Bidirectional Optimality and Signaling Games

Aniello De Santo

Math Reading Group
Spring 2016

Outline

Optimality Theory

- Mappings between levels of representations:

  Input-output mapping based on notions of grammatical preference expressed through constraints

- A theory of constraint interactions (Prince & Smolensky 1993)
- Constraints are *ranked*: NO-MIXED-VOICING < FAITH-VOICING

Optimality Theory

- Mappings between levels of representations:
  - ⇝ Input-output mapping based on notions of grammatical preference expressed through constraints
- A theory of constraint interactions (Prince & Smolensky 1993)
- Constraints are *ranked*: NO-MIXED-VOICING < FAITH-VOICING

/kaet + z /	NO-MIXED-VOICING	FAITH-VOICING
a. [kaetz]	*!	
☞ b. [kaets]		*

Optimality Theory

- Mappings between levels of representations:
  - $\rightsquigarrow$ Input-output mapping based on notions of grammatical preference expressed through constraints
- A theory of constraint interactions (Prince & Smolensky 1993)
- Constraints are *ranked*: NO-MIXED-VOICING < FAITH-VOICING

- *Unidirectional optimality*
  - Phonology (constraints interact along a single dimension)
- Bidirectional OT
  - Syntax
  - Semantics
  - Pragmatics

OT in Pragmatics

Forms and Meanings are matched by language users in production and interpretation.

OT System

Given a set of Forms $\mathcal{F}$ and a set of meanings $\mathcal{T}$, an *OT-System* is a pair $< \mathcal{G}, \preceq >$ where:

- $\mathcal{G} \subseteq \mathcal{F} \times \mathcal{T}$ is a generator
- $\preceq \subseteq \mathcal{G} \times \mathcal{G}$ is a preference ordering on elements of $\mathcal{G}$

OT in Pragmatics

Forms and Meanings are matched by language users in production and interpretation.

OT System

Given a set of Forms $\mathcal{F}$ and a set of meanings $\mathcal{T}$, an *OT-System* is a pair $< \mathcal{G}, \preceq >$ where:

- $\mathcal{G} \subseteq \mathcal{F} \times \mathcal{T}$ is a generator
- $\preceq \subseteq \mathcal{G} \times \mathcal{G}$ is a preference ordering on elements of $\mathcal{G}$

Some Notation:

We write

- $\mathcal{G}(t, f)$ for $< t, f > \in \mathcal{G}$
- $t_1 \prec_f t_2$ whenever $< t_1, f > \prec < t_2, f >$ (and viceversa)

Unidirectional OT

Unidirectionally Optimal pairing

Fixed a dimension (e.g. t):
$< t, f >$ is *unidirectionally optimal* iff $\mathcal{G}(t, f) \wedge \neg\exists f'$ s.t. $f' \prec_t f$

Unidirectional OT

Unidirectionally Optimal pairing

Fixed a dimension (e.g. t):
$< t, f >$ is *unidirectionally optimal* iff $\mathcal{G}(t, f) \wedge \neg\exists f'$ s.t. $f' \prec_t f$

But...

Optimal communication is a *dynamic* process involving *both* speaker and hearer!

Optimal communication is a *dynamic* process ...

$$OT_{speaker} = \{< t, f > \in \mathcal{G} \mid \neg \exists f' : < t, f' > \in \mathcal{G} \wedge < t, f' > \prec < t, f >$$

and

$$OT_{hearer} = \{< t, f > \in \mathcal{G} \mid \neg \exists t' : < t', f > \in \mathcal{G} \wedge < t', f > \prec < t, f >$$

Bidirectional OT

Optimal communication is a *dynamic* process ...

$$OT_{speaker} = \{< t, f > \in \mathcal{G} \mid \neg\exists f' : < t, f' > \in \mathcal{G} \wedge < t, f' > \prec < t, f >$$

and

$$OT_{hearer} = \{< t, f > \in \mathcal{G} \mid \neg\exists t' : < t', f > \in \mathcal{G} \wedge < t', f > \prec < t, f >$$

