Cooperative and non-cooperative discourse

Nicholas Asher & Eric McCready

CNRS, Institut de Recherche en Informatique de Toulouse
Université Paul Sabatier, Toulouse
Aoyama Gakuin University, Tokyo.

August 2015

Outline of course

- Lecture 1: Before the Fall: Gricean communication and its troubles
- Lecture 2: After the Fall: Strategic communication in Meaning Exchange Games
- Lecture 3: More on Meaning Exchange Games
- Lecture 4: game semantics of dialogue applied to linguistic phenomena: acknowledgments, corrections, politeness.
- Lecture 5: The evolution of cooperation reexamined from a linguistic point of view.

Outline of Lecture 1

- Gricean communication
- problems with gricean communication
- other game theoretic analyses—trust games
- some outstanding puzzles
- moving to a different framework based on a more logical conception of game

Conversation as sequential exchange

from Wittgenstein:

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

What principles might govern such exchanges?

Some starting points

- taking "language games" slogan seriously
- conversation is a rational activity between agents designed to further their ends.
- how do these ends affect the realization of linguistic signals, their structure, their semantics?

From Discourse Semantics to Games

- 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 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.

Before the Fall: Gricean communication

- a sketch of a general theory of communication between rational agents
- agents abide by a series of “maxims” that ensure cooperative exchanges.

The cooperative principle

A familiar quotation (Grice 1975):

> *Make your conversational contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged.*

This is the COOPERATIVE PRINCIPLE.

The Maxims

The cooperative principle is usually spelled out via four “Maxims of Conversation".

- Sincerity—say what you believe to be true.
- Relation—make your contribution contribute to solving the goals of the conversation
- Manner—make your contribution appropriate in complexity to requirements of conversational goals.
- Quantity—make you contribution as informative as required for the goals of the conversation.

The use of the maxims:

- Normative: indicate how communicative behavior should proceed and “describe" actions of idealized cooperative agents
- Used to derive *implicatures*

Implicature: expressed content which is not part of the literal meaning of the sentence.

Examples

John ate some of the doughnuts.

This sentence implicates that John didn’t eat all the doughnuts.

- What is the meaning of *implicate*? This is content that the sentence seems to express, but that is not, strictly speaking, implied by it, as seen by the fact that it is cancellable.

Examples

John ate some of the donuts. In fact, he ate them all.

Gricean reasoning about implicature

- A said that John ate some of the donuts.
- The sentence 'John ate all of the donuts' implies what A said, and so is strictly stronger than it.
- By Quantity then if A had known that John ate all the donuts she should have said that.
- But she didn't. Why not? Likely because doing so would have given a Quality violation, as it would not have been true; we conclude that John didn't eat all the donuts.

Evaluation

- this is all quite vague (do the right thing)
- More importantly what is left open is what are the conversational goals.
- Many assume some sort of cooperative principle, e.g., Asher & Lascarides 2003)
- If the point of a prior contribution by an interlocutor is some goal ϕ, then adopt that goal as a conversational goal and adjust the maxims to it.

Making Grice precise: aligned utility functions

- Why is it rational to adhere to Gricean maxims?
- Asher & Lascarides (*Semantics and Pragmatics* 2013) have a partial answer: if two players are *Grice cooperative* and have aligned utility functions, then their adherence to the maxims is rational.
- A utility function is a function from propositions or states of the world (goals) to real numbers that completely orders conversational goals.
- Two players with aligned utility functions assign the same ranking to conversational goals.
- A & L also offer a precisification of Grice's maxims that yields the same rankings on conversational goals (they assume strong cooperativity).
- define Grice Cooperativity in terms of conditional preferences:

$$(\phi : \psi \succ_a \neg\psi) > (\phi : \psi \succ_b \neg\psi)$$

- This ensures aligned preferences as they are revealed, provides a symmetric payoff game, and entails the formalizations of Gricean cooperativity of Perrault 1987, Asher & Lascarides 2003.
- Content cooperativity is ensured

Signaling games

- Main framework in the literature
- A sender S has knowledge of his type $t \in T$.
- He sends a message m.
- Messages have a 'conventional' meaning $\|m\| \subseteq T$
- A receiver R takes an action a upon receiving m.
- a sequential game.
- Both have outcomes given by utility functions $U_{S,R}(t,m,a)$.
- The classical solution concept used is Perfect Bayesian Equilibrium.

