Strategic reasoning in conversations under imperfect information

Nicholas Asher and Soumya Paul

IRIT, Équipe MELODI
Toulouse, France

Logic and Applications (LAP 2016)
IUC, Dubrovnik

19 September 2016

Conversations as a strategic activity

Bensten-Quayle 1988 Vice Presidential debate

Christie-Rubio 2016 Republican primary debate

Coleman's spokesman Sheehan's 2008 press conference

Conversations as a strategic activity

1. **Prosecutor**: *Do you have any bank accounts in Swiss banks, Mr. Bronston?*
2. **Bronston**: *No, sir.*
3. **Prosecutor**: *Have you ever?*
4. **Bronston**: *The company had an account there for about six months, in Zurich.*
5. **Prosecutor**: *Thank you Mr. Bronston.*

[Solan-Tiersma'05, *Speaking of crime: the language of criminal justice*]

Conversations as a strategic activity

1. **Prosecutor**: *Do you have any bank accounts in Swiss banks, Mr. Bronston?*
2. **Bronston**: *No, sir.*
3. **Prosecutor**: *Have you ever?*
4. **Bronston**: *The company had an account there for about six months, in Zurich.*
5. **Prosecutor**: *Thank you Mr. Bronston.*

[Solan-Tiersma'05, *Speaking of crime: the language of criminal justice*]

Conversations as a strategic activity

- **Justin**: *Have you been seeing Valentino this past week?*
- **Janet**: *Valentino has mononucleosis.*

Puzzling observations

- In situations where the agents' interests are broadly opposed (eg. political debates, marital disputes), people still act somewhat cooperatively.
- They answer more often than not their interlocutor's questions.
- You'd better attend to what your opponent says and gauge exactly what might be meant if you hope to win a debate.
- For eg. Bronston's overall interests are opposed to that of the Prosecutor. Yet he is somewhat cooperative, and exploits the implicatures of (4) to answer his questions.

Unique characteristics of conversations

- Inherently 'turn-based' - it matters who says what.
- Can be either symmetric or asymmetric.
- A move carries more semantic content than is usually assumed:

Unique characteristics of conversations

- Inherently 'turn-based' - it matters who says what.
- Can be either symmetric or asymmetric.
- A move carries more semantic content than is usually assumed:
  - Misdirection: *The company had an account there for about six months, in Zurich.*
  - Ambiguity: *Are you coming to the party? I'm tired.*
  - Implicature: *I brought my umbrella today.*
  - Credibility, lying: *I proved* P $\neq$ NP.

Unique characteristics of conversations

- Inherently 'turn-based' - it matters who says what.
- Can be either symmetric or asymmetric.
- A move carries more semantic content than is usually assumed:
  - Misdirection: *The company had an account there for about six months, in Zurich.*
  - Ambiguity: *Are you coming to the party? I'm tired.*
  - Implicature: *I brought my umbrella today.*
  - Credibility, lying: *I proved* $\mathsf{P} \neq \mathsf{NP}$. *I proved* $\mathsf{P} = \mathsf{NP}$.

Unique characteristics of conversations

- Inherently 'turn-based' - it matters who says what.
- Can be either symmetric or asymmetric.
- A move carries more semantic content than is usually assumed:
  - Misdirection: *The company had an account there for about six months, in Zurich.*
  - Ambiguity: *Are you coming to the party? I'm tired.*
  - Implicature: *I brought my umbrella today.*
  - Credibility, lying: *I proved* $\mathsf{P} \neq \mathsf{NP}$. *I proved* $\mathsf{P} = \mathsf{NP}$.
- The presense of a 'Jury'.
- Absence of a set-end or horizon.

