Strategic Conversation and Meaning Exchange Games

Nicholas Asher

CNRS, Institut de Recherche en Informatique de Toulouse
Université Paul Sabatier

October 2013
Language games

1. A: Slab
2. B: (does an action)
3. A: Mortar
4. ...

Nicholas Asher (CNRS) Strategic Conversation and Meaning Exchange Games October 2013 2 / 32
What are we interested in?

- how to characterize message exchanges
- in a strategic context where user’s preferences may be incompatible.
- how to think about a wide range of goals that speakers have in conversation
- and how agents might strategize to achieve these in contexts of imperfect information
Background

- most work done to date on cooperative dialogue, where participants’ conversational goals align.

- Gricean heritage—conversation is cooperative and follows the conversational maxims.

- Formalisations of Gricean principles to date has involved strong cooperativity of intentions (if you make it manifest that you want a true and informative answer to a question, I make it my business to give you one if I’m able to do it).

- We want to investigate situations in which such cooperativity is not the case.
Real conversations can have many purposes, not just information exchange:

- bargaining
- showing off or self-promotion
- putting others down
- persuasion regardless of the facts.
- talking to misdirect or to conceal information.
An example of misdirection (Solan and Tiersma)

(1a) Prosecutor: Do you have any bank accounts in Swiss banks, Mr. Bronston?

(1b) Bronston: No, sir.

(1c) P: Have you ever?

(1d) B: The company had an account there for about six months, in Zurich.

commentary

The locutionary content of (d) is true. But Bronston succeeds in deflecting the prosecutor’s enquiry by exploiting a misleading implicature, or what one might call a *misdirection* in (d), which implicates that Bronston never had any Swiss bank account and this is false.
formal models of textual meaning from dynamic semantics with rhetorically structured discourse contexts now well developed

but a strategic conversation must have as a goal a play a conversation of a certain type

So we need to characterize conversations and winning or desirable subsets thereof..

in terms of the moves they make and/or the states they visit or revisit.

Here we take the content of the message is fixed (assuming Asher & Lascarides 2013). So signaling games will play no role here
First try: asymmetric bargaining and exchange games/ trust games.

- An exchange game a formal model of two or more agents sending goods to one another.
- Why asymmetric: speaker places his fate in the hands of the hearer when making a request or asking a question.
- Asking a question
- Trust games depict a scenario where Player X has an initial option to defer (A) to Player Y for a potentially larger payoff (C) for both.
- Player Y could defect (D) on Player X and keep more money for himself.
- $U_x(DA) < U_x(\neg A); U_x(CA) > U_x(\neg A); U_y(DA) > U_y(CA) > U_y(\neg A)$
- For a one-shot game, deference will not occur for a rational Player X.
standard trust game

\[ X \]
\[ Y \]
\[ \neg A \rightarrow A \]
\[ 0,0 \]
\[ D \rightarrow H \]
\[ -1,2 \rightarrow 1,1 \]
Large conversational games

- Why do people respond to questions when it is not in their *prima facie* interest to do so?
- What we observe in single linguistic exchanges (e.g. misdirection examples like Bronston’s) is players do not defect.
- Two ways to get reputation effects: repeated games, infinitary games.
- Repeated games require fiddling with the utility function of initial games (and what about the end games?)
- Repeated games miss out lots of linguistic detail.
- Also set end points are a big problem.
- Use backward induction and defect at the initial exchange with simple repeated exchange games.
A picture

\[
\begin{array}{c}
\text{x} \\
(0,0) \\
\text{y} \\
(-1,2) \\
\text{x} \\
(0,0) \\
\text{y} \\
(-1,2) \\
\text{x} \\
(0,0) \\
\text{y} \\
(-1,2) & (1,1)
\end{array}
\]
Models again

- At each turn, an agent can say anything and perhaps go on indefinitely long.
- In principle conversations have no intrinsic definite bound.
- So model them as infinite games of a particular type.
- Players win strategic conversations by pursuing objectives over an indefinite number of turns.
- Winning objective is often a property of the string of moves themselves. (Candidate replied well to all the questions)
Infinite games

- Banach Mazur, Gale-Stewart, Wadge, Lipschitz games
- two players each take turns choosing finite sequences of elements from a fixed set $A$
- the game has a winning condition $\text{win}$ which is a subset of the set of infinite strings over $A$, $A^\omega$
- ties to descriptive set theory, topology, and formal verification of reactive systems in CS
Structure of a BM game

- 2 players each play a finite sequence of moves from a fixed set $A$
- Players alternative indefinitely, building strings in $A^\omega$
- A BM game contains a winning condition for player 0 $\text{win}$.
- The set of infinite strings forms a metrisable topological space and so we can characterize various winning conditions in terms of basic open sets, unions of basic open sets, complements of unions of basic open sets,...
Example:

Let $X = \{a, b\}$. Then $abX^\omega$ is an open set and so is $abX^\omega \cup baX^\omega$. The complement of the set $abX^\omega$ is the set $A$ of all strings that do not have $ab$ as their prefix. This is a closed set.
Linguistic BM games

- alphabet $X$: basic discourse moves as given by SDRT—EDUs, discourse relations between EDUs, CDUs, discourse relations between CDUs and EDUs.
- each SDRS is a finite sequence of elements from our alphabet.
- The elements of the topological space are infinite strings over $X$. These are all the possible conversations using this alphabet.
- Thus formally,

**Definition**

The set of conversations is a BM game $BM(X^\omega, \text{win})$ where $X$ is a countable set of constants as described above.
A player, at turn $i$, picks a finite sequence $x_i$ of such moves, with $x_i \in X^*$. At any point in the conversation, these finite sequences of moves concatenate and give us a finite conversational play $x$.

Given a finite conversational play $x$, what the players can still say, or how the conversation can continue, is represented by all the infinite sequences that continue $x$, i.e. $xx^\omega$.

$B(X) = \{xx^\omega : x \in X^*\}$ are our basic open sets. The set of open sets in $X^\omega$ correspond to all the possible ways a conversation can go from a particular point.
Example

For instance suppose the sequence in (1) and that Player 2 responds either with (4) or (5)

4 P: I’m interested in whether you, Mr. Bronston, ever had a bank account, (f) not about your company (g) Please answer the question.

5 P: Thank you, Mr. Bronston (f’) I now would like to move to the question of your involvement in the Victory offshore trading company located in the Bahamas. (g’) Were you an officer of this company or not?

$O(1 \ast 4) \cup O(1 \ast 5)$ is the set of all continuations of the two possible sequences.
Closed sets

- The closed sets in this topology are the complements of the open sets. Let $xX^\omega$ be an open set, which represents a finite conversation $x$ and its all possible continuation thereafter.
- The complement of $xX^\omega$, $\overline{xX^\omega}$, is a closed set and represents all possible ways the conversation could have gone had the players, instead of sticking to the initial conversation $x$, deviated from it at some earlier point.
- Suppose player 0 gives the story in Figure 1 and has the play associated with (1). Then $1X^\omega$ is the set of all the ways that the conversation can continue, and $\overline{1X^\omega}$, is all the ways the conversation could have gone had player 0 deviated at some point from (1).

$$O(1 \ast 4) \cap O(1 \ast 5) = O(1 \ast 4) \cup O(1 \ast 5)$$

The ways the conversation could have deviated either from $1 \ast 4$ or from $1 \ast 5$. 
Strings and information states

- our representation of dialogue is a pair (or n tuple for n agents) of SDRSs with shared structure representing a pair of information states; at stage $n$ such a represent the commitments of each participant.
- For any SDRSs (any finite sequences $X^*$ for $X \subset A$) $\sigma$ and $\tau$, $\sigma \equiv \tau$ iff they have the same consistent continuations.
- If $\sigma \equiv \tau$, then $\sigma \equiv \tau$, where $\equiv$ is the notion of mutual dynamic entailment defined for the SDRS language.
- Using this we can speak of winning conditions on information states, which we can always lift to properties of elements in $A^*$.
A player’s conversational strategy is a function from the set of finite plays in $X^\omega$ to a finite sequence of discourse moves by her.

The objective of Player 0 is to achieve an infinite sequence of discourse moves that are in the winning condition $win$ for the conversational game $BM(X^\omega, win)$.

The objective of Player 1 is the dual of Player 0, that is, to make the play stay outside of the sequences in $win$. Thus Player 1 tries to achieve a sequence in the complement set $X^\omega \setminus win$. 
Winning conditions, $\Sigma^0_1$

- $\Sigma_1$ winning conditions can be characterized by unions of basic open sets. Equivalent to *reachability* conditions.

- Suppose $R$ is a subset of $X$. Then $\text{reachable}(R) = \{ x \in X^\omega \mid R \subseteq \text{occ}(x) \}$ is the set of all strings which contain at least one element of $R$.

- Prosecutor in the Bronston e.g. (1) is pursuing $a\Sigma_1$ objective getting, a response that defeasibly implies an answer to their question (IQAP in SDRT)
Winning conditions, $\Pi_1^0$

- $\Pi_1$ conditions are complements of $\Sigma_1$ conditions. Equivalent to safety.
- Suppose $S$ is a subset of $X$ (the ‘safe’ set). Then $safe(S) = \{x \in X^\omega \mid occ(x) = S\}$ is the set of all strings which contains elements from $S$ alone. That is, the strings remain in the safe set and do not move out of it.
- Bronston is pursuing a $\Pi_1$ objective. He wants to keep from giving responses that entail direct answers to the questions posed by Prosecutor.
Misdirection revisited, multiple BM games

- Interestingly, Bronston and Prosecutor are not playing exactly complementary winning conditions.
- Prosecutor is happy with the $\Sigma_1$ objective where an answer is defeasibly implied.
- Bronston is happy with the $\Pi_1$ objective where an answer is not entailed. $\Sigma_1^P \neq \Pi_1^B$

They are playing different BM games with different winning conditions.

Misdirection allows for $B$ to win his condition while allowing a weaker $\Sigma_1$ condition to obtain as well—e.g., $\Pi_1^B \subset \Sigma_1^P$.

Note that $\Pi_1^B$ is itself a $\Sigma_1$ condition.

This explains why it’s in $B$ and $P$’s interest to make the dialogue moves observed—question with IQAP. (missing in Asher & Lascarides, 2013), and why the implicature exists (exists in both games, but irrelevant to one winning condition)! We explain the utilities for these local moves in terms of the global goals of the two players.
Misdirection defined

- Misdirection by 1 occurs when 0 and 1 are playing different games $g_1$ and $g_2$ (different win conditions) and the complement of 1’s winning condition in $g_2$ entails 0’s winning condition in $g_1$.
- The audience, posterity or in the case of Bronston, the court, decides which winning condition is the appropriate one.
- By playing different games, both players can win; they can also both lose and one can win.
Comparisons to previous work

- the only previous account of misdirection we know is in Asher & Lascarides 2013.
- they analyze misdirection in terms of larger and smaller games—accounts for plausible deniability.
- unclear why it’s really in the Prosecutor’s interest to accept an IQAP.
- the new account makes sense of this.
Debates and multiple games

- a political debate could often involve two games, if the participants choose different winning conditions.
- a safe strategy, making more likely a win for each side.
Winning conditions, $\Sigma_2^0$

- $\Sigma_2$ conditions are unions of $\Pi_1$ sets. Co-Buchi conditions.
- In terms of temporal logic, we could think of a small fragment of such winning conditions in terms of a formula of the form $\Diamond \Box \phi$—eventually you will always be in a state (or property of strings) characterized by $\phi$.
- Information seeking dialogues, where eventually 0 establishes a truth that is no longer contested are examples of conversational games with $\Sigma_2$ winning conditions.
- Bargaining dialogues are similar; they aim at a stable exchange state; once in the bargain, they don’t leave that state.
- Conversations may have several goals.
Buchi conditions, $\Pi_2^0$

- $\Pi_2$ conditions are complements of Co-Buchi conditions.
- In terms of temporal logic, we could think of a fragment of winning conditions in this class in terms $\square\diamond\phi$—you are always able to visit a state (or property of strings) characterized by $\phi$.
- Clinton’s “it’s the economy, stupid”. (keep on coming back to the economy).
- not information seeking, but perhaps very useful in persuasion games or debates.
Beyond Buchi

- we can no longer characterize winning conditions that are complete for a given level in the BH hierarchy using FO(<) formulas
- though we do have examples of conversational winning strategies at higher levels, we take up a different theme, that of incomplete information.
- In what follows,
  - explore incomplete knowledge of moves by a participant (you don’t know what your opponents moves are fully)
  - explore how a message may have a different effect at different information states
Playing with different moves (Asher & Paul, 2013)

**Theorem**

Let $X$ and $Y$ be two alphabets such that $X \varsubsetneq Y$. We have the following in the Borel hierarchy:

\[
\begin{align*}
\Sigma_1 & \rightarrow \Sigma_2 & \Sigma_3 & \rightarrow \Sigma_4 & \cdots & \Sigma_\omega & \Sigma_{\omega+1} & \Sigma_\omega_1 \\
\Pi_1 & \rightarrow \Pi_2 & \Pi_3 & \rightarrow \Pi_4 & \cdots & \Pi_\omega & \Pi_{\omega+1} & \Pi_\omega_1
\end{align*}
\]

Proof idea: investigate encodings of winning conditions in the reduced vocabulary within the expanded vocabulary.
Conclusions and projects

- The last theorem points to an interesting problem about when the players are not playing with the same set of moves. Strategies may be complex in this setting.
- A theory of strategic conversation combining techniques from linguistics and games.
- Infinitary games explain many instances of strategic behavior.
- They make sense of why agents with opposed preferences nevertheless engage in some cooperative linguistic behavior.
- A precise characterization of information seeking vs. non-information seeking dialogues.
- Together with a formal theory of discourse interpretation, we can analyze successful and non-successful strategies.
What’s needed

- various forms of imperfect knowledge (we’ve assumed here complete knowledge of the game tree and of all possible moves except for the slide on Asher & Paul 2013).
- all our games are “determined” (following Martin 1975) but how does this translate to conversations?? Some winning conditions are unobtainable.
- We need preferences on various winning conditions. We’ve said nothing here about how agents decide on winning conditions.
- we’ve said little about substantive linguistic constraints on possible continuations—e.g., how do incoherent responses affect played? The agent may be much less likely to attain a winning condition.