Signaling game generates implicatures

Some/all

a) Did every student pass the test?

b) Some passed ($\rightsquigarrow$ Not everyone passed)

Some/all game

	m_{some}	$m_{\forall}$	$a_{\exists\neg\forall}$	$a_{\forall}$
$t_{\exists\neg\forall}$	✓	–	1,1	0,0
$t_{\forall}$	✓	✓	0,0	1,1

Sketch of an analysis

Assume S sends something truthful. If R is rational and receives m_{some} she is uncertain on the state, but if he receives $m_{\forall}$, she knows that the state is $t_{\forall}$ and takes $a_{\forall}$. If S knows that R is rational, S always sends $m_{\forall}$ in state $t_{\forall}$. Finally, if R knows that S knows that he his rational, he knows by counterfactual reasoning upon receiving m_{some} that the state is not $t_{\forall}$.

Commentary

- In our game set up $U_S = U_R$ (aligned utility functions)
- the set of interpretive actions is limited (Bloch, Univ. Michigan Pragmatics Colloquium, 2009)

What happens if utilities are not aligned?

- this happens abstractly in 0 sum games.
- but also in conversations between antagonists in a debate, conversations between prosecutor and defendant, or a lawyer and an expert witness for the opposing side.
- Crawford and Sobel (1982) show that in 0 sum signaling games with cheap talk, the optimal equilibrium for a signaling game is a babbling equilibrium.

The non technical argument

- Suppose *A* and *B* are in a signaling game where interests are opposed.
- Then if *A* conveys a message, it is in her rational interest to do so.
- If it is in A's interest, it is not in *B*'s interest to react to the message, and so she should ignore it. A's message is just noise.
- If communication has a cost, then *A* should not send a message to *B*, since she knows that *B* rationally should ignore it.

Rationality and optimal action

This situation is not limited to signaling.

- Consider a game of buying and selling with nonaligned preferences.

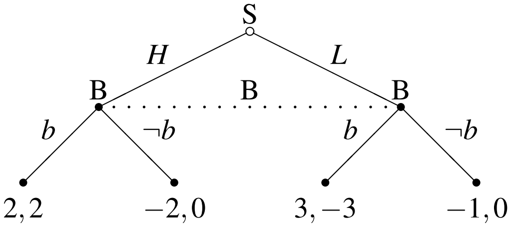


Figure: Buying watches off the street

- Higher payoff for seller with low-quality object, but the converse for the buyer.

Best option (equilibrium): $(L, \neg b)$.

- Noncooperative.

Now we allow the seller to advertise (cost = 0.5).

- A kind of communication, which might or might not be credible.

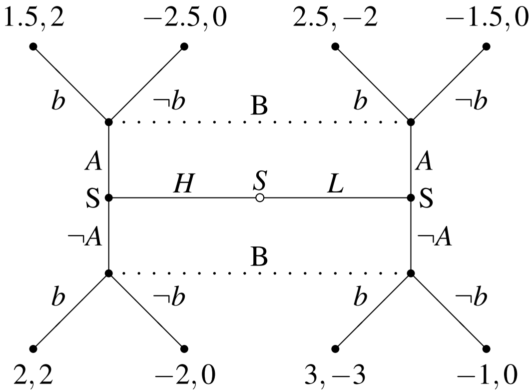


Figure: Advertising street sales

Equilibrium: $(\neg A; L, \neg b)$.

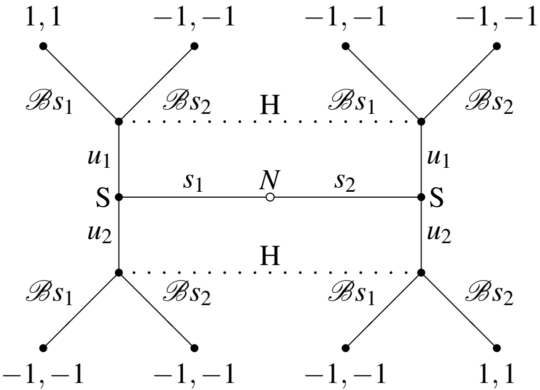
- No incentive for truthful signal;
- no incentive to believe;
- no incentive to buy; so no incentive to communicate either (see also skyrms10).
- If I try to sell you a watch on the street by claiming that it is a Rolex (= high quality?) you are not likely to believe me.