$< t, f >$ is bidirectionally optimal iff

- f is *speaker optimal* for t ($< t, f > \in OT_{speaker}$)
- t is *hearer optimal* for f ($< t, f > \in OT_{hearer}$)

Bidirectional OT

Optimal communication is a *dynamic* process ...

$$OT_{speaker} = \{< t, f > \in \mathcal{G} \mid \neg \exists f' : < t, f' > \in \mathcal{G} \wedge < t, f' > \prec < t, f >$$

and

$$OT_{hearer} = \{< t, f > \in \mathcal{G} \mid \neg \exists t' : < t', f > \in \mathcal{G} \wedge < t', f > \prec < t, f >$$

$< t, f >$ is bidirectionally optimal iff

- f is *speaker optimal* for t ($< t, f > \in OT_{speaker}$)
- t is *hearer optimal* for f ($< t, f > \in OT_{hearer}$)

This is a *strong* definition, which defines the set:

$$BiOT_{str} = OT_{speaker} \cap OT_{hearer}$$

Bidirectional OT

$< t, f >$ is *strongly* bidirectionally optimal iff

- f is *speaker optimal* for t ($< t, f > \in OT_{speaker}$)
- t is *hearer optimal* for f ($< t, f > \in OT_{hearer}$)

Bidirectional OT

$< t, f >$ is *strongly* bidirectionally optimal iff

- f is *speaker optimal* for t ($< t, f > \in OT_{speaker}$)
- t is *hearer optimal* for f ($< t, f > \in OT_{hearer}$)

$< t, f >$ is *weakly* optimal iff

- there is no *weakly* optimal $< t, f^{'} >$ s.t. $< t, f^{'} > \prec < t, f >$
- there is no *weakly* optimal $< t^{'}, f >$ s.t. $< t^{'}, f > \prec < t, f >$

Bidirectional OT

$< t, f >$ is *strongly* bidirectionally optimal iff

- f is *speaker optimal* for t ($< t, f > \in OT_{speaker}$)
- t is *hearer optimal* for f ($< t, f > \in OT_{hearer}$)

$< t, f >$ is *weakly* optimal iff

- there is no *weakly* optimal $< t, f^{'} >$ s.t. $< t, f^{'} > \prec < t, f >$
- there is no *weakly* optimal $< t^{'}, f >$ s.t. $< t^{'}, f > \prec < t, f >$

If you are troubled by the recursive definition, fear not! Here comes an example.

Example: Division of Pragmatic Labor

- Horn(1984)
- a.k.a. M-Implicature (Levinson, 2000)
- unmarked form pairs with unmarked meaning
- marked form pairs with marked meaning

Example: Division of Pragmatic Labor

- Horn(1984)
- a.k.a. M-Implicature (Levinson, 2000)
- unmarked form pairs with unmarked meaning
- marked form pairs with marked meaning

Unmarked

a. James killed Rob.
$\rightsquigarrow$ *James killed Rob in a stereotypical way.*

Marked

b. James caused Rob to die.
$\rightsquigarrow$ *James killed Rob in a non-stereotypical way.*

Example: BiOT Implementation

Precedence Relationship

Speaker Optimization $\rightarrow$ Simplicity of expression
Hearer Optimization $\rightarrow$ Stereotipicality of meaning

Example: BiOT Implementation

Precedence Relationship

Speaker Optimization $\rightarrow$ Simplicity of expression
Hearer Optimization $\rightarrow$ Stereotipicality of meaning

Strong Optimality

- $< t, f > \in \mathcal{G}$
- $\neg \exists f' : < t, f' > \in \mathcal{G} \wedge < t, f' > \prec < t, f >$
- $\neg \exists t' : < t', f > \in \mathcal{G} \wedge < t', f > \prec < t, f >$

