Game Theory and Gricean Pragmatics Lesson II

Anton Benz

Zentrum für Allgemeine Sprachwissenschaften
ZAS Berlin

Course Overview

- Lesson 1: Introduction
  - From Grice to Lewis
  - Relevance Scale Approaches
- Lesson 2: Signalling Games
  - Lewis‘ Signalling Conventions
  - Parikh‘s Radical Underspecification Model
- Lesson 3: The Optimal Answer Approach I
- Lesson 4: The Optimal Answer Approach II
  - Comparison with Relevance Scale Approaches
  - Decision Contexts with Multiple Objectives

Signalling Games

Lesson II – April, 4th

Overview of Lesson II

- Lewis on Conventions
  - Examples of Conventions
  - Signalling conventions
  - Meaning in Signalling systems
- Approaches based on Signalling Games
- Parikh‘s Radical Underspecification Approach

Lewis on Conventions (1969)

Lewis on Conventions

- **Lewis Goal:** Explain the conventionality of language meaning.
- **Method:** Meaning is defined as a property of certain solutions to signalling games.
- **Achievement:** Ultimately a reduction of meaning to a regularity in behaviour.

Lewis on Conventions

1. Some Examples of Conventions
2. Lewis' Definition of Convention
3. Signalling Games and Conventions
4. Meaning in Signalling Games

Examples of Conventions

Examples of Conventions I

Driving Left or Right

- All drivers have an interest to avoid crashes.
- If two drivers meet driving in opposite directions, then they have to agree who drives on which side of the street.
- In each region or country developed a convention which tells the drivers which side to choose.

Driving Left or Right

	Left	Right
Left	1, 1	0, 0
Right	0, 0	1, 1

Examples of Conventions II

Rousseau’s stag hunters

There is a party of hunters.

- They have the possibility to hunt stag together or hunt rabbit individually.
  - If they hunt stag together, they are provided with meat for several days.
  - If they hunt individually, then they can only hunt rabbit which provides them with meet for only one day.
- They have only success hunting stag if everybody joins in. ⇒ If one hunter drops out, then all others who still go for stag will achieve nothing.

Rousseau's stag hunters

	Stag	Rabbit
Stag	2, 2	0, 1
Rabbit	1, 0	1, 1

Examples of Conventions III
Lewis' fire collectors

There is a party of campers looking for fire wood.

- It does not matter to anyone which area he searches but
- everyone has an interest not to search the same place which has already been searched by another member of the party.

Lewis' fire collectors

	North	South
North	0, 0	1, 1
South	1, 1	0, 0

Lewis' Definition of Convention

(Lewis, 2002, p. 58)

A regularity R in the behaviour of members of a population P when they are agents in an recurrent situation S is a convention if and only if it is true that, and is common knowledge in P that, in any instance of S among member of P,

1. everyone conforms to R;
2. everyone expects everyone else to conform to R;
3. everyone prefers to conform to R under the condition that the others do, since S is a coordination problem and uniform conformity to R is a coordination equilibrium in S.

Analysis of Conventions

- Conventions are solutions to a **coordination problem**.
- The coordination problem is a **recurrent** coordination problem.
- A convention consists in a **regularity in behaviour**.

- Everyone **expects** the others to follow the convention.
- A true convention has to be an **arbitrary** solution to the coordination problem.

- *In order to count as a true convention, it must be in everybody's interest that everybody follows the convention.*

Representations of Regularities of Behaviour

A regularity in behaviour can be represented by an agent's **strategy**:

- A function that tells for each type of situation which action the agent will perform.

$$S : \text{Situation-type} \rightarrow \text{Actions}$$

Signalling Conventions

(preliminary – simple cases)

The Coordination Problem in Communication

- The speaker wants to communicate some meaning M.
- In order to communicate this he chooses a **form** F.
- The hearer interprets the form F by choosing a **meaning** M'.
- Communication is successful if M=M'.