The puzzle: why does anyone bother to advertise at all?

- More generally: why bother to communicate?

The same holds for games of communication with nonaligned preferences.

Game with conventionalized signals and signaling costs.

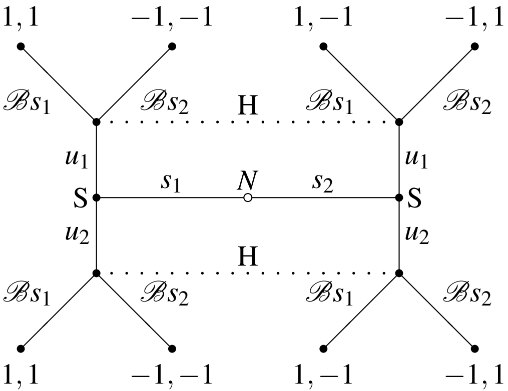


Equilibrium: match indices on situation, utterance, and object of belief. = Incentive for trust.

- Similar preferences lead to trust at equilibrium – similarly for buying and selling.

A game with nonaligned preferences:

- Hearer, as before, wants to believe something true.
- Speaker wants for hearer to think she is in state s_1.
- The obvious way to do so is to use signal u_1 in every circumstance. If the hearer believes the conventional content, she will conclude she is in s_1, and the speaker will benefit.
- However, if the state is actually s_2, the hearer will not gain anything.
- Result: the hearer has no reason to believe conventional content.



Equlibrium: (u_1, p), p the state deemed most probable by the hearer. = Failure of communication.

Upshot: when preferences are not aligned, no cooperation at equilibrium.

- More sophisticated characterization for signaling (Farrell 1990, Rabin 1990):
  - credible information transmission possible to the extent that interests are aligned
- More alignment = more credibility.

How then fares the cooperative principle in general? Not so well.

- There are many noncooperative settings.

Puzzling observations

- in situations where agents' interests are broadly opposed (e.g., 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.
- implicatures seem to work even in non-cooperative settings.

A guiding example: Bronston (Solan and Tiersma, 2005)

(2a) Do you have an account in that bank, Mr. Bronston?.

(2b) Bronston: No, sir.

(2c) P: Have you ever?

(2d) B: The company had an account there for about six months, in Zurich.

remark

Bronston's overall interests are opposed to the prosecutor (P). Yet he is "rhetorically cooperative" (Asher & Lascarides 2013), and exploits implicatures (2d) to answer questions.

Refining signaling games

- there is a large space between 0 sum games and fully cooperative Gricean games.
- we can construct a utility function for B so that he is rhetorically cooperative.
- But why that utility function?
- Not straightforward how to reason about non-Gricean cooperative situations.

Why non-Gricean cooperative situations are complicated

- Consider a sophisticated signaling game framework like that of Michael Franke (2009):
- the game structure is determined by the sender types (bijection between sender types t and receiver actions a_t)
- utility profile: S and R get rewarded iff they coordinate on actual sender type.
- a good level of abstraction, in which a_t may mean R believes that S is of type t or that S is publicly committed to being of type t.
- For Grice cooperative games, these distinctions collapse due to sincerity. So the framework seems just the right level of abstraction for Grice cooperative games.

The framework in a non-Gricean cooperative setting

- beliefs and commitments come apart.
- interpreting a_t as an action that R comes to believe $t(S)$: So if S sends $\neg$ *bank*, and R believes $\neg$ *bank*, then no charge of perjury can be genuinely brought forward.
- interpreting a_t as an action that S publicly commits to $t(S)$: So if S sends $\neg$ *bank*, *R* can level charge of perjury.

the upshot

- more details needed in a Gricean account about what a_t means precisely.
- payoffs will depend on these interrpetations and on subsequent actions so possible outcomes must take into account all possible continuations of the conversation.