Weak optimality

- $< t, f > \in \mathcal{G}$
- $\neg \exists f' : < t, f' > \prec < t, f > \wedge < t, f' >$ is weakly optimal
- $\neg \exists t' : < t', f > \prec < t, f > \wedge < t', f >$ is weakly optimal

Example: M-Implicature Generalized

$$\begin{tikzcd}
(t_1, f_1) & (t_1, f_2) \arrow[l] & (t_1, f_3) \arrow[l] \\
(t_2, f_1) \arrow[u] & (t_2, f_2) \arrow[l] \arrow[u] & (t_2, f_3) \arrow[l] \arrow[u] \\
(t_3, f_1) \arrow[u] & (t_3, f_2) \arrow[l] \arrow[u] & (t_3, f_3) \arrow[l] \arrow[u]
\end{tikzcd}$$

$OT_{strong} = \{(t_1, f_1)\}$

$OT_{weak} = \{(t_1, f_1), (t_2, f_2), (t_3, f_3)\}$

- How to interpret optimality?
  - pragmatic online reasoning? (e.g. Hendriks et al., 2010)
  - diachronic, evolutionary optimization? (e.g. Blutner and Zeevat, 2008)

Open Issue(s)

- How to interpret optimality?
  - pragmatic online reasoning? (e.g. Hendriks et al., 2010)
  - diachronic, evolutionary optimization? (e.g. Blutner and Zeevat, 2008)

Maybe OT can be mapped to more *tangible* concepts such as beliefs, preferences, rationality, learning, etc.

Outline

1. Optimality Theory
   - Introduction
   - Bidirectional OT and Pragmatics
   - Running Example
2. Game Theory and OT
   - Signaling Games
   - Parallels to BiOT
   - Strategies and Typology

Signaling Games: Essential Properties

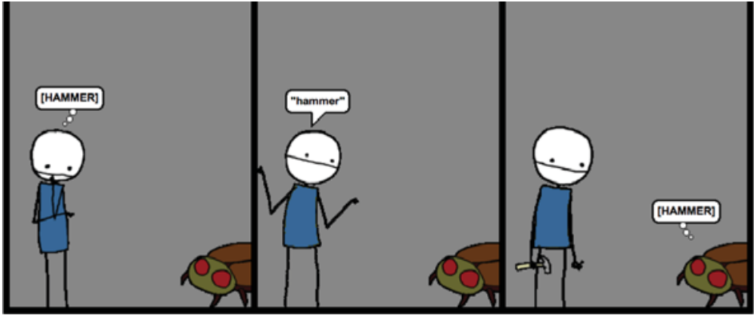
- dynamic/sequential games with incomplete info;
- models simplest case of information flow between two agents :

  1. the sender chooses a message given a state t of the word;
  2. the receiver tries to guess at the state of the word t given the observed message;

Signaling Games: Essential Properties

- dynamic/sequential games with incomplete info;
- models simplest case of information flow between two agents :

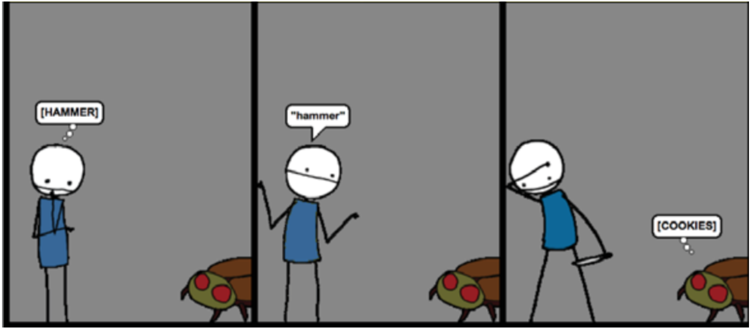
  1. the sender chooses a message given a state t of the word;
  2. the receiver tries to guess at the state of the word t given the observed message;
  3. the communication is successful if $t_{sender} = t_{receiver}$



Signaling Games: Essential Properties

- dynamic/sequential games with incomplete info;
- models simplest case of information flow between two agents :