The Signalling Game

- Let **F** be a set of forms and ***M*** a set of meanings.
- The speaker's signalling strategy is a function

$$S : \boldsymbol{M} \rightarrow \boldsymbol{F}$$

- The hearer's interpretation strategy is a function

$$H : \boldsymbol{F} \rightarrow \boldsymbol{M}$$

- Speaker and hearer have success if always

$$S(M) = F \Rightarrow H(F) = M$$

Lewis' Signalling Convention

- A solution to the signalling game is a strategy pair (S,H).
- A strategy pair (S,H) with

$$S : \boldsymbol{M} \rightarrow \boldsymbol{F} \text{ and } H : \boldsymbol{F} \rightarrow \boldsymbol{M}$$

is a **signalling convention** if

$$H \circ S = id|_{\boldsymbol{M}}$$

Meaning in Signalling Games

Meaning in Signalling Conventions

Lewis (IV.4,1996) distinguishes between

- indicative signals
- imperative signals

Two different definitions of **meaning**:

- **Indicative:**

  A form F signals that M if S(M)=F

- **Imperative:**

  A form F signals to interpret it as H(F)

- Two possibilities to define meaning.
- Coincide for signalling conventions in simple signalling games.
- Lewis **defines** truth conditions of signals F as $S^{-1}(F)$.

The Paul Revere Examples

A scene from the American War of independence:

The sexton of the Old North Church informs Paul Revere about the movements of the British troops, the redcoats. The only possibility to communicate with each other is by use of lanterns. A possible signalling strategy of the sexton may look as follows:

A Possible Signalling Strategy

1. If the redcoats are observed staying home, hang no lantern in the belfry;
2. If the redcoats are observed setting out by land, hang one lantern in the belfry;
3. If the redcoats are observed setting out by sea, hang two lanterns in the belfry.

An Interpretation Strategy

1. If no lantern is observed hanging in the belfry, go home;
2. If one lantern is observed hanging in the belfry, warn the countryside that the redcoats are coming by land;
3. If two lanterns are observed hanging in the belfry, warn the countryside that the redcoats are coming by sea.

Representation of strategies

	stay	land	sea	*states*
S	0	1	2	*lanterns*

	0	1	2	*lanterns*
H	stay	land	sea	*states*

- The strategy pair is obviously a signalling convention.
- It solves the coordination problem.
- It is arbitrary.

Meaning of the Signals

Given the signalling convention before:

- 0 lanterns in the belfry *means* that the British are staying home.
- 1 lantern in the belfry *means* that the British are setting out by land.
- 2 lantern in the belfry *means* that the British are setting out by sea.

Signalling Games and Grice‘an Pragmatics

Game and Decision Theoretic Approaches to Gricean Pragmatics

Distinguish between Approaches based on:

- Classical Game Theory
  - Radical Underspecification Approach (P. Parikh).
  - Optimal Answer Approach (Benz).
- Evolutionary Game Theory
  - E.g. v. Rooij, Jäger

- Decision Theory
  - Relevance Approaches
  - E.g. Merin: Argumentative View
  - v. Rooij: Non-Argumentative View

Explanation of Implicatures

Relevance Scale Approaches (e.g. Rooij)

1. Propositions are ordered by a linear pre-order $\leq$.
2. The speaker chooses an answer A such that A is the most relevant proposition which S believes to be true.
3. Implicature F +> ψ is explained if it is known that S knows whether ψ and

$$H(F) < \neg\psi$$

Explanation of Implicatures

Diachronic Approach (e.g. Jäger)

1. Start with a signalling game **G** and a first strategy pair (S,H).
2. Diachronically, a stable strategy pair (S',H') will evolve from (S,H).
3. Implicature F +> ψ is explained if

$$H'(F) \models \psi$$

Explanation of Implicatures

Radical Underspecification Approach (Parikh)

1. Start with a signalling game **G** which allows many candidate interpretations for critical forms.
2. Impose pragmatic constraints and calculate equilibria that solve this game.
3. Implicature F +> ψ is explained if it holds for the solution (S,H):

$$H(F) \models \psi$$

Explanation of Implicatures

Optimal Answer Approach (Benz, v. Rooij)

1. Start with a signalling game where the hearer interprets forms by their literal meaning.
2. Impose pragmatic constraints and calculate equilibria that solve this game.
3. Implicature F +> φ is explained if for all solutions (S,H):

$$S^{-1}(F) \models \varphi$$

Contrast

In the optimal answer approach:

- Implicatures emerge from **indicated meaning** (in the sense of Lewis).
- Implicatures are **not** initial candidate interpretations.
- Speaker does **not** maximise relevance.
- No diachronic process.

