ABDUCTION IN ARGUMENTATION FRAMEWORKS AND ITS USE IN DEBATE GAMES

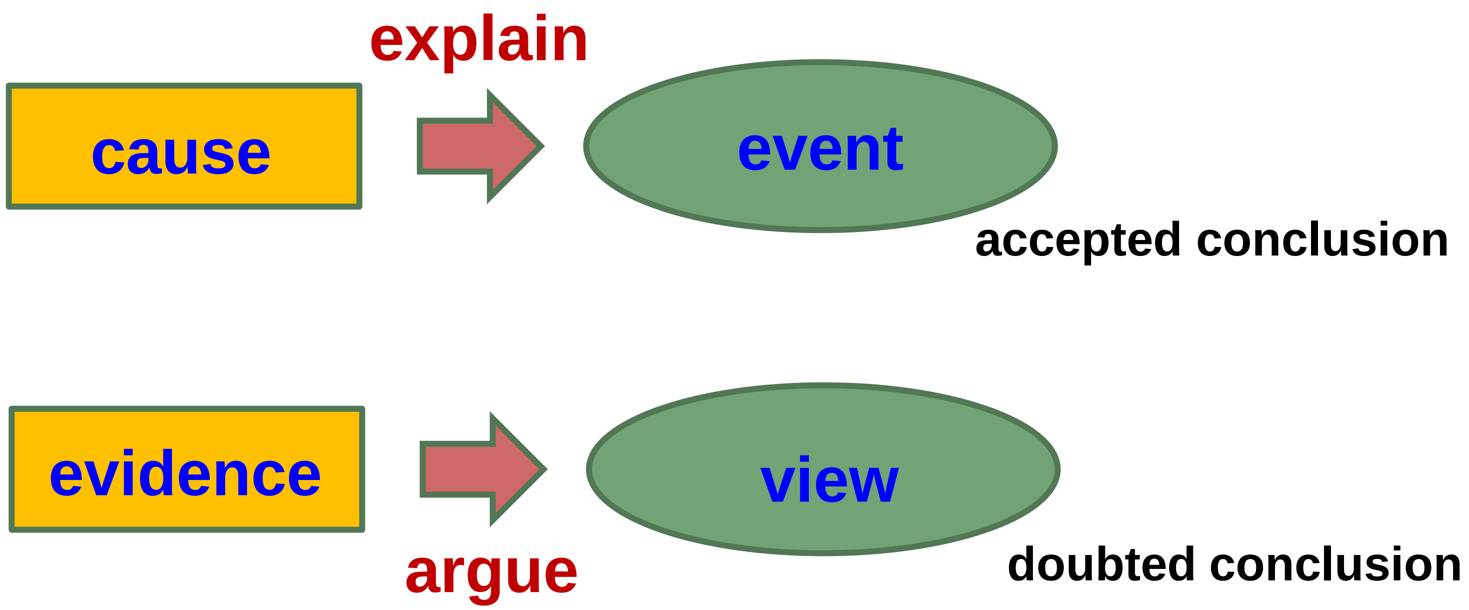
CHIAKI SAKAMA
WAKAYAMA UNIVERSITY, JAPAN

AAA 2013, Kanagawa Japan, October 2013

DIFFERENT ROLES OF ARGUMENT AND EXPLANATION

"The purpose of an **explanation** is to show ***why*** and ***how*** some phenomenon occurred or some event happened; the purpose of an **argument** is to show ***that*** some view or statement is correct or true."