  1. the sender chooses a message given a state t of the word;
  2. the receiver tries to guess at the state of the word t given the observed message;
  3. the communication fails if $t_{sender} \neq t_{receiver}$



Signaling Games: Informal Definition

$$< \{S, R\}, \mathcal{T}, Pr, \mathcal{M}, \mathcal{A}, U_s, U_r >$$

- a sender S and a receiver R
- set of types $t \in \mathcal{T}$
- each $t \in \mathcal{T}$ occurs with some prior probability $\mathrm{Pr(t)} \in \Delta(\mathrm{t})$
- sender S knows actual type t
- receiver R doesn't but knows prior distribution Pr $\in \Delta(\mathrm{t})$
- S chooses a message m $\in$ M
- R observes m and chooses an action a $\in$ A
- both S and R receive payoffs depending on t, m and a

Signaling Games: Trivial Parallels

$$< \{S, R\}, \mathcal{T}, Pr, \mathcal{M}, \mathcal{A}, U_s, U_r >$$

- a sender S and a receiver R $\rightarrow$speaker and hearer
- set of states $t \in \mathcal{T}$ $\rightarrow$meanings
- each $t \in \mathcal{T}$ occurs with some prior probability Pr(t) $\in \Delta$(t)
- sender S knows actual state t
- receiver R doesn't but knows prior distribution Pr $\in \Delta$(t)
- S chooses a message m $\in$ M $\rightarrow$forms
- R observes m and chooses an action a $\in$ A $\rightarrow \mathcal{A} = \mathcal{T}$
- both S and R receive payoffs depending on t, m and a

Signaling Games: Open Questions

$$< \{S, R\}, \mathcal{T}, Pr, \mathcal{M}, \mathcal{A}, U_s, U_r >$$

- a sender S and a receiver R $\rightarrow$speaker and hearer
- set of states $t \in \mathcal{T}$ $\rightarrow$meanings
- each $t \in \mathcal{T}$ occurs with some prior probability $\mathrm{Pr(t)} \in \Delta(\mathrm{t})$
- sender S knows actual state t
- receiver R doesn't but knows prior distribution $\mathrm{Pr} \in \Delta(\mathrm{t})$
- S chooses a message $\mathrm{m} \in \mathrm{M}$ $\rightarrow$forms
- R observes m and chooses an action $\mathrm{a} \in \mathrm{A}$ $\rightarrow \mathcal{A} = \mathcal{T}$
- both S and R receive payoffs depending on t, m and a
- preference pre-ordering?

Signaling Games: Open Questions

$$< \{S, R\}, \mathcal{T}, Pr, \mathcal{M}, \mathcal{A}, U_s, U_r >$$

- a sender S and a receiver R $\rightarrow$speaker and hearer
- set of states $t \in \mathcal{T}$ $\rightarrow$meanings
- each $t \in \mathcal{T}$ occurs with some prior probability $\mathrm{Pr(t)} \in \Delta(\mathrm{t})$
- sender S knows actual state t
- receiver R doesn't but knows prior distribution $\mathrm{Pr} \in \Delta(\mathrm{t})$
- S chooses a message $\mathrm{m} \in \mathrm{M}$ $\rightarrow$forms
- R observes m and chooses an action $\mathrm{a} \in \mathrm{A}$ $\rightarrow \mathcal{A} = \mathcal{T}$
- both S and R receive payoffs depending on t, m and a
- preference pre-ordering?