Parikh‘s Radical Underspecification Approach

Prashant Parikh (2001)

The Use of Language

Signalling games

The general case

- We consider only signalling games with two players:
  - a speaker S,
  - a hearer H.
- Signalling games are Bayesian games in extensive form; i.e. players may have private knowledge.

Private knowledge

- We consider only cases where the **speaker** has private knowledge.
- Whatever the **hearer** knows is common knowledge.
- The private knowledge of a player is called the player’s **type**.
- It is assumed that the hearer has certain expectations about the speaker’s type.

Signalling Game

A signalling game is a tuple:

$$\langle N, \Theta, p, (A_1, A_2), (u_1, u_2) \rangle$$

- N: Set of two players S,H.
- Θ: Set of types representing the speakers private information.
- p: A probability measure over Θ representing the hearer’s expectations about the speaker’s type.

- (A_1,A_2): the speaker's and hearer's action sets.
- (u_1,u_2): the speaker's and hearer's payoff functions with

$$u_i: A_1 \times A_2 \times \Theta \rightarrow \mathbf{R}$$

Playing a signalling game

1. At the root node a type is assigned to the speaker.
2. The game starts with a move by the speaker.
3. The speaker’s move is followed by a move by the hearer.
4. This ends the game.

Strategies in a Signalling Game

- Strategies are functions from the agents information sets into their action sets.
- The speaker's information set is identified with his type $\theta \in \Theta$.
- The hearer's information set is identified with the speaker's previous move $a \in A_1$.

$$S : \Theta \to A_1 \text{ and } H : A_1 \to A_2$$

Resolving Ambiguities

Prashant Parikh

An Application

The Standard Example

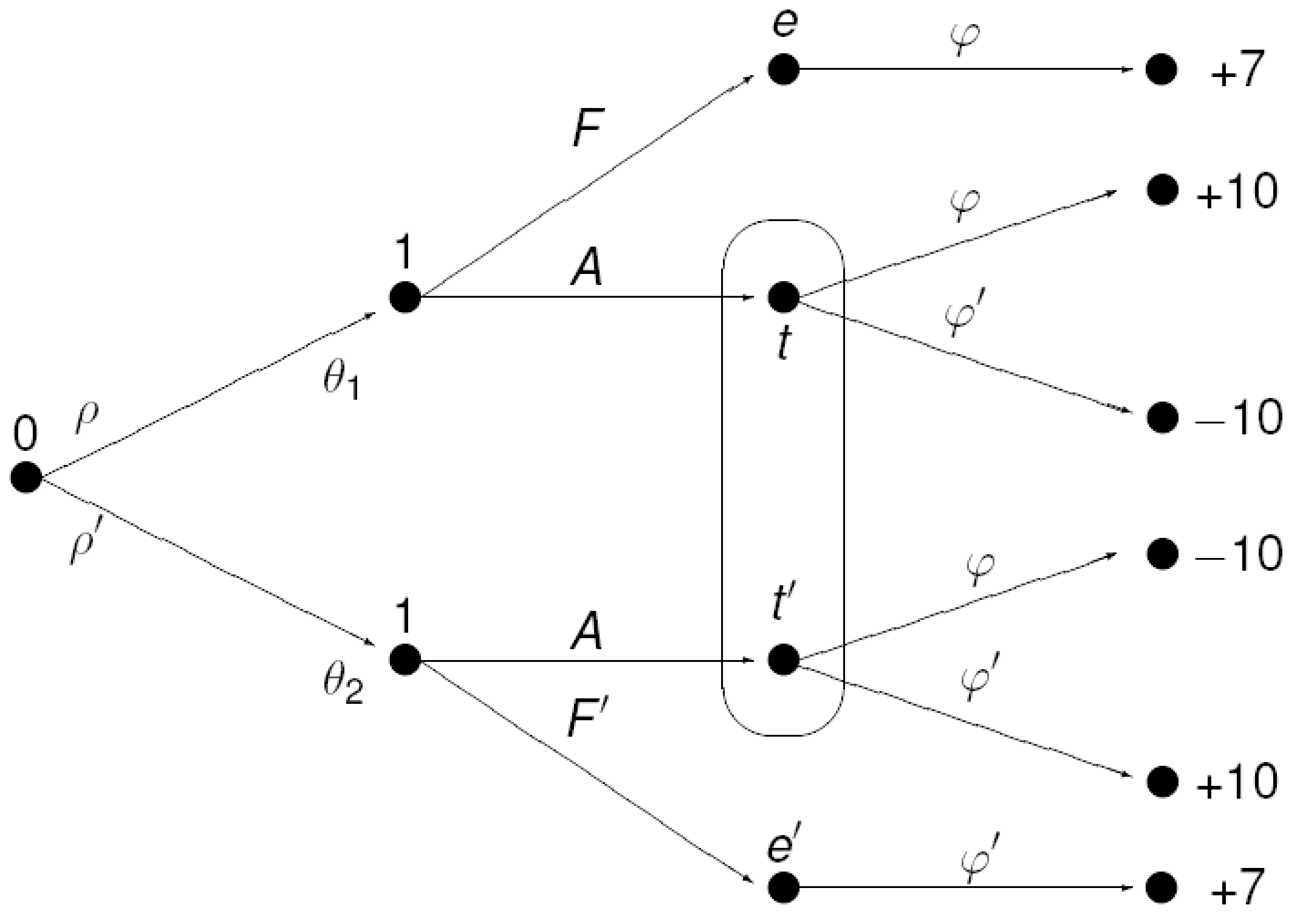
a) Every ten minutes a man gets mugged in New York. (A)
b) Every ten minutes some man or other gets mugged in New York. (F)
c) Every ten minutes a particular man gets mugged in New York. (F')

