

PHIL 808V: Foundations of Game Theory

Instructor:	Eric Pacuit (pacuit.org)
Semester:	Spring 2025
Email:	epacuit@umd.edu
Course Website:	https://umd.instructure.com/courses/1380924
Office:	Skinner 1103A
Office Hours:	TBA
Class Times:	Thursdays, 3:00pm - 5:30pm
Class Location:	SKN 1116

Description

Game theory is a powerful analytical tool that has been successfully applied across a wide range of disciplines, including economics, sociology, anthropology, political science, biology, and philosophy. Early research in game theory focused on developing “solution concepts” based on an intuitive understanding of “rational” behavior in strategic situations. These solutions were argued to be both predictively accurate and normatively justified to some degree. By the late 1990s, however, foundational questions began to challenge many of the basic assumptions underlying game theory. What notions of rationality are necessary for the validity of its predictions and recommendations? If human behavior diverges from these ideals, does game theory lose its relevance? Or can its core insights be adapted and reinterpreted to preserve their value?

This seminar is divided into two modules, each offering a different perspective on these foundational questions:

1. **Epistemic Game Theory:** This module presents key results that characterize classic solution concepts based on the players’ rationality and their mutual beliefs about each other’s rationality.
2. **Strategic Reasoning and Social Dynamics:** This module explores the interplay between reasoning and social dynamics in shaping strategic behavior. It combines insights from evolutionary game theory, which studies the evolution of strategies in populations, with an analysis of the processes of deliberation and reasoning that influence choices in strategic situations.

Some previous exposure to game and decision theory will be helpful, but is not required (I will do my best to provide the necessary background in game and decision theory). This is an interdisciplinary topic, and so our readings will be taken from economics, logic, philosophy and cognitive science journals.

Course work

If you are taking this seminar for credit, there are two main requirements in addition to regular attendance and completing the assigned readings:

1. **Questions:** Before each session, submit a question or comment about the readings for the week or the material discussed in the previous lecture. Your submission could include: A request for further explanation of mathematical concepts or notation, a question about the proof of a theorem we are discussing, or a philosophical question or comment related to the material from the previous or upcoming lecture. Questions or comments are due by midnight on the Tuesday prior to class, giving me time to review them beforehand. While these submissions will not be graded, you are required to submit at least 10-12 throughout the semester.
2. **Project:** The primary requirement is a project, which can take several forms. Here are some options:
 - Write a seminar paper exploring a philosophical topic related to the course material.
 - Write a more technical paper, discussing a formal result about the course material or establishing a new result.
 - Implement and analyze an agent-based model connected to topics discussed during the semester.
 - Create 5-6 handouts (approximately three pages each) on a topic related to the course material that could be used for a presentation. You would then select one handout to present, either to me or the class.

I am open to other project ideas as well. Whatever you choose, we should finalize your project plan by the 11th or 12th week of the semester to ensure you have sufficient time to complete it before the semester ends or shortly thereafter.

Schedule and Readings

The following is a tentative schedule for the course (this may change based on the interests of the students or if we need to spend more time on a particular topic). Some of the readings contain a lot of mathematics (especially the economics paper). I will explain the proofs when it is useful, otherwise we will focus on understanding the philosophical implications of the mathematical results.

Week 1 (Thursday 1/30): Introductory Remarks, A Brief Introduction to Game Theory

- M. Osborne (2004). *An Introduction to Game Theory*, Oxford University Press. The first 6 chapters are available for free at

<https://www.economics.utoronto.ca/osborne/igt/index.html>.

- M. Risse (2000). What is Rational about Nash Equilibria? *Synthese*, 124(3), pp. 361-384.

Additional Reading:

- K. Leyton-Brown and Y. Shoham (2008). *Essentials of Game Theory: A Concise, Multidisciplinary Introduction*, Morgan & Claypool.
- D. Ross (2024). “Game Theory”, The Stanford Encyclopedia of Philosophy (Winter 2024 Edition), <https://plato.stanford.edu/archives/win2024/entries/game-theory/>.
- A. Colman (2003). Cooperation, psychological game theory, and limitations of rationality in social interaction. *Behavioral and Brain Sciences*, 26, pp. 139-153.
- A. Rubinstein (1991). Comments on the Interpretation of Game Theory, *Econometrica*, 59, pp. 909-924.

Epistemic Game Theory

Primary Readings

- [EPGTH] Paolo Galeazzi, Eric Pacuit, and Olivier Roy (forthcoming). *Epistemic Game Theory*, Cambridge University Press
- [EPGTH-SEP] Eric Pacuit and Olivier Roy (2025). Epistemic Game Theory, *Stanford Encyclopedia of Philosophy* (the latest version will be available on the course website).