Message Exchange games

$$\begin{array}{llllll} & & & & x_3 & y_3\ x_4 \cdots \\ x_1 & y_1 & x_2 & y_2 & & \\ & & & & x_3' & y_3'\ x_4' \cdots \end{array}$$

- $x_j \in V_0^+$, V_0: vocabulary of Player 0
- $y_j \in V_1^+$, V_1: vocabulary of Player 1

Message Exchange games

$$x_3\ y_3\ x_4 \cdots$$

$$x_1\ y_1\ x_2\ y_2$$

$$x'_3\ y'_3\ x'_4 \cdots$$

- $x_j \in V_0^+$, V_0: vocabulary of Player 0
- $y_j \in V_1^+$, V_1: vocabulary of Player 1
- $win_0, win_1 \subset (V_0 \cup V_1)^\omega$ fixed by the Jury.

Message Exchange games

$$x_1\ y_1\ x_2\ y_2 \begin{cases} x_3\ y_3\ x_4 \cdots \in win_0 \\ x'_3\ y'_3\ x'_4 \cdots \notin win_0 \end{cases}$$

- $x_j \in V_0^+$, V_0: vocabulary of Player 0
- $y_j \in V_1^+$, V_1: vocabulary of Player 1
- $win_0, win_1 \subset (V_0 \cup V_1)^\omega$ fixed by the Jury.

Message Exchange games

$$x_1\ y_1\ x_2\ y_2 \begin{cases} x_3\ y_3\ x_4 \cdots \in win_0 \\ x'_3\ y'_3\ x'_4 \cdots \notin win_0 \end{cases}$$

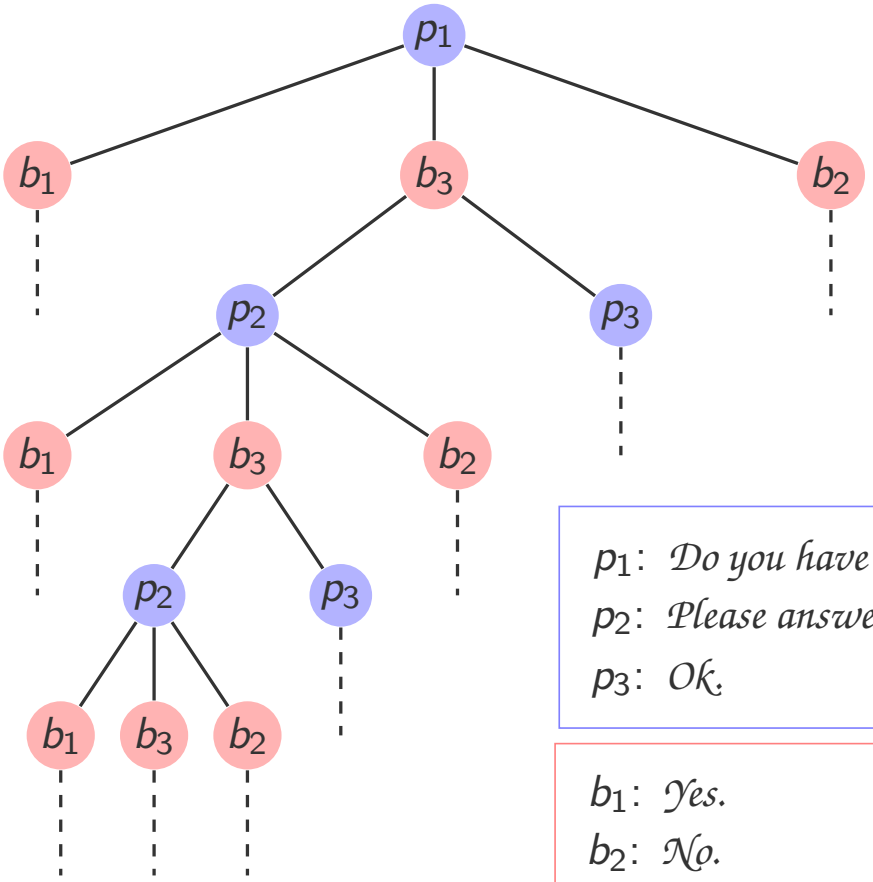
- $x_j \in V_0^+$, V_0: vocabulary of Player 0
- $y_j \in V_1^+$, V_1: vocabulary of Player 1
- $win_0, win_1 \subset (V_0 \cup V_1)^\omega$ fixed by the Jury.
- Strategy $\sigma_0 : (V_0 \cup V_1)^* V_1^+ \to V_0^+$

Message Exchange games

The vocabulary is defined using the theory of SDRT [Asher,Lascarides '03].

