PHI858K: Logic and Probabilistic Models of Belief Change

Instructor:	Eric Pacuit (pacuit.org)
Semester:	Spring 2016
Email:	epacuit@umd.edu
Course Website:	myelms.umd.edu/courses/1181012
Office:	Skinner 1103A
Office Hours:	W 2 - 3.30 PM
Class Times:	Th 4:30 - 7:00 PM
Class Location:	SKN 1116

Course Description

Reasoning about the knowledge and beliefs of a single agent or group of agents is an interdisciplinary concern spanning computer science, game theory, philosophy, linguistics and statistics. Inspired, in part, by issues in these different "application" areas, many different notions of knowledge and belief have been identified and analyzed in the formal epistemology literature. The main challenge is not to argue that one particular account of belief or knowledge is *primary*, but, rather, to explore the logical space of definitions and identify interesting relationships between the different notions. A second challenge (especially for students) is to keep track of the many different formal frameworks used in this broad literature (typical examples include modal logics of knowledge and belief, the theory of subjective probability, but there are many variants, such as the Dempster-Shafer belief functions and conditional probability systems). This foundational course will introduce students to key methodological, conceptual and technical issues that arise when designing a formalism to make precise intuitions about the beliefs of a group of agents, and how these beliefs may change over time. There are two central questions that I will address is this course: 1. What is the precise relationship between the different formalisms describing an agent's beliefs (e.g., what is the relationship between an agent's graded beliefs and full beliefs?); and 2. How should a agent change her beliefs in response to new evidence?

In this course, I will introduce the main formalisms that can describe an agents' beliefs and how those beliefs change over time. Rather than focusing solely on the technical details of a specific formalism, I will pay special attention to the key foundational questions (of course, introducing formal details as needed). There are good reasons for taking an "issue-oriented" approach to introducing formal models of belief and belief change (especially at a summer school such as NASSLLI). Many of the recent developments concerning formal models of belief have been driven by analyzing concrete examples. These range from toy examples, such as the infamous muddy children puzzle or the Monte hall dilemma to philosophical quandaries, such as the lottery and preface paradoxes, to everyday examples

of social interaction. Different formal frameworks are then judged, in part, on how well they conform to the analyst's intuitions about the relevant set of examples. Thus, in order to appreciate the usefulness and limits of the different formal frameworks, it is important to understand the issues that motivate the key technical developments.

The topics that will be discussed in this course include:

1. Formal models of belief: Modal logics of knowledge and belief, Subjective probability
2. Connections between graded beliefs and full beliefs (Leitgeb's stability theory of beliefs [18])
3. Problems for modeling beliefs: lottery paradox and preface paradox [15]
4. Updating probabilities: Conditioning, Jeffrey Conditioning, Adam's Conditioning, minimizing Kullback-Leibler distance [5, 6, 11, 13, 19]
5. Dealing with probability 0 events: Conditional probability systems, lexicographic probability systems, non-standard probabilities [14]
6. Problems for updating (graded) beliefs: The Judy Benjamin problem [11, 12]
7. Problems for the theory of belief revision: counterexamples to the AGM principles, Gärdenfors' triviality result (incompatibility of the Ramsey test and the AGM postulates) [21]
8. Iterated belief revision [4, 18, 21]
9. Problems involving beliefs and time: Absent-Minded Driver and the sleeping beauty problem
10. Belief change in a social environment: Aumann's agreeing to disagree theorem and its generalizations, opinion pooling [8] (e.g., the Lehