Logic and Artificial Intelligence

Lecture 0

Eric Pacuit

Visiting Center for Formal Epistemology, CMU

Center for Logic and Philosophy of Science
Tilburg University

ai.stanford.edu/∼epacuit
e.j.pacuit@uvt.nl

August 29, 2011

Practicalities

Practicalities

- Course website
  http://ai.stanford.edu/∼epacuit/classes/logicai-cmu.html
  - Weekly readings will be posted
  - Slides will be posted
  - Pay attention to the schedule (midterm, canceled classes, etc.)

Practicalities

- Course website
  http://ai.stanford.edu/∼epacuit/classes/logicai-cmu.html
  - Weekly readings will be posted
  - Slides will be posted
  - Pay attention to the schedule (midterm, canceled classes, etc.)

- Weekly lecture + discussion

Practicalities

- Course website
  http://ai.stanford.edu/∼epacuit/classes/logicai-cmu.html
  - Weekly readings will be posted
  - Slides will be posted
  - Pay attention to the schedule (midterm, canceled classes, etc.)
- Weekly lecture + discussion
- Office Hours: Wednesdays 11 - 12 and by appointment (e.j.pacuit@uvt.nl)
- Office: Room 161B, Baker Hall

Practicalities: Grading

Practicalities: Grading

1. Problem sets (40%)
   - distributed periodically throughout the semester (4-6 total)
2. Midterm exam (20%)
3. Final exam (40%)

Practicalities: Literature

Practicalities: Literature

Contemporary research papers published in academic journals and chapters from recent books (consult the schedule for details).

Practicalities: Literature

Contemporary research papers published in academic journals and chapters from recent books (consult the schedule for details).

No required textbook. This means:

- Ask questions! Don't let me go too fast!
- Watch out for differences in notation
- Important to work through the problem sets (*what will be on the exams??*)

What is this course about?

What is this course about?

R. Thomason. *Logic and AI*. Stanford Encyclopedia of Philosophy (2008).

What is this course about?

R. Thomason. *Logic and AI*. Stanford Encyclopedia of Philosophy (2008).

More accurate course title: Logics of Rational Agency

We are interested in reasoning about rational (and not-so rational) agents engaged in some form of social interaction.

We are interested in reasoning about rational (and not-so rational) agents engaged in some form of social interaction.

- Philosophy (social epistemology, philosophy of action)
- Game Theory
- Social Choice Theory
- AI (multiagent systems)

We are interested in reasoning about rational (and not-so rational) agents engaged in some form of social interaction.

What is a "rational agent"? What *are we modeling?*

We are interested in reasoning about rational (and not-so rational) agents engaged in some form of social interaction.

What is a "rational agent"? What *are we modeling?*

- has consistent preferences (complete, transitive)
- (acts as if she) maximizes expected utility
- reacts to observations
- revises beliefs when learning a *surprising* piece of information
- understands higher-order information
- plans for the future
- asks questions
- ????

We are interested in reasoning about rational (and not-so rational) agents engaged in some form of social interaction.

- playing a (card) game
- having a conversation
- executing a *social procedure* (voting, making a group decision)
- ....

Goal: incorporate/extend existing game-theoretic/social choice analyses

We are interested in reasoning about rational (and not-so rational) agents engaged in some form of social interaction.

There is a jungle of logical frameworks!

- logics of informational attitudes (knowledge, beliefs, certainty)
- logics of action & agency
- temporal logics/dynamic logics
- logics of motivational attitudes (preferences, intentions)
- deontic logics

(*Not to mention various game-theoretic/social choice models and logical languages for reasoning about them*)

We are interested in reasoning about rational (and not-so rational) agents engaged in some form of social interaction.

- How can we compare different logical frameworks addressing similar aspects of rational agency and social interaction?
- How should we combine logical systems which address different aspects of social interaction towards the goal of a comprehensive (formal) theory of rational agency?
- How does a logical analysis contribute to the broader discussion of rational agency and social interaction within philosophy and the social sciences?

Game Theory

Game Theory

"We wish to find the mathematically complete principles which define 'rational behavior' for the participants." (pg. 31)

J. von Neumann and O. Morgenstern. *Theory of Games and Economic Behavior*. Princeton University Press, 1944.

Game Theory

"We wish to find the mathematically complete principles which define 'rational behavior' for the participants." (pg. 31)

J. von Neumann and O. Morgenstern. *Theory of Games and Economic Behavior*. Princeton University Press, 1944.

"Game theory is a bag of analytical tools designed to help us understand the phenomena that we observe when decision-makers interact." (pg. 1)

M. Osborne and A. Rubinstein. *Introduction to Game Theory*. MIT Press, 2004.

Game Situations

		Bob	
		L	R
Ann	U	1,2	0,0
	D	0,0	2,1

Economic "data": *feasible* options (i.e., actions), desirability (i.e., utilities), structural properties of the interactive situation (i.e., game *forms*: extensive, strategic, simultaneous moves, stochastic, etc.)