- V is of the form $\pi : \phi$ where
  - $\pi \in$ DU: the set of discourse constituent labels.
  - ϕ is a formula from some fixed language eg. some higher order logic for describing elementary discourse move contents.
- There is also a set of relations $\mathcal{R} \subset \text{DU} \times \text{DU}$, eg. question-answer-pair (qap).
- These relations specify when a discourse move can be played 'coherently' in a context.
- Can be described as a graph structure.

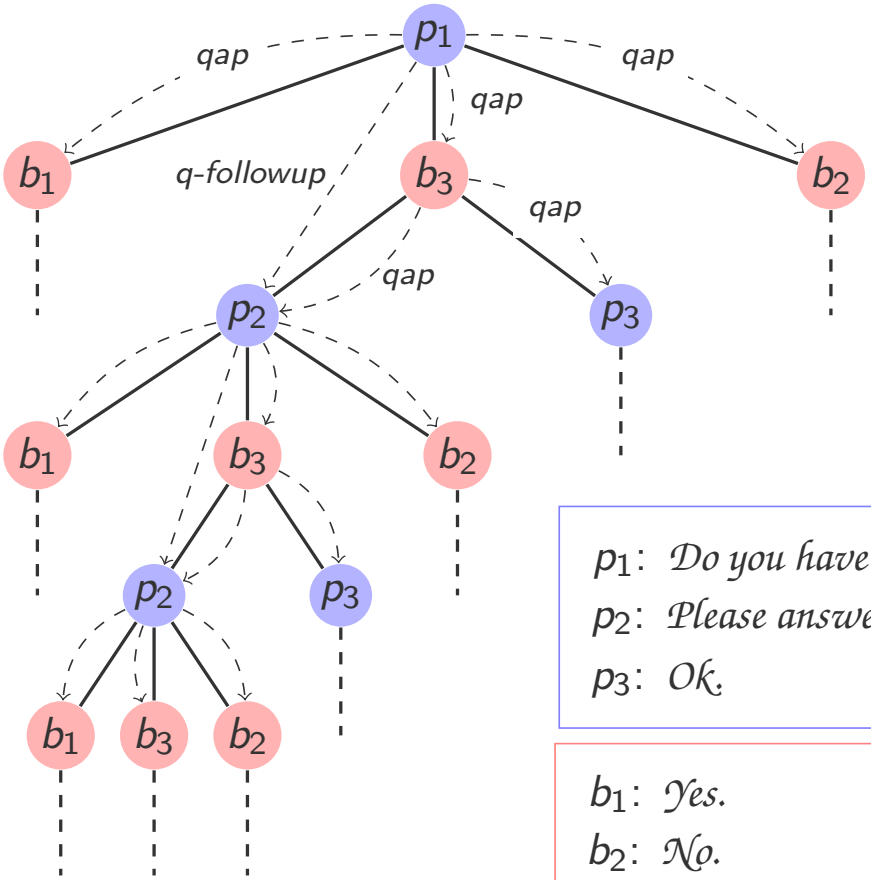
Example of an ME game



p_1: *Do you have any bank accounts in Swiss Banks, Mr. Bronston?*
p_2: *Please answer my question directly.*
p_3: *Ok.*