- How to read the quantifiers in a)?

Abbreviations

- φ: Meaning of `every ten minutes some man or other gets mugged in New York.'
- φ': Meaning of `Every ten minutes a particular man gets mugged in New York.'
- θ_1: State where the speaker knows that φ.
- θ_2: State where the speaker knows that φ'.

A Representation



The Strategies

	θ_1	θ_2
S	A	A
S'	A	F'
S''	F	A
S'''	F	F'

	A	F	F'
H	φ	φ	φ'
H'	φ'	φ	φ'

The Strategies

Speaker: $\mathcal{S}_1 = \{S, S', S'', S'''\}$

Hearer: $\mathcal{S}_2 = \{H, H'\}$

The Payoffs

θ_1	H	H'
S	10	-10
S'	10	-10
S''	7	7
S'''	7	7

θ_2	H	H'
S	-10	10
S'	7	7
S''	-10	10
S'''	7	7

The Payoffs

Left: In situation θ_1

Right: In situation θ_2

Expected Payoffs

	H	H'
S	8	-8
S'	9.7	-8.3
S''	5.3	7.3
S'''	7	7

The Expected Payoffs

Probability of θ_1: $\rho = 0.9$

Probability of θ_2: $\rho' = 0.1$

Core Equilibrium Concepts

- Nash Equilibrium

  A strategy pair (S,H) is a Nash equilibrium iff there are no strategies S´, H´ such that

  - the speaker prefers playing (S´,H) over (S,H),
  - the hearer prefers playing (S,H´) over (S,H).

- Pareto Nash Equlibrium

  A Nash equilibrium (S,H) is a Pareto Nash equilibrium iff there is no Nash equilibrium (S´,H´) such that both players prefer playing (S´,H´) over (S,H).