Many possible mappings! (Franke 2009)

DoPL: we need a model

- The speaker wants to communicate t_1 to the hearer, who wants to know it:

$$U_s(t_s, f, t_r) = U_r(t_s, f, t_r) = \begin{cases} 1 & \text{if } t_s = t_r \\ 0 & \text{otherwise} \end{cases}$$

DoPL: we need a model

- The speaker wants to communicate t_1 to the hearer, who wants to know it:

$$U_s(t_s, f, t_r) = U_r(t_s, f, t_r) = \begin{cases} 1 & \text{if } t_s = t_r \\ 0 & \text{otherwise} \end{cases}$$

- Complexity of forms is represented through *cost*:

$$c_r(f^*) = c_s(f^*) = 0.1$$

Costs maps to preference:

$$< t, f > \;\; \prec \;\; < t, f' > \text{ iff } c(t, f) \leqslant c(t, f')$$

DoPL: we need a model

- The speaker wants to communicate t_1 to the hearer, who wants to know it:

$$U_s(t_s, f, t_r) = U_r(t_s, f, t_r) = \begin{cases} 1 & \text{if } t_s = t_r \\ 0 & \text{otherwise} \end{cases}$$

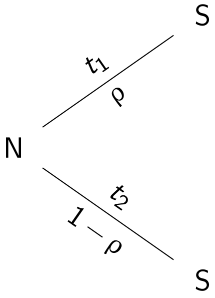
- Complexity of forms is represented through *cost*:

$$c_r(f^*) = c_s(f^*) = 0.1$$

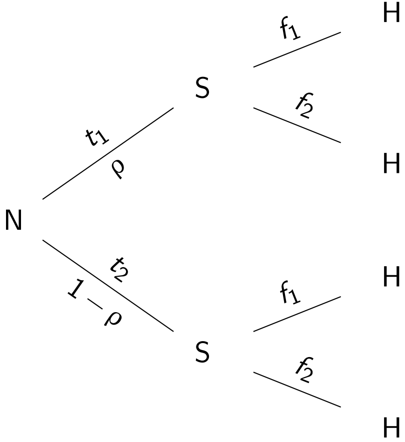
- The hearer has to guess whether t_1 or t_2 is the case:

$$Pr(t_1) = \rho \qquad\qquad Pr(t_2) = 1 - \rho$$

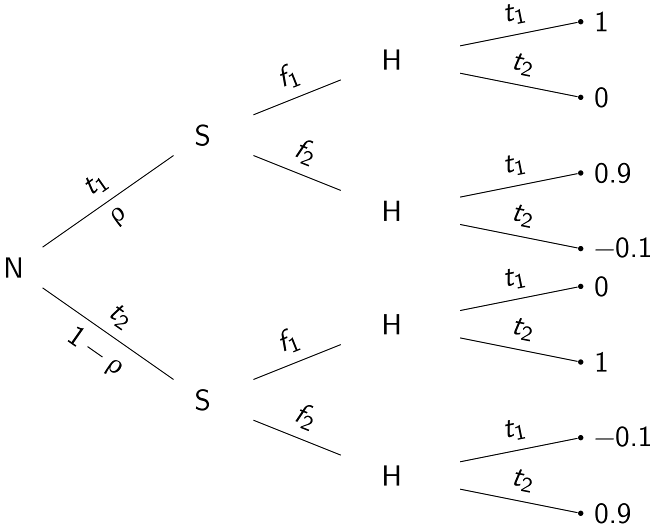
DoPL: let's play!



DoPL: let's play!



DoPL: let's play!



DoPL: Stategies

- The behavior of Speaker and Hearer is modeled by *strategies*!
- The tree shows the strategic situation with which the interlocutors are faced.

The speaker knows t. A *pure sender strategy* is a function:

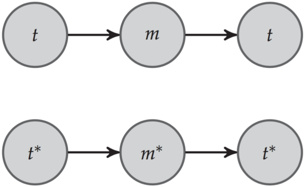
$$s \in F^T$$

The hearer observes f. A *pure hearer strategy* is a function:

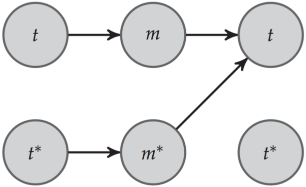
$$h \in T^F$$

Joint Strategy Profile

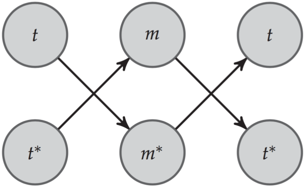
$< s, h >$ characterizes the players' joint behavior given a game