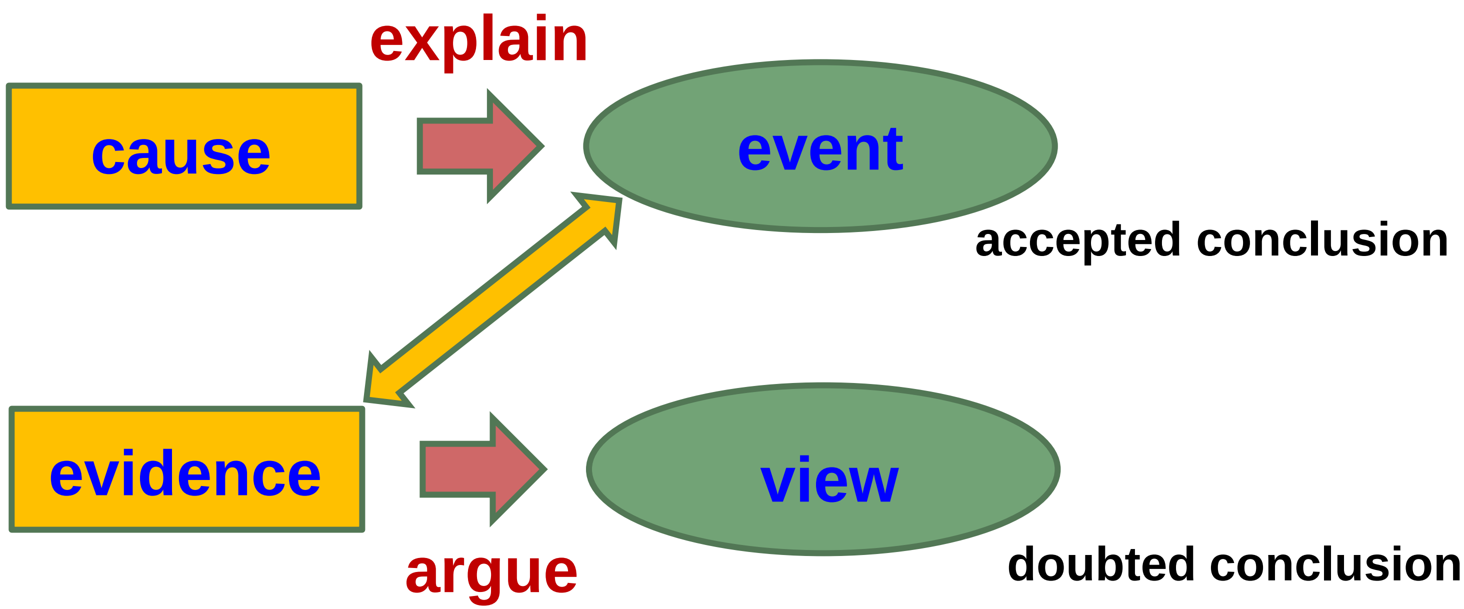
— ***Critical Thinking*** by W. Hughes (1992)



COMPLEMENTARY ROLES OF ARGUMENT AND EXPLANATION

“arguments and explanations have a complementary relationship and reasoning is normally perceived as incomplete when one occurs in the absence of the other” — ***Informal Logic*** by G. R. Mayes (2010)

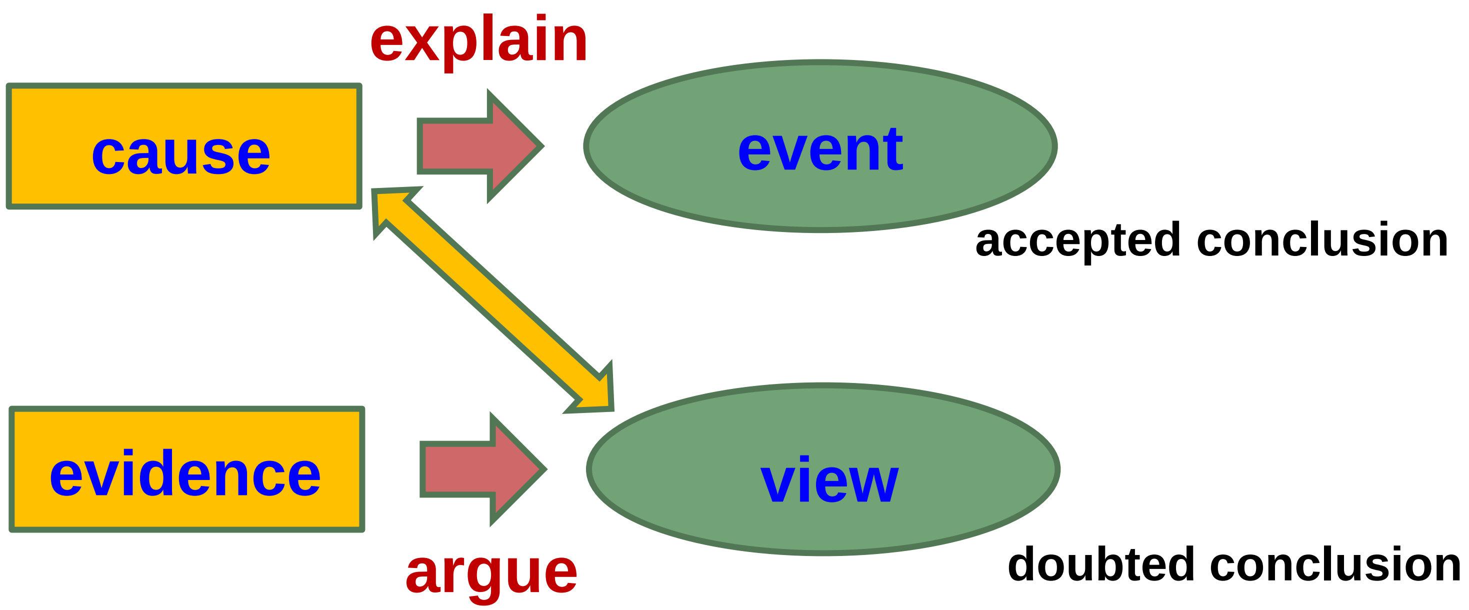
Explanation helps to understand how evidence occurs.