b_1: *Yes.*
b_2: *No.*
b_3: *The company had an account...*

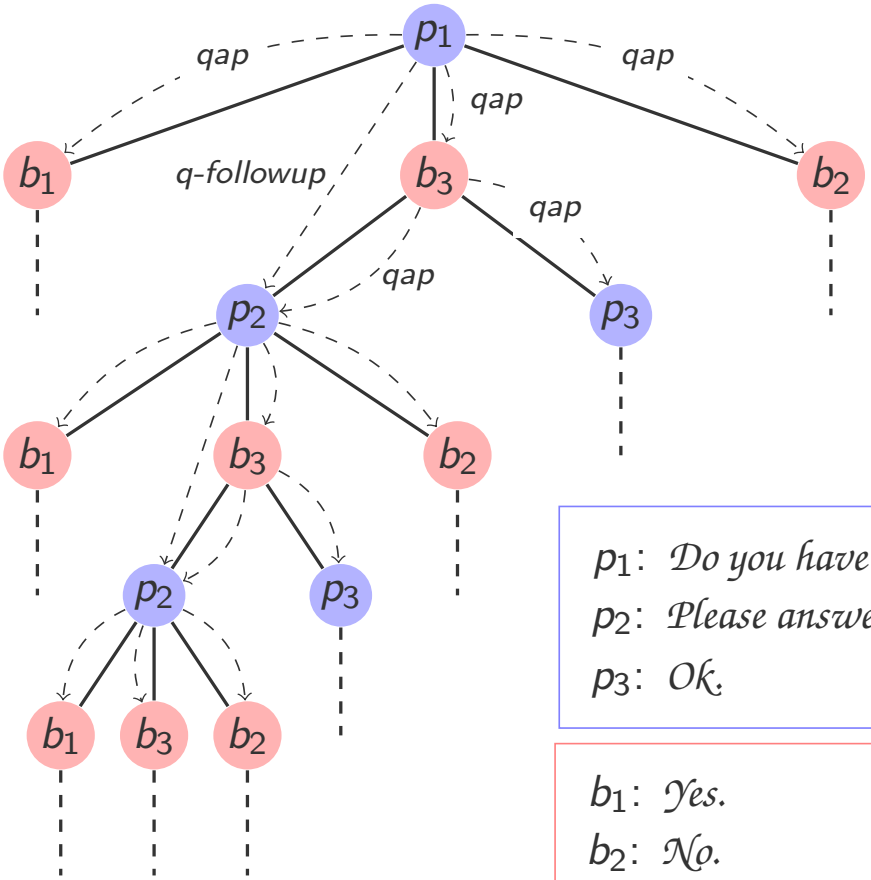
Example of an ME game



p_1: *Do you have any bank accounts in Swiss Banks, Mr. Bronston?*
p_2: *Please answer my question directly.*
p_3: *Ok.*

b_1: *Yes.*
b_2: *No.*
b_3: *The company had an account...*

Example of an ME game



p_1: *Do you have any bank accounts in Swiss Banks, Mr. Bronston?*
p_2: *Please answer my question directly.*
p_3: *Ok.*

b_1: *Yes.*
b_2: *No.*
b_3: *The company had an account...*

Prosecutor has a winning strategy!

Properties of ME games I

- Potentially unbounded, no set end.
- Messages come with a conventionally associated meaning due to the constraints enforced by the Jury: an agent who asserts ϕ commits to its content.
- We can add a language of simple propositional modal logic over the language of SDRT [Asher, Venant'15]:

$$\neg\phi \mid \phi_1 \vee \phi_2 \mid C_i\phi,\ i \in \{0,1\} \mid C^*\phi$$

- Commitment is modeled as a Kripke modal operator via an alternativeness relation in a pointed model with a distinguished (actual) world w_0.

Properties of ME games II

- Each discourse move defines an action, that updates the model's commitment structure in the style of PAL.
- Entailment relation $\models$ that ensures $\phi \models C^* C_i \phi$.
- Usual axioms.
- Allows movement from sequence of discourse moves to sequence of updates on the model.