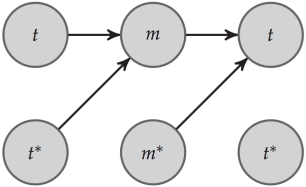
Horn Convention



Unstable Pattern



Anti-Horn Convention

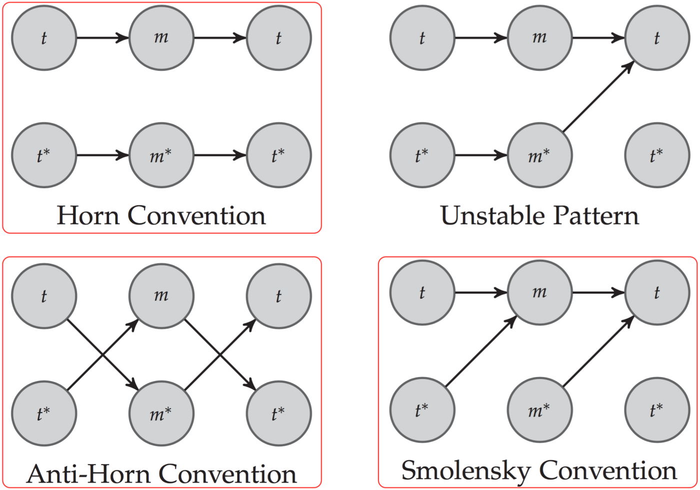


Smolensky Convention

Nash equilibrium (intuition)

Arrangement of strategies, one for each player, such that no player would benefit from unilateral deviation (i.e., no player would be better off doing something else if everybody else keeps doing the same thing).

DoPL: Nash Equilibria



Ok, cool... Why should we care?

- OT is a theory at the cognitive level;
- It is not a theory of strategic interaction;

- Game theory tell us about stable signaling equilibria...
- ... and how they can be/have been reached;

Ok, cool... Why should we care?

- OT is a theory at the cognitive level;
- It is not a theory of strategic interaction;

- Game theory tell us about stable signaling equilibria...
- ... and how they can be/have been reached;

BiOT

- insists on communicative success as a criterion to select optimal pairs
- brings a notion of strategic interaction to OT
- It is then an hybrid combining both aspects

But...

- BiOT is (still) a very top-level theory;
- More concrete reasoning/evolution schemes from Game Theory can show some of its limitations.

BiOT and Game Theory: State of the Art

The exact interpretation of BiOT is still open!

- Static Games (Dekker and van Rooij, 2000)
- Iterated Best Response Reasoning (Franke, 2009)
- Reinforcement Learning (Franke and Jager, 2011)
- Evolutionary Game Theory (Jager 2007)

Selected References

1 Dekker, Paul and Robert van Rooij (2000). *Bi-Directional Optimality Theory: An Application of Game Theory*. In: Journal of Semantics 17, pp. 217?242.

2 Franke, Michael and Gerhard Jager (2011). *Bidirectional Optimization from Reasoning and Learning in Games*. In: Journal of Logic, Language and Information.

3 Franke, Michael (2009). *An Epistemic Interpretation of Bidirectional Optimality Based on Signaling Games*. In: ZAS Papers in Linguistics 51, 2009: 111?134.

4 Horn, Laurence R. (1984). *Towards a New Taxonomy for Pragmatic Inference: Q-based and R-based Implicature*. In: Meaning, Form, and Use in Context. Ed. by Deborah Shiffrin. Washington: Georgetown University Press, pp. 11?42.

5 Levinson, Stephen C. (2000). *Presumptive Meanings. The Theory of Generalized Conversational Implicature*. Cambridge, Massachusetts: MIT Press.

6 Lewis, David (1969). *Convention. A Philosophical Study*. Harvard University Press.

Thank you!