Dishonest Arguments in Debate Games

"The science of Dialectic, in one sense of the word,
is mainly concerned to tabulate and analyse dishonest stratagems"
Arthur Schopenhauer, The Art of Controversy (1896)

Chiaki Sakama

Wakayama University, Japan

COMMA 2012, Vienna

Purpose

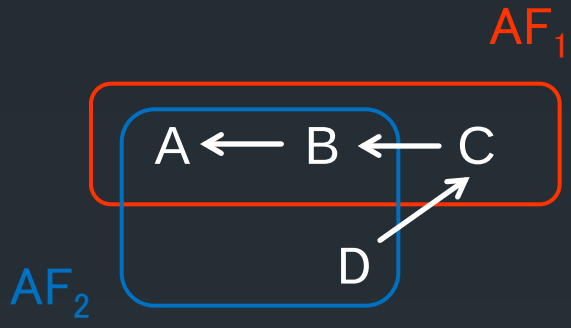
- People use dishonest arguments in daily life, while formulation of dishonest arguments has received little attention in formal argumentation.
- We introduce a debate game between two players in which a player may provide false or inaccurate arguments as a tactic to win the game.
- We formulate a debate game using formal argumentation and investigate situation where a player may provide dishonest arguments.

Debate Game

- The universal AF $UAF=(Ar, att)$ contains all arguments constructed from available information in the universe.
- A player i has his $AF_i = (Ar_i, att_i)$ as a sub-AF of the UAF s.t. $Ar_i \subseteq Ar$ and $att_i = att \cap (Ar_i \times Ar_i)$.
- Each player 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.

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) }.



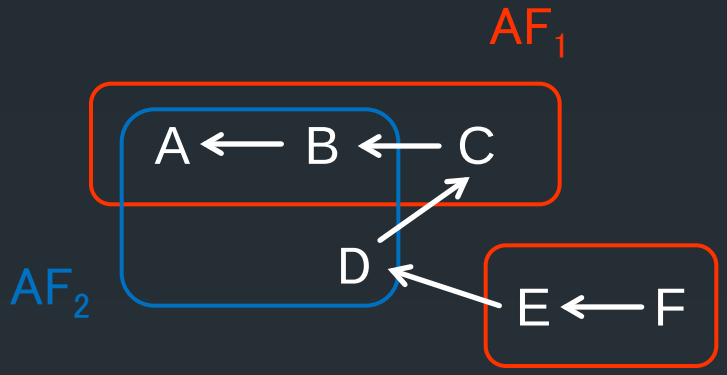
- A debate game for the argument A between two players proceeds as follows:

  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) }).



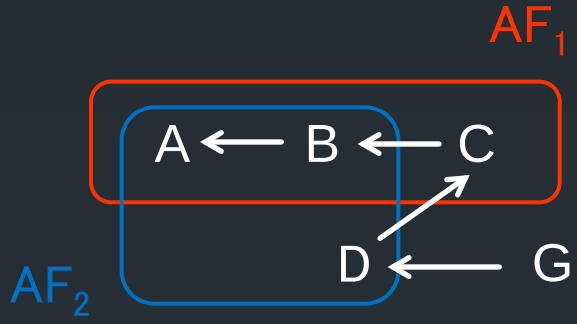
- AF_1 has the complete labelling:
  { in(A), out(B), in(C), out(E), in(F) }
- AF_2 has the complete labelling:
  { out(A), in(B), in(D) }.
- Suppose the debate game for the argument A :

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 .

Dishonest Arguments

- A player lies if he brings in(A) while believing out(A) or undec(A) in his (complete) labelling.
- A player bullshits if he brings in(A) while none of in(A), out(A), nor undec(A) is in his (complete) labelling.
  Note: We assume a bullshitter understands what arguments are possible in the UAF but does not know whether it really holds or not.
- A player is dishonest if he lies or bullshits in a game.

Contributions

- We discuss conditions when a honest player has a chance to win a debate game and when a player has a reason to behave dishonestly.
- We provide a best-practice strategy for a debate game that prescribes when to behave dishonestly and which dishonesty (lies or bullshit) a player should use at first.
- We argue the possibility of detecting dishonest arguments of the opponent player.