Expected Payoffs

	H	H'
S	8	-8
S'	9.7	-8.3
S''	5.3	7.3
S'''	7	7

Pareto Nash equilibrium

Nash Equilibria

The Expected Payoffs

Probability of θ_1: $\rho = 0.9$

Probability of θ_2: $\rho' = 0.1$

Analysis

- There are two Nash equilibria

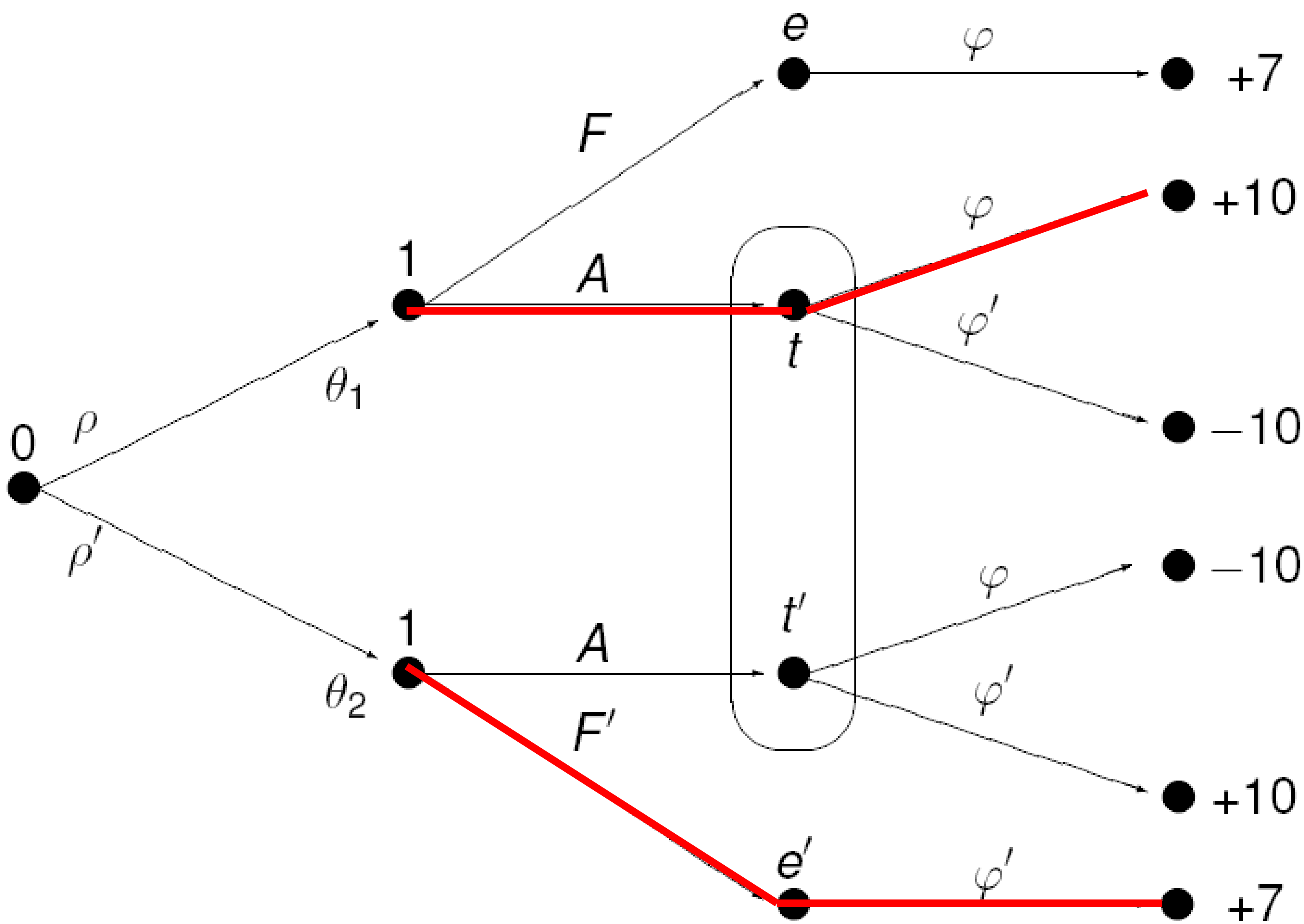
  (S',H) and (S'',H')

- The first one is also a Pareto Nash equilibrium.
- With (S',H) the utterance (A) should be interpreted as meaning (F):

(A) Every ten minutes a man gets mugged in New York.

(F) Every ten minutes some man or other gets mugged in New York.

The Pareto Optimal Solution



General Characteristics

- There is a form A that is ambiguous between meanings φ and φ'.
- There are more complex forms F, F' which can only be interpreted as meaning φ and φ'.
- The speaker but not the hearer knows whether φ (type θ_1) or φ' (type θ_2) is true.

- It is assumed that interlocutors agree on a Pareto Nash equilibria (S,H).
- The actual interpretation of a form is the meaning assigned to it by the hearer's strategy H.

Implicatures

Classification of Implicatures

Parikh (2001) distinguishes between:

- **Type I implicatures:** There exists a decision problem that is directly affected.
- **Type II implicatures**: An implicature adds to the information of the addressee without directly influencing any immediate choice of action.

Examples of Type I implicatures

1. A stands in front of his obviously immobilised car.

   A: I am out of petrol.