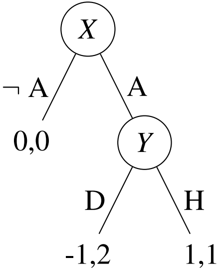
Another problem

- another problem: an agent's conversational objectives might include exposing the "bad" non-cooperative behavior of an opponent, who seeks to hide her non-cooperativity.
- asymmetry of R and S: S reveals something about his type in sending a message, but what about S's taking an action?
- m is sent with non-zero probability and the receiver uses a as a response to m with a non- zero probability iff there is a sender type t' such that a is a best response to m in t' in any equilibrium of both the sender and the receiver and the receiver's posterior probability reflects that after the sending of m, t' is more likely than t.
- The only basis for a receiver to ever accept a misleading answer is that he judges it more likely that his opponent is cooperative than not cooperative, never that he has other reasons to avoid confrontation.

Alternative to signaling games: 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
- 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.
- For a one-shot game, deference will not occur for a rational Player X.

standard trust game



Payoff Structure

$U_x(DA) < U_x(\neg A); U_x(CA) > U_x(\neg A); U_y(DA) > U_y(CA) > U_y(\neg A)$

Conversational trust games

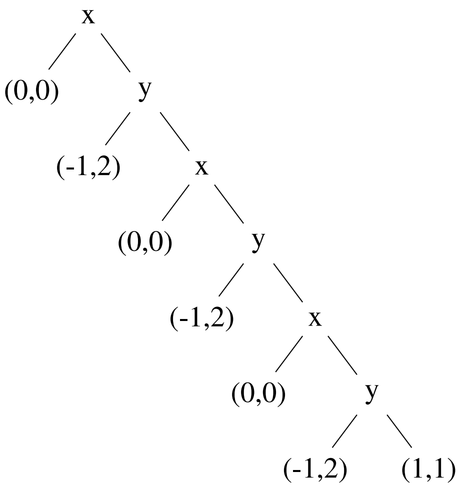
- Trust Games in Normal and Extensive Form: Player X has the option to Ask(A) Player Y for Help.
- Y can Help (H) or Defect(D).
- But why do people answer questions even when it's not in their general interest to do so?

Thus we see that the same problem – trust and cooperativity – arises in a variety of settings.

Backward Induction

- Consider an iterated trust game, say n games sequentially arranged for some fixed n.
- Reason backward from the optimal state.

A picture



The problem of iterated games of fixed length

- Player 1 will defect at penultimate move $m-1$. But player 0 knows this
- So player 0 will not play move $m-2$. But player 1 knows this so she will defect at $m-3$....
- So player 0 will not play the first move. No conversation predicted.

Other backwards induction phenomena

- Surprise examination paradox
- Centipede game

Repeated games with reputation effects still fall prey to the backward induction argument above.

- If it is known when the repetitions will cease, there is no need to worry about reputation on the final iteration;
- no cooperation expected there;
- and so on through the turns back to the initial one.

Some mechanism is needed to avoid this counterintuitive result.

Avoiding backwards induction results

- Finite but unbounded sequences? If there is no set limit to the sequence of moves, then there is no backwards induction problem.
- Conceptually complex (Trachtenbrodt)

Our solution

- Play as if conversations had no set end.
- I.e. conversations are ω length sequences.
- Finite sequences can be modeled as infinite sequences (with a repeating null element).

More motivations for infinitary games

- How do we characterize P's goal in the Bronston example?
- He wants Bronston to commit either to having an account or to not having an account.
- But he will be happy if he EVENTUALLY gets an answer; he does not need an answer on the next turn or even the turn after.
- His conversational goal is naturally expressed in terms of a linear temporal logic (LTL) formula, $\Box\phi$.
- LTL models involve ω length linear sequences.

Infinitary games

- Banach Mazur, Gale-Stewart, Wadge, Lipschitz games
- two players each take turns choosing finite sequences of elements or just one element from a fixed set A called the *vocabulary*
- the game has a winning condition *Win* which is a subset of the set of infinite strings over A, A^ω
- ties to descriptive set theory, topology, and formal verification of reactive systems in CS

Acknowledgements

This research was partially supported by:

- ERC Grant 269427 to Asher
- JSPS Kiban C Grant to McCready