Who is game theory about?

Who is game theory about?

1. **Classical view**: idealized world with *perfectly rational agents*

2. **Humanistic view**: real people in interactive situations

L. Samuelson. *Comments on Game Theory*. Game Theory: 5 Questions, Automatic Press, 2007.

Who is game theory about?

1. **Classical view**: idealized world with *perfectly rational agents*

    - The game itself it taken to be a literal description of the strategic interaction

"We adhere to the classical point of view that the game under consideration fully describes the real situation — that any (pre) commitment possibilities, any repetitive aspect, any probabilities of error, or any possibility of jointly observing some random event, have already been modeled in the game tree." (pg. 1005)

E. Kohlberg and J.-F. Mertens. *On the strategic stability of equilibria*. Econometrica, 54, pgs. 1003 - 1038, 1986.

L. Samuelson. *Comments on Game Theory*. Game Theory: 5 Questions, Automatic Press, 2007.

Who is game theory about?

1. **Classical view**: idealized world with *perfectly rational agents*
    - The game itself it taken to be a literal description of the strategic interaction
    - Any appropriate concept of equilibrium should be an *implication* of the information provided in the modeled interpreted through an assumption of perfect rationality.
2. **Humanistic view**: real people in interactive situations

L. Samuelson. *Comments on Game Theory*. Game Theory: 5 Questions, Automatic Press, 2007.

Who is game theory about?

1. **Classical view**: idealized world with *perfectly rational agents*

   - The game itself it taken to be a literal description of the strategic interaction
   - Any appropriate concept of equilibrium should be an *implication* of the information provided in the modeled interpreted through an assumption of perfect rationality.

2. **Humanistic view**: real people in interactive situations

   - the mathematical structures are *models* of interactive situations
   - the appropriate notion of equilibrium is part of the specification of the model

L. Samuelson. *Comments on Game Theory*. Game Theory: 5 Questions, Automatic Press, 2007.

But, the game models are missing something...

Formally, a game is described by its strategy sets and payoff functions.

But, the game models are missing something...

Formally, a game is described by its strategy sets and payoff functions. But in real life, may other parameters are relevant; there is a lot more going on. Situations that substantively are vastly different may nevertheless correspond to precisely the same strategic game.

But, the game models are missing something...

Formally, a game is described by its strategy sets and payoff functions. But in real life, may other parameters are relevant; there is a lot more going on. Situations that substantively are vastly different may nevertheless correspond to precisely the same strategic game. For example, in a parliamentary democracy with three parties, the winning coalitions are the same whether the parties hold a third of the seats, or, say, 49%, 39%, and 12 % respectively.

But, the game models are missing something...

Formally, a game is described by its strategy sets and payoff functions. But in real life, may other parameters are relevant; there is a lot more going on. Situations that substantively are vastly different may nevertheless correspond to precisely the same strategic game. For example, in a parliamentary democracy with three parties, the winning coalitions are the same whether the parties hold a third of the seats, or, say, 49%, 39%, and 12 % respectively. But the political situations are quite different.

But, the game models are missing something...

Formally, a game is described by its strategy sets and payoff functions. But in real life, may other parameters are relevant; there is a lot more going on. Situations that substantively are vastly different may nevertheless correspond to precisely the same strategic game. For example, in a parliamentary democracy with three parties, the winning coalitions are the same whether the parties hold a third of the seats, or, say, 49%, 39%, and 12 % respectively. But the political situations are quite different. The difference lies in the attitudes of the players, in their expectations about each other, in custom, and in history, though the rules of the game do not distinguish between the two situations. (pg. 72)

R. Aumann and J. H. Dreze. *Rational Expectation in Games*. American Economic Review, 98, pgs. 72-86, 2008.

What about a logical analysis?

What about a logical analysis?

- Which aspects of social situations should we focus on? Knowledge, Beliefs, Group Knowledge, Preferences, Desires, Ability, Actions, Intentions, Goals, Obligations, etc.

What about a logical analysis?