   B: There is a garage around the corner.

   +>The garage is open and sells petrol.

1. Assume that speaker S and hearer H have to attend a talk just after 4 p.m. S utters the sentence:

   S: It's 4 p.m. (A)

   +> S and H should go for the talk. (ψ)

A model for a type I implicature

The Example

1. Assume that speaker S and hearer H have to attend a talk just after 4 p.m. S utters the sentence:

   S: It's 4 p.m. (A)

   +> S and H should go for the talk. (ψ)

The possible worlds

The set of possible worlds Ω has elements:

- s_1: it is 4 p.m. and the speaker wants to communicate the implicature ψ that it is time to go for the talk.
- s_2: it is 4 p.m. and the speaker wants to communicate only the literal content φ.

The Speaker's types

- Assumption: the speaker knows the actual world.
- Types:
  - $\theta_1 = \{s_1\}$: speaker wants to communicate the implicature ψ.
  - $\theta_2 = \{s_2\}$: speaker wants to communicate the literal meaning φ.

Hearer's expectations about speaker's types

- Parikh's model assumes that it is much more probable that the speaker wants to communicate the implicature ψ.
- Example values:

$$p(\theta_1) = 0.7 \text{ and } p(\theta_2) = 0.3$$

The speaker's action set

The speaker chooses between the following forms:

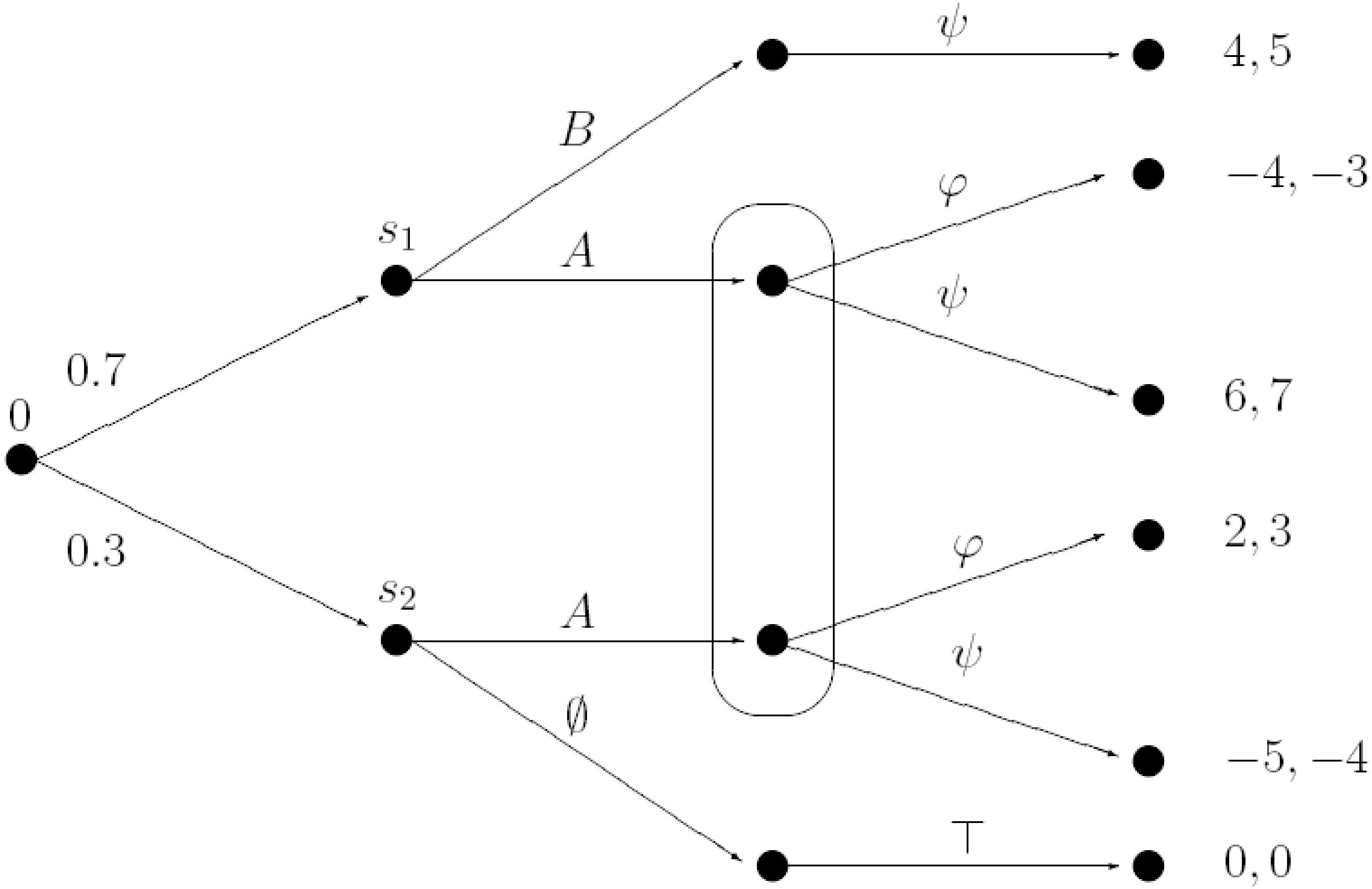
1. A ≡ It's 4 pm. ([A] = φ)
2. B ≡ It's 4 pm. Let's go for the talk. ([B] = $\psi \wedge \varphi$)