Week 2 (Thursday 2/6): Background: Decisions, Games, Beliefs, and Rationality

- Chapters 1 and 2 of [EPGTH]

Additional Reading:

- Sections 1 and 2 of [EPGTH-SEP]
- R. Stalnaker (1996). Knowledge, Belief and Counterfactual Reasoning in Games *Economics and Philosophy*, 12(2), pp. 133-163

Week 3 (Thursday 2/13): Nash Equilibrium and Mixed Strategies

- Chapter 3 of [EPGTH]

Additional Reading:

- R. Aumann and A. Brandenburger (1995). Epistemic Conditions for Nash Equilibrium, *Econometrica*, 63, pp. 1161-1180.
- K. Zollman (2022). On the normative status of mixed strategies, in *Reflections on the Foundations of Probability and Statistics*, Theory and Decision Library A: vol 54, pp. 207 - 239.
- T. Icard (2021). Why Be Random? *Mind*, 517, pp. 111-139.

Week 4 (Thursday 2/20): Fundamental Theorem of Epistemic Game Theory

- Chapter 4 of [EPGTH]
- J. Hillas and D. Samet (2020). Dominance rationality: A unified approach, *Games and Economic Behavior*, 119, pp. 189 - 196.

Additional Reading:

- Section 3.1 of [EPGTH-SEP]
- K. Apt (2007). The Many Faces of Rationalizability, *The B.E. Journal of Theoretical Economics*, 7(1), Article 18.
- K. Apt and J. Zvesper (2010). The Role of Monotonicity in the Epistemic Analysis of Strategic Games, *Games*, 1(4), pp. 381 - 394.
- L. Samuelson, Dominated strategies and common knowledge, *Game and Economic Behavior*, 4(2), pp. 284 - 313.

Week 5 (Thursday 2/27): Correlated Equilibrium

- Chapter 5 of [EPGTH]
- R. Aumann (1987). Correlated Equilibrium as an Expression of Bayesian Rationality, *Econometrica*, 55, pp. 1 - 18.

Additional Reading:

- A. Brandenburger and A. Friedenberg (2008). Intrinsic Correlation in Games, *Journal of Economic Theory*, 141, pp. 28 - 67.
- P. Vanderschraaf (1995). Convention as correlated equilibrium, *Erkenntnis*, 42(1), pp. 65 - 87.

Week 6 (Thursday 3/6): Sequential Games and Backward Induction

- Chapter 6 of [EPGTH]
- J. Halpern (2001), Substantive rationality and backward induction, *Games and Economic Behavior*, 37, pp. 425 - 435.

Additional Reading:

- R. Aumann (1995). Backward Induction and Common Knowledge of Rationality, *Games and Economic Behavior*, 8, pp. 6-19.
- R. Aumann (1998). On the Centipede Game, *Games and Economic Behavior*, 23, pp. 97-105.
- R. Stalnaker (1998). Belief Revision in Games: Forward and Backward Induction, *Mathematical Social Sciences*, 36, pp. 31 - 56.
- D. Samet (2013), Common belief of rationality in games of perfect information, *Games and Economic Behavior*, 79, pp. 192 - 200.

Week 7 (Thursday 3/13): Levels of Higher-Order Reasoning

- Chapter 7 of [EPGTH]
- J. Wright and K. Leyton-Brown (2019). Level-0 Models for Predicting Human Behavior in Games, *Journal of Artificial Intelligence Research*, 64, pp. 357-383.

Additional Reading:

- D. O. Stahl and P. W. Wilson (1995). On players models of other players: Theory and experimental evidence, *Games and Economic Behavior*, 10, pp. 218 - 254.
- T. Hedden and J. Zhang (2002). What do you think I think you think?: Strategic reasoning in matrix games, *Cognition*, 85, pp. 1 - 36.
- B. Meijering, H. van Rijn, N.A. Taatgen, and R. Verbrugge (2011), I do know what you think I think: Second-order theory of mind in strategic games is not that difficult. In *Proceedings of the 33rd Annual Conference of the Cognitive Science Society*, Cognitive Science Society, Austin, TX, pp. 2486 - 2491.
- B. Meijering, H. van Rijn, N.A. Taatgen and R. Verbrugge (2012). What eye movements can tell about theory of mind in a strategic game, *PLoS ONE*, 7:9.

Week 8 (Thursday 3/20): Spring Break - No Lecture

Strategic Reasoning and Social Dynamics

Week 9 (Thursday 3/27): Fairness

- B. Skyrms (2009). Evolution and the Social Contract, in *The Tanner Lectures on Human Values*, 28, pp. 47-69 (available as Chapter 1 in *Social Dynamics* by Brian Skyrms)
- B. Skyrms (2008). Trust, Risk, and the Social Contract, *Synthese*, 160, pp. 21-25 (available as Chapter 2 in *Social Dynamics* by Brian Skyrms)

- B. Skyrms and J. Alexander (1999). Bargaining with Neighbors: Is Justice Contagious? *Journal of Philosophy*, 96, pp. 588-598 (available as Chapter 3 in *Social Dynamics* by Brian Skyrms)

Additional Reading:

- K. Zollman (2008). Explaining fairness in complex environments. *Politics, Philosophy & Economics*, 7(1), pp. 81-97.

Week 10 (Thursday 4/3): Unfairness

- C. O'Connor (2024). Why Natural Social Contracts Are Not Fair, in Michael Moehler, and John Thrasher (eds), *New Approaches to Social Contract Theory: Liberty, Equality, Diversity, and the Open Society*, Oxford University Press.
- J. Bruner (2019). Minority (dis)advantage in population games, *Synthese*, 196, pp. 413 - 427.