- Which aspects of social situations should we focus on? Knowledge, Beliefs, Group Knowledge, Preferences, Desires, Ability, Actions, Intentions, Goals, Obligations, etc.
- One grand system, or many smaller systems that loosely "fit" together?

What about a logical analysis?

- Which aspects of social situations should we focus on? Knowledge, Beliefs, Group Knowledge, Preferences, Desires, Ability, Actions, Intentions, Goals, Obligations, etc.
- One grand system, or many smaller systems that loosely "fit" together?
- *Combining* systems is hard! (conceptually and technically)

What about a logical analysis?

- Which aspects of social situations should we focus on? Knowledge, Beliefs, Group Knowledge, Preferences, Desires, Ability, Actions, Intentions, Goals, Obligations, etc.
- One grand system, or many smaller systems that loosely "fit" together?
- *Combining* systems is hard! (conceptually and technically)
- Logics *of* rational agents in social situations.

  vs.

  Logics *about* rational agents in social situations.

What about a logical analysis?

- Which aspects of social situations should we focus on? Knowledge, Beliefs, Group Knowledge, Preferences, Desires, Ability, Actions, Intentions, Goals, Obligations, etc.
- One grand system, or many smaller systems that loosely "fit" together?
- *Combining* systems is hard! (conceptually and technically)
- Logics *of* rational agents in social situations.

  vs.

  Logics *about* rational agents in social situations.
- Normative vs. Descriptive

The point of view of this model is not normative; it is not meant to advise the players what to do. The players do whatever they do; their strategies are taken as given.

The point of view of this model is not normative; it is not meant to advise the players what to do. The players do whatever they do; their strategies are taken as given. Neither is it meant as a description of what human beings actually do in interactive situations.

The point of view of this model is not normative; it is not meant to advise the players what to do. The players do whatever they do; their strategies are taken as given. Neither is it meant as a description of what human beings actually do in interactive situations. The most appropriate term is perhaps "analytic"; it asks, what are the implications *of rationality in interactive situations? Where does it lead?*

The point of view of this model is not normative; it is not meant to advise the players what to do. The players do whatever they do; their strategies are taken as given. Neither is it meant as a description of what human beings actually do in interactive situations. The most appropriate term is perhaps "analytic"; it asks, what are the implications *of rationality in interactive situations? Where does it lead? This question may be as important as, or even more important than, more direct "tests" of the relevance of the rationality hypothesis.* (pg. 622)

R. Aumann. *Irrationality in Game Theory*. in: *Aumann's Collected Papers, Volume 1*, Chapter 35, 1992.

Ingredients of a Logical Analysis of Rational Agency

Ingredients of a Logical Analysis of Rational Agency

- What are the basic building blocks?

Ingredients of a Logical Analysis of Rational Agency

- What are the basic building blocks? (the nature of time (continuous or discrete/branching or linear), how (primitive) *events* or *actions* are represented, how *causal* relationships are represented and what constitutes a *state of affairs*.)

Ingredients of a Logical Analysis of Rational Agency

- What are the basic building blocks? (the nature of time (continuous or discrete/branching or linear), how (primitive) *events* or *actions* are represented, how *causal* relationships are represented and what constitutes a *state of affairs*.)
- Single agent vs. many agents.

Ingredients of a Logical Analysis of Rational Agency

- What are the basic building blocks? (the nature of time (continuous or discrete/branching or linear), how (primitive) *events* or *actions* are represented, how *causal* relationships are represented and what constitutes a *state of affairs*.)
- Single agent vs. many agents.
- What the the primitive operators?
  - Informational attitudes
  - Motivational attitudes
  - Normative attitudes

Ingredients of a Logical Analysis of Rational Agency

- What are the basic building blocks? (the nature of time (continuous or discrete/branching or linear), how (primitive) *events* or *actions* are represented, how *causal* relationships are represented and what constitutes a *state of affairs*.)
- Single agent vs. many agents.
- What the the primitive operators?
  - Informational attitudes
  - Motivational attitudes
  - Normative attitudes
- Static vs. dynamic

Ingredients of a Logical Analysis of Rational Agency

$\Rightarrow$ informational attitudes (eg., knowledge, belief, certainty)

$\Rightarrow$ time, actions and ability

$\Rightarrow$ evaluative/motivational attitudes (eg., preferences)

$\Rightarrow$ pro-attitudes (eg., intentions)

$\Rightarrow$ group notions (eg., common knowledge and coalitional ability)

$\Rightarrow$ normative attitudes (eg., obligations, reasons)

Next Lecture: Introduction to Epistemic Logic