COMPLEMENTARY ROLES OF ARGUMENT AND EXPLANATION

“arguments and explanations have a complementary relationship and reasoning is normally perceived as incomplete when one occurs in the absence of the other” — ***Informal Logic*** by G. R. Mayes (2010)

Argument helps to know whether some view (or hypothesis) is true or not.



CONTRIBUTIONS

- **Argumentation framework** by Dung (1995) provides an abstract framework for argumentative reasoning, but it does not have a mechanism of explanatory reasoning.
- We propose an **abductive framework** based on Dung's argumentation framework.
- We apply the proposed framework to reasoning by players in **debate games**.

RELATED WORKS

	argumentation framework	abduction
Kakas & Moraitis, 2002	supportive argument	normal
Bex et al, 2007	evidential reasoning	normal
Dung et al, 2009	assumption-based AF	normal
Wakaki et al, 2010	Dung's AF associated with abductive logic programming	normal
Seselja et al, 2013	explanatory AF	normal
Our framework	Dung's abstract AF	extended

Normal abduction: addition of hypotheses

Extended abduction (Inoue & Sakama, IJCAI-95):

addition of hypotheses + deletion of hypotheses

ARGUMENTATION FRAMEWORK

(DUNG 1995; CAMINADA & GABBAY 2009)

- An **argumentation framework (AF)** is a pair (Ar, att) where Ar is a finite set of arguments and $att \subseteq Ar \times Ar$.
  An argument A **attacks** an argument B iff $(A,B) \in att$.

- $AF=(Ar, att)$ is represented by a directed graph where vertices are arguments in Ar and arcs from A to B exist whenever $(A,B) \in att$.

- A **labelling** of AF is a function $\mathtt{L}: Ar \rightarrow \{ in, out, undec \}$.

- A labelling $\mathtt{L}$ of AF is **complete labelling** if for each $A \in Ar$,
  - $\mathtt{L}(A)=in$ iff $\mathtt{L}(B)=out$ for every $B \in Ar$ s.t. $(B, A) \in att$.
  - $\mathtt{L}(A)=out$ iff $\mathtt{L}(B)=in$ for some $B \in Ar$ s.t. $(B, A) \in att$.
  - $\mathtt{L}(A)=undec$ iff $\mathtt{L}(A) \neq in$ and $\mathtt{L}(A) \neq out$.

ABDUCTION IN ARGUMENTATION FRAMEWORK

Alice: "I think Mary can speak Japanese because she has stayed in Japan."

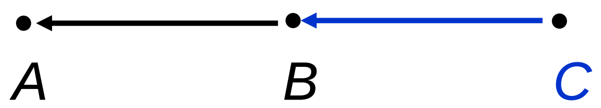
Bob: "I don't think so because her staying in Japan was too short to learn Japanese."

- The situation is represented by the argumentation framework $AF=(\{A,B\}, \{(B,A)\})$ where *A* and *B* represent arguments by Alice and Bob, respectively.

- *AF* has the complete labelling $\{ out(A), in(B) \}$ which means that the argument *A* is **rejected** and the argument *B* is **accepted**.

ABDUCTION IN ARGUMENTATION FRAMEWORK

- In another day, Bob observes that Mary speaks Japanese.
- To explain this, he assumes an argument *C* that Mary studied Japanese hard to be able to speak it well.
- The revised argumentation becomes *AF'* =({*A*,*B*,*C*}, {(*C*,*B*),(*B*,*A*)})



- *AF'* has the complete labelling { *in*(*A*), *out*(*B*), *in*(*C*) } where *A* and *C* are accepted and *B* is rejected.
- It illustrates the situation in which **a new argument is introduced** to explain a new observation.

ABDUCTION IN ARGUMENTATION FRAMEWORK

Alice: "I think the new iPhone will be selling well."

Bob: "I don't think so because few people will get interested in this new model."

- The situation is represented by the argumentation framework $AF=(\{A,B\}, \{(B,A)\})$ where *A* is rejected and *B* is accepted.
- Later it is observed that the new iPhone breaks the sales record. Bob then withdraws his argument *B* and the revised AF becomes $AF' =(\{A\}, \{\})$. Then, *A* is now accepted in *AF'*.
- It illustrates the situation in which **a previously believed argument is removed** in face of a new observation.

UNIVERSAL AF

- The **universal argumentation framework (UAF)** is an argumentation framework (U, att_U) where U is the set of all possible arguments and $att_U \subseteq U \times U$ is the fixed attack relations over U.

- An **agent** has $AF=(Ar, att)$ where $Ar \subseteq U$ and $att = att_U \cap (Ar \times Ar)$.

- The agent has a belief on the labelling of every argument in Ar based on the attack relations in att.

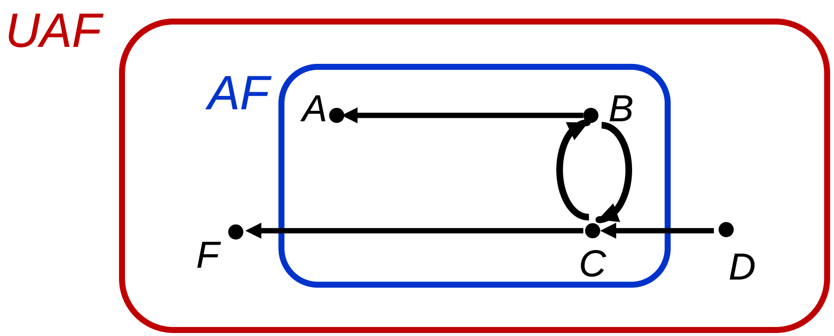
- The agent can recognize the possibility of arguments in $U \setminus Ar$, but does not know whether those arguments are valid or not.

OBSERVATION AND EXPLANATION

- Given $UAF=(U, att_U)$, an **observation** by AF is either
  $O=in(A)$ (there is an evidence for A) or
  $O=out(A)$ (there is an evidence against A)
  for some $A \in U$ such that $(A, A) \notin att_U$.
- An observation O is **skeptically** (resp. **credulously**) **explained** by $E=(I,J)$ under the labelling of AF_E if O is included in every (resp. some) labelling of the argumentation framework $AF_E=(Ar_E, att_E)$ s.t.
  - $Ar_E = (Ar \setminus J) \cup I$ where $I \subseteq U \setminus Ar$ and $J \subseteq Ar$
  - $att_E = att_U \cap (Ar_E \times Ar_E)$.
- An **explanation** $E=(I, J)$ of an observation O is **minimal** if $I' \subseteq I$ and $J' \subseteq J$ imply $I'=I$ and $J'=J$ for any explanation (I', J') of O. An explanation (I, J) is **empty** if $I=J=\{\ \}$.
- O is skeptically (or credulously) **justified** by AF iff O has the skeptical (or credulous) empty explanation.

EXAMPLE

- Let *UAF*=({ *A*,*B*,*C*,*D*,*F* }, { (*B*,*A*), (*B*,*C*), (*C*,*B*), (*D*,*C*), (*C*,*F*) }) and *AF*=({ *A*,*B*,*C* }, { (*B*,*A*), (*B*,*C*), (*C*,*B*) }).



- *AF* has 3 complete labellings: { *in*(*A*), *out*(*B*), *in*(*C*) }, { *out*(*A*), *in*(*B*), *out*(*C*) }, and { *undec*(*A*), *undec*(*B*), *undec*(*C*) }.
- Observations O_1=*in*(*A*) and O_2=*out*(*A*) have the single minimal credulous explanation *E*=({ },{ }).
- O_2 has 2 minimal skeptical explanations E_1=({ }, {*C*}) and E_2=({*D*}, { }).
- O_3=*in*(*F*) has 2 minimal skeptical explanations E_3=({*F*},{*C*}) and E_4=({*D*,*F*},{ }).

MOST PREFERRED EXPLANATION

- An explanation which minimally changes the labelling of the original AF is called a **preferred explanation**. A preferred explanation is **most preferred** if it is also a minimal explanation.

- If an observation *O* has an explanation in *AF*, then there is a most preferred explanation of *O* in *AF*.

