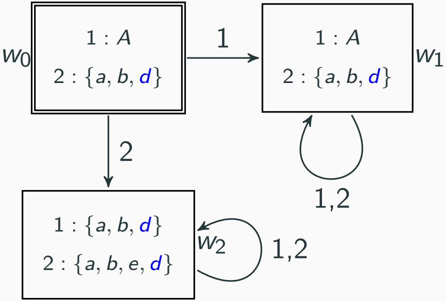


Assume goal $= \{\text{in}\}$

Figure 2: $(M, w_0)^{+d}$

- $\{d\}$ is persuasive from 1's perspective in (M, w_0)
- $\{d\}$ is not actually persuasive

Plain Persuasion and Epistemic Persuasion



Assume goal = {in}

Figure 2: $(M, w_0)^{+d}$

- $\{d\}$ is persuasive from 1's perspective in (M, w_0)
- $\{d\}$ is not actually persuasive
- goal = {in} is not achievable in $(A, \mathcal{D}_1^{+d}(w_0), \mathcal{D}_2^{+d}(w_0), \rightsquigarrow)$,

A Logic for Argument Disclosure

Let $A \neq \emptyset$ and finite and define $\mathcal{L}^{+!}(A)$

$$\varphi ::= \mathsf{owns}_i(a) \mid \neg\varphi \mid \varphi \wedge \varphi \mid \Box_i\varphi \mid [+a]\varphi \mid [a!]\varphi$$

$$a \in A \quad i \in \{1, 2\}$$

A Logic for Argument Disclosure

Let $A \neq \emptyset$ and finite and define $\mathcal{L}^{+!}(A)$

$$\varphi ::= \mathsf{owns}_i(a) \mid \neg\varphi \mid \varphi \wedge \varphi \mid \Box_i\varphi \mid [+a]\varphi \mid [a!]\varphi$$

$$a \in A \quad i \in \{1, 2\}$$

$$\begin{array}{lll}
(M, w) \vDash \mathsf{owns}_i(a) & \text{iff} & a \in \mathcal{D}_i(w) \\
(M, w) \vDash \Box_i\varphi & \text{iff} & (M, w') \vDash \varphi \quad \forall w' \text{ s.t. } w\mathcal{R}_i w' \\
(M, w) \vDash [a!]\varphi & \text{iff} & (M, w)^{a!} \vDash \varphi \\
(M, w) \vDash [+a]\varphi & \text{iff} & (M, w)^{+a} \vDash \varphi
\end{array}$$

where $(M, w)^{a!} = ((W, \mathcal{R}, \mathcal{D}^{a!}), w)$ and $\mathcal{D}_i^{a!}(v) = \mathcal{D}_i(v) \cup \{a\}$ for all $v \in W$

Axioms	
All propositional tautologies	(Taut)
$\vdash \Box_i(\varphi \rightarrow \psi) \rightarrow (\Box_i\varphi \rightarrow \Box_i\psi)$	(K)
$\vdash \mathsf{owns}_i(a) \rightarrow \Box_i \mathsf{owns}_i(a)$	(PI)
$\vdash \neg\mathsf{owns}_i(a) \rightarrow \Box_i \neg\mathsf{owns}_j(a)$	(GNI)
Rules	
From $\varphi \rightarrow \psi$ and φ, infer ψ	MP
From φ infer $\Box_i\varphi$	NEC

Table 1: Axioms for the static fragments

Reduction Axioms

$\vdash [+a]\varphi \leftrightarrow (\text{owns}_1(a) \rightarrow [a!]\varphi) \wedge (\neg\text{owns}_1(a) \rightarrow \varphi)$	(Def+)
$\vdash [a!]\text{owns}_i(a) \leftrightarrow \top$	(Atoms$^{=}$)
$\vdash [a!]\text{owns}_i(b) \leftrightarrow \text{owns}_i(b)$ where $a \neq b$	(Atoms$^{\neq}$)
$\vdash [a!]\neg\varphi \leftrightarrow \neg[a!]\varphi$	(Negation)
$\vdash [a!](\varphi \wedge \psi) \leftrightarrow ([a!]\varphi \wedge [a!]\psi)$	(Conjunction)
$\vdash [a!]\Box_i\varphi \leftrightarrow \Box_i[a!]\varphi$	(Box)
From $\varphi \leftrightarrow \psi$, infer $\delta \leftrightarrow \delta[\varphi/\psi]$	SE

Table 2: Reduction Axioms for $\mathcal{L}^{+!}(A)$

Plain Persuasiveness and Epistemic Persuasiveness

Proposition

Given a pointed model (M, w) *for* $(A, \rightsquigarrow, A_1, A_2, a)$*, let* $B \subseteq A$ *be persuasive from the speaker's perspective. Let*
$A_i := \{a_i \in A \mid M, w \models \mathsf{owns}_2(a_i) \wedge \neg\Box_1 \mathsf{owns}_2(a_i)\}$.

If $A_i \not\rightsquigarrow (\mathcal{D}_2^{+B}(w) \setminus A_i)$ *then* B *is persuasive.*

New Advances and Future Work

Two main pending tasks

Two main pending tasks

1. Capturing persuasive sets and EB-persuasive sets in the object language:

Two main pending tasks

1. Capturing persuasive sets and *EB*-persuasive sets in the object language:
    - Including new kind of variables in the propositional fragment to talk about the attack relation (following (Doutre et al., 2014, 2017))and the belonging of each argument to certain subsets

Two main pending tasks

1. Capturing persuasive sets and *EB*-persuasive sets in the object language:
    - Including new kind of variables in the propositional fragment to talk about the attack relation (following (Doutre et al., 2014, 2017))and the belonging of each argument to certain subsets
2. Dropping the assumption of credulous agents

Two main pending tasks

1. Capturing persuasive sets and *EB*-persuasive sets in the object language:
    - Including new kind of variables in the propositional fragment to talk about the attack relation (following (Doutre et al., 2014, 2017))and the belonging of each argument to certain subsets
2. Dropping the assumption of credulous agents
    - Studying new forms of update where the hearer behaves more sceptically (for instance by privately learning an attacker of the communicated argument)

Thanks for your attention!

References

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

Sylvie Doutre, Andreas Herzig, and Laurent Perrussel. A dynamic logic framework for abstract argumentation. In *Fourteenth International Conference on the Principles of Knowledge Representation and Reasoning*, 2014.

References ii

Sylvie Doutre, Faustine Maffre, and Peter McBurney. A dynamic logic framework for abstract argumentation: adding and removing arguments. In *International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*, pages 295–305. Springer, 2017.

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. *COMMA*, 75:177–184, 2012.

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

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