1. $\varnothing$ ≡ *silence*.

The hearer's action set

- The hearer interprets utterances by meanings.
- Parikh's model assumes that an utterance can be interpreted by any meaning χ which is stronger than its literal meaning φ.

The Game Tree



The Utility Functions

Parikh decomposes the utility functions into four additive parts:

1. A utility measure that depends on the complexity of the form and processing effort.
2. A utility measure that depends on the correctness of interpretation.
3. A utility measure that depends on the value of information.
4. A utility measure that depends on the intrinsic value of the implicated information.

Utility Value of Information

- Derived from a decision problem.
- Hearer has to decide between:
  - going to the talk
  - stay

probability	state	going	staying
0.2	time to go	10	-10
0.8	not time to go	-2	10

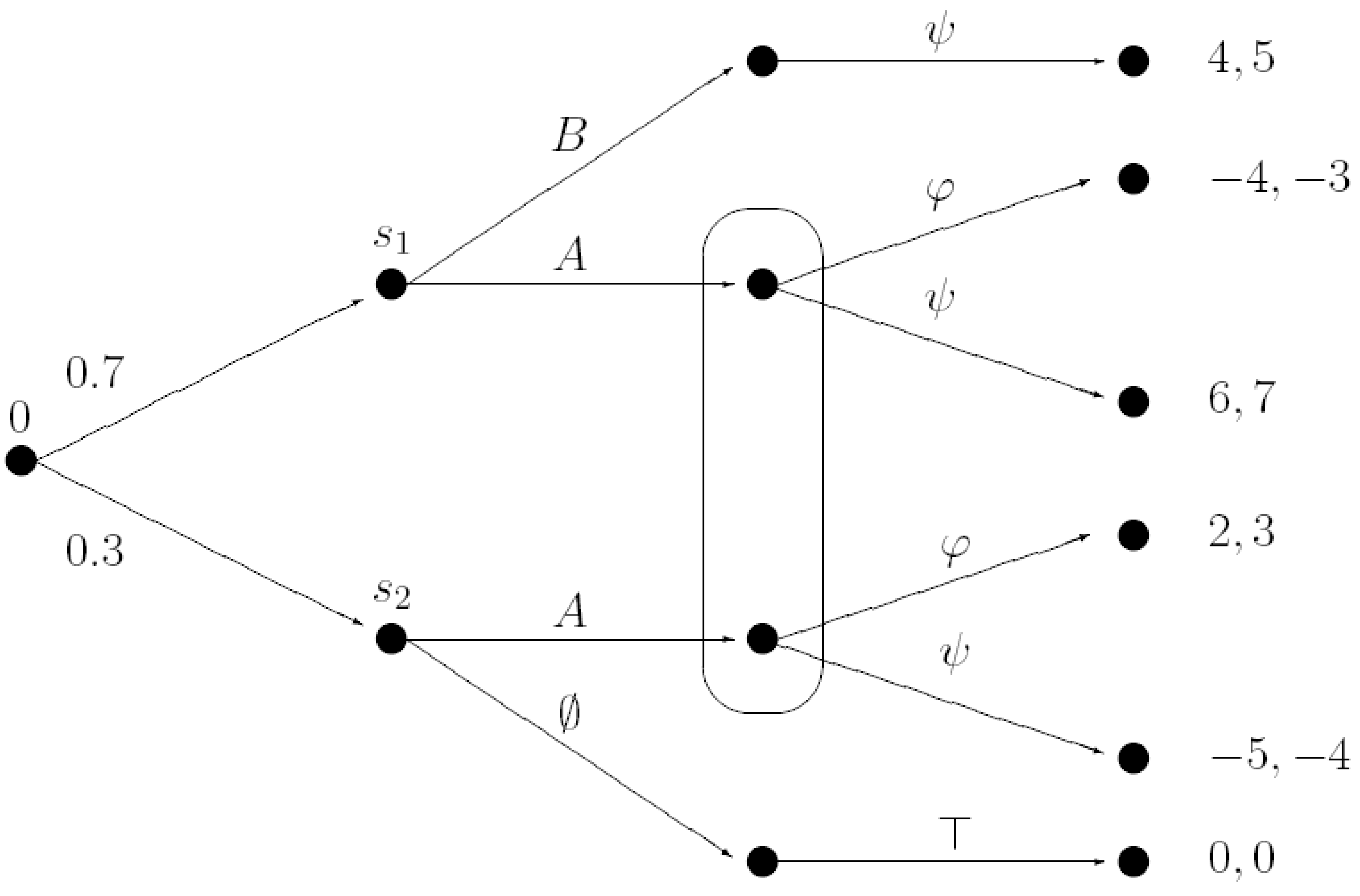
Utility Value of Information

- Before learning ‘It’s 4 p.m.’:
  - EU(leave) = 0.2×10 + 0.8×(-2) = 0.4
  - EU(not-leave) = 0.2×(-10) + 0.8×10 = 6
- After learning ‘It’s 4 p.m.’(A), hence that it is time to leave:
  - EU(leave|A) = 1×10 = 10
  - EU(not-leave|A) = 1×(-10) = -10