- Computation of skeptical/credulous explanations in *AF* is done by transforming *AF* into a **logic program** (see the paper for detail).

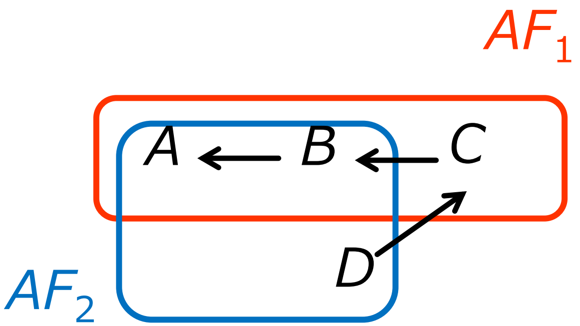
DEBATE GAME (SAKAMA, COMMA 2012)

A **debate game** provides an abstract model of debates between 2 players.

- Each player has background knowledge as an argumentation framework.
- Two players exchange a **claim** of the form:
  (*in*(*A*), _): "an argument *A* is labelled *in*" or
  (*out*(*A*), *in(B*)): "*A* is labelled *out* because *B* is labelled *in*".
- Each player can **learn** a new argument posed by the opponent, then **revises** its own AF by incorporating the new arguments and the corresponding attack relations.
- A player may claim **inaccurate** or even **false** arguments as a tactic to win a debate.

EXAMPLE

- *UAF* = ({ *A*, *B*, *C*, *D* }, { (*D*,*C*), (*C*,*B*), (*B*,*A*) }).
AF_1 = ({ *A*, *B*, *C* }, { (*C*,*B*), (*B*,*A*) }).
AF_2 = ({ *A*, *B*, *D* }, { (*B*,*A*) }).



- AF_1 has the complete labelling:
{ *in*(*A*), *out*(*B*), *in*(*C*) }

- AF_2 has the complete labelling:
{ *out*(*A*), *in*(*B*), *in*(*D*) }.

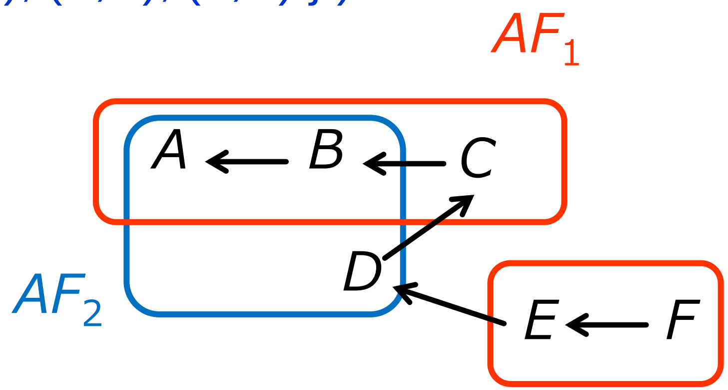
- A debate game for the argument *A* between two players proceeds as follows:

AF_1 : (*in*(*A*), _) "I claim that *A* is *in*"
AF_2 : (*out*(*A*), *in*(*B*)) "*A* is *out* because *B* is *in*"
AF_1 : (*out*(*B*), *in*(*C*)) "*B* is *out* because *C* is *in*"
⇒ Player 2 revises her AF as AF_2 =({ A,B,C,D }, { (D,C), (C,B), (B,A) }).
AF_2 : (out(C), in(D)) "*C* is *out* because *D* is *in*"
⇒ Player 1 revises his AF as AF_1 =({ A,B,C,D }, { (D,C), (C,B), (B,A) }).

- The player 1 cannot refute AF_2 , then the player 2 wins the game.

EXAMPLE

- *UAF* = ({ *A, B, C, D, E, F* }, { (*F,E*), (*E,D*), (*D,C*), (*C,B*), (*B,A*) }).
AF_1 = ({ *A, B, C, E, F* }, { (*F,E*), (*C,B*), (*B,A*) }).
AF_2 = ({ *A, B, D* }, { (*B,A*) }).



- AF_1 has the complete labelling:
{ *in*(*A*), *out*(*B*), *in*(*C*), *out*(*E*), *in*(*F*) }

- AF_2 has the complete labelling:
{ *out*(*A*), *in*(*B*), *in*(*D*) }.