Past approaches

- Signaling games and its cousins
  - Signaling Games [Spence'73].
  - Persuasion Games [Rubinstein-Glazer'04].
  - Cheap Talk [Crawford-Sobel'82, Farrell'87].
  - Long Cheap Talk [Aumann-Hart'03, Aumann-Maschler'95, Forges'90].
- Limitations of the signaling framework
  - Does not take into account the linguistic structure of messages.
  - Works well when the interests are aligned (Gricean players) or partially aligned.
  - Non-cooperative setting: problem of interpretation.
  - With costs: no exchange in equilibrium; without costs: 'babbling equilibrium' [Crawford-Sobel'82].
  - Inherently asymmetric.

Advantages of our approach

- Exploits the rich linguistic content of messages.
- The interpretation problem is solved extrinsically
  Prosecutor: *I want to know if* **you** *ever had an account in Swiss banks. Answer my question directly.*
- Avoids undesirable backward induction arguments.
- The set of plays has a rich topological structure.
- The player goals based on linguistic evaluations by the Jury can be characterized in terms of their structural complexity
  - Consistency, Coherence, Responsiveness etc. are all FO($<$) definable.
  - CNEC is not FO($<$) definable.
- Indicative of the strategic complexity of the player goals.

Deciding the winner finitely

Questions

- How does the Jury determine the winning sets?
- When and how does the Jury decide to stop the game?

Enter Weighted ME games

- The Jury assigns a score or weight to every move of each player.
- The weights are history dependent.
- The final score is the 'discounted sum' of the individual weights.
- Discounting reflects the fact: "play your best cards first".
- Permits the Jury to stop the game after a fixed finite number of turns with the assurance that neither player can gain or lose more than a certain amount.

Imperfect information

- It is crucial that the players are uncertain about the Jury conditions.
- They are also uncertain about the strategies of the other players.
- What constitutes rational play? equilibrium concepts?
- Type theory [Harsanyi '68].

Types and beliefs I

- Each player i has a (possibly infinite) set of types T_i.
- Each type t_i of Player i has a (first-order) belief function $\beta_i(t_i)$ which assigns to t_i a probability distribution over the types of the other players and the Jury.
- The higher-order beliefs can be defined in a standard way.
- Each type t_i starts the game with an initial belief, called the 'prior belief'.

Types and beliefs II

- After each move, all the players dynamically update their beliefs through Bayesian updates.
- 'optimal strategies', 'best-response', 'rationality', 'common belief in rationality' etc. can then be defined.
- Equilibrium concepts: NE, iterated removal of dominated strategies, correlated equilibrium, rationalizability etc. can be explored in terms of the beliefs.

Back to Bronston-Prosecutor

(I) The company had an account in the Swiss banks **but** Bronston himself did not.

Types of the Jury:

tj_1	accepts (I) and clears him of all charges
tj_2	does not accept (I) and he is charged with contempt of court
tj_3	does not accept (I) but does not charge him yet either

Beliefs of the players:

	tp_1	tp_2	tp_3	tj_1	tj_2	tj_3
tb_1	0	1	0	1	0	0
tb_2	0	0	1	0	1	0
tb_2	0	1	0	0	0	1

	tb_1	tb_2	tb_3	tj_1	tj_2	tj_3
tp_1	1	0	0	1	0	0
tp_2	1	0	0	0	1	0
tp_2	0	0	1	0	0	1

σ_1	*yes*
σ_2	*no*
σ_3	indirect

τ_1	followup
τ_2	acknowledgement

Rational type-strategies:
Bronston: $(\cdot, \sigma_3)$
Prosecutor: $(tp_1, \tau_2), (tp_2, \tau_1), (tp_2, \tau_2), (tp_3, \tau_2)$

Inexistence results

Theorem (Hellman, Levy '03,'12,'13)

A Bayesian game with a non-separable type space may not have a (measurable) equilibrium.

- Guaranteeing equilibria:
  - Restricting types to subsets definable in a countable language (FO definable, natural language).
  - Restricting types to subsets generated by resource bounded agents.
    - Limited memory.
    - Limited computation power.
  - Debates usually have a 'moderator' to assign turns.