- Utility value of learning ‘It’s 4 p.m.’ (A):
  - UV(A) = EU(leave|A) - EU(not-leave) = 10 – 6 = 4

Other Utilities

- Intrinsic Value of Implicature: 5
- Cost of misinterpretation -2
  - In addition, Parikh assumes that in case of miscommunication the utility value of information is lost (*)
- Various costs due to complexity and processing effort.
  - Higher for speaker than hearer.

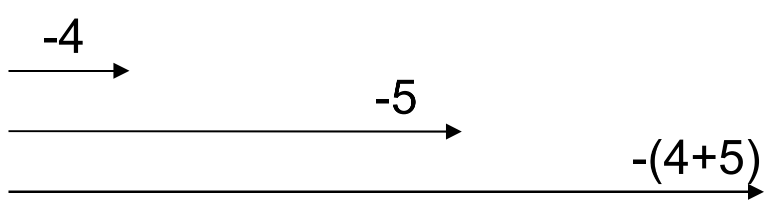
The Game Tree



Some Variations of the Payoffs

	(a)	(b)	(c)	(d)
$\langle \theta_1, B, \psi \rangle$	$4, 5$	$0, 1$	$-1, 0$	$-5, -4$
$\langle \theta_1, A, \varphi \rangle$	$0, 1$	$-4, -3$	$0, 1$	$-4, -3$
$\langle \theta_1, A, \psi \rangle$	$6, 7$	$2, 3$	$1, 2$	$-3, -2$
$\langle \theta_2, A, \varphi \rangle$	$2, 3$	$-2, -1$	$2, 3$	$-2, -1$
$\langle \theta_2, A, \psi \rangle$	$-1, 0$	$-5, -4$	$-1, 0$	$-5, -4$
$\langle \theta_2, \emptyset, \top \rangle$	$0, 0$	$0, 0$	$0, 0$	$0, 0$

a) without (*)
b) minus utility value
c) minus intr. val. of implic.
d) minus both



Result

In all variations it turns out that the strategy pair (S,H) with

- S(θ_1) = It's 4 p.m., S(θ_2) = *silence*, and
- H(It's 4 p.m) = [It's 4 p.m] $\wedge$ [Let's go to the talk]

is Pareto optimal.