- Suppose the debate game for the argument *A* :

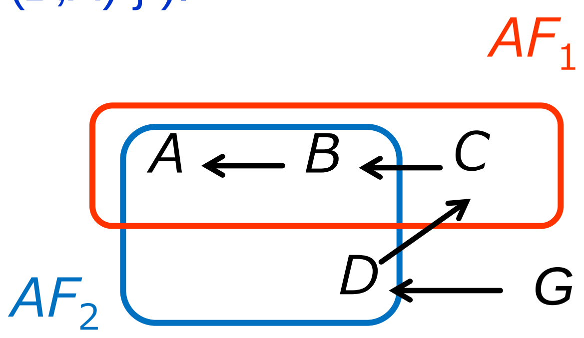
AF_1 : (*in*(*A*), _) "I claim that *A* is *in*"
AF_2 : (*out*(*A*), *in*(*B*)) "*A* is *out* because *B* is *in*"
AF_1 : (*out*(*B*), *in*(*C*)) "*B* is *out* because *C* is *in*"
AF_2 : (*out*(*C*), *in*(*D*)) "*C* is *out* because *D* is *in*"
AF_1 : (*out*(*D*), *in*(*E*)) "*D* is *out* because *E* is *in*"

- The player 2 cannot refute AF_1 , then the player 1 wins the game.

- The player 1 provides a false argument on *E* because *E* is *out*.

EXAMPLE

- $UAF = (\{A, B, C, D, G\}, \{(G,D), (D,C), (C,B), (B,A)\})$.
$AF_1 = (\{A, B, C\}, \{(C,B), (B,A)\})$.
$AF_2 = (\{A, B, D\}, \{(B,A)\})$.



- AF_1 has the complete labelling:
{ *in*(*A*), *out*(*B*), *in*(*C*) }

- AF_2 has the complete labelling:
{ *out*(*A*), *in*(*B*), *in*(*D*) }.

- Suppose the debate game for the argument *A* :

AF_1 : (*in*(*A*), _) "I claim that *A* is *in*"
AF_2 : (*out*(*A*), *in*(*B*)) "*A* is *out* because *B* is *in*"
AF_1 : (*out*(*B*), *in*(*C*)) "*B* is *out* because *C* is *in*"
AF_2 : (*out*(*C*), *in*(*D*)) "*C* is *out* because *D* is *in*"
AF_1 : (*out*(*D*), *in*(*G*)) "*D* is *out* because *G* is *in*"

- The player 2 cannot refute AF_1 , then the player 1 wins the game.

- The player 1 provides an inaccurate argument on *G* because *G* is not in AF_1 .

(DIS)HONEST CLAIMS

Let $UAF=(U, att_U)$ and $AF_i=(Ar,att)$ an AF of a player P_i.

- A claim $(in(A), \_)$ or $(out(B), in(A))$ is **honest** wrt AF_i if $A \in Ar$ and $\mathbf{L}(A)=in$ for some complete labelling $\mathbf{L}$ of AF_i .
- A claim $(in(A), \_)$ or $(out(B), in(A))$ is a **lie** wrt AF_i if $A \in Ar$ and $\mathbf{L}(A) \neq in$ for any complete labelling $\mathbf{L}$ of AF_i .
- A claim $(in(A), \_)$ or $(out(B), in(A))$ is **bullshit** wrt AF_i if $A \in U \setminus Ar$.
- A claim is **dishonest** if it is either a lie or bullshit.

COMPUTING (DISHONEST) CLAIMS BY ABDUCTION

Given a claim (*in*(*A*), _) or (*out*(*B*), *in*(*A*)) made by the player P_1,

- If *O*=*out*(*A*) has the **empty credulous explanation** in AF_2, then a player P_2 can make a **honest claim** (*out*(*A*), *in*(*C*)) that refutes the claim by P_1.
- Else if *O*=*out*(*A*) has no empty credulous explanation but has a **non-empty credulous explanation** *E* in AF_2, then a player P_2 cannot make a honest claim but can make a **dishonest claim** (*out*(*A*), *in*(*C*)) that refutes the claim by P_1.
- Otherwise, if *O*=*out*(*A*) has **no explanation**, then P_2 cannot refute the claim by P_1 and **loses** the game.

STRATEGIES FOR PLAYERS

- Try to make a **honest claim at first**.
- If there is no honest claim that can refute the opponent, make a **dishonest claim**. In this case,
  - **Avoid lie detection** (do not hide arguments which have already been used in previous claims).
  - **Prefer bullshit to lies** (arguments used in lies are believed false while truthfulness of arguments used in bullshit are unknown).