Additional Reading:

- C. O'Connor (2019). *The Origins of Unfairness: Social Categories and Cultural Evolution*, Oxford University Press.
- J. Bruner and C. O'Connor (2018). Power, Bargaining, and Collaboration. in *Scientific Collaboration and Collective Knowledge*, Oxford University Press, pp. 135 - 158.

Week 11 (Thursday 4/10): Cooperation and the Prisoner's Dilemma

- J. Bruner (2024). How mixed strategies make a difference in the one-shot prisoner's dilemma, *Analysis*, <https://doi.org/10.1093/analys/anad098>
- J. Barrett (2015). The Prisoner's Dilemma and the coevolution of descriptive and predictive dispositions, in *The Prisoner's Dilemma*, M. Peterson (ed.), Cambridge University Press, pp. 85 - 100.
- S. Kuhn and E. Pacuit (2025). Advocacy Games, manuscript.

Additional Reading:

- V. Capraro and J. Halpern (2019). Translucent players: explaining cooperative behavior in social dilemmas, *Rationality and Society*, 31, pp. 371-408.
- V. Capraro, J. Halpern, and M. Perc (2024). From outcome-based to language-based preferences, *Journal of Economic Literature*, 62(1), pp. 115 - 154.

Week 12 (Thursday 4/17): Deliberational Decision Theory

- S. Huttegger (2023). Reconciling Causal and Evidential Decision Theory, *Philosopher's Imprint*, 23:20 (2023)

- H. S. Shin (1989). Two Notions of Ratifiability and Equilibrium in Games, in *Foundations of Decision Theory*, Blackwell.

Additional Reading:

- C. Hare and B. Hedden (2015). Self-Reinforcing and Self-Frustrating Decisions, *Noûs*, 50, pp. 604 - 628.
- B. Armendt (2019). Causal Decision Theory and Decision Instability, *The Journal of Philosophy*, 116, pp. 263 - 277.
- W. Harper (1986). Mixed Strategies and Ratifiability in Causal Decision Theory, *Erkenntnis*, 24:1, pp. 25 - 36.

Week 13 (Thursday 4/24): Strategic Reasoning I

- B. Skyrms (1990). Chapter 2 “Dynamic Deliberation: Equilibria” and Chapter 3 “Dynamic Deliberation: Stability” in *The Dynamics of Rational Deliberation*, Harvard University Press.
- J. McKenzie Alexander (2010). Local Interactions and the Dynamics of Rational Deliberation, *Philosophical Studies*, 147, pp. 102 - 121.
- P. Vanderschraaf and B. Skyrms (2003). Learning to Take Turns, *Erkenntnis*, 59(3), pp. 311-348.

Additional Reading:

- R. Jeffrey (1992). Review of the dynamics of rational deliberation by Brian Skyrms, *Philosophy and Phenomenological Research*, 52(3), pp. 734 - 737.
- J. McKenzie Alexander (2009). Social Deliberation: Nash, Bayes, and the Partial Vindication of Gabriele Tarde, *Episteme*, 6(2): pp. 164 - 184.
- E. Pacuit (2015). Dynamic Models of Rational Deliberation in Games, in *Models of Strategic Reasoning*, LNCS 8972, pp. 3 - 33.

Week 14 (Thursday 5/1): Strategic Reasoning II: Forward Induction Reasoning

- A. Perea and M. Meier (2023). Forward induction in a backward inductive manner.
- E. Pacuit and A. Knoks (2018). Deliberational dynamics in context, *LOFT 2018*, <https://pacuit.org/static/4458f1ff849e1979b219c529290adb64/fi-bi-skyrms-delib.pdf>.

Additional Reading:

- A. Perea (2010). Backward Induction versus Forward Induction Reasoning, *Games*, 1, pp. 168-188.
- P. Battigalli, E. Catonini, and J. Manili (2022). Belief Change, Rationality, and Strategic Reasoning in Sequential Games. available at SSRN: <https://ssrn.com/abstract=4206944>.

- M. Mariotti (1995). Is bayesian rationality compatible with strategic rationality? *The Economic Journal*, 105(432), pp. 1099 - 1109.

Week 15 (Thursday 5/8): Strategic Reasoning III: Large Language Models

- K. Gandhi, D. Sadigh, and N. D. Goodman (2023). Strategic Reasoning with Language Models, <https://arxiv.org/abs/2305.19165>.

Additional Reading:

- F. Guo (2023). GPT in Game Theory Experiments, <https://arxiv.org/abs/2305.05516>.
- Y. Zhang, S. Mao, T. Ge, X. Wang, A. de Wynter, Y. Xia, W. Wu, T. Song, M. Lan, and F. Wei (2024). LLM as a Mastermind: A Survey of Strategic Reasoning with Large Language Models, <https://arxiv.org/abs/2404.01230>.
- T. Kojima, S. S. Gu, M. Reid, Y. Matsuo, and Y. Iwasawa (2022). Large Language Models are Zero-Shot Reasoners. *Advances in Neural Information Processing Systems*, 35, pp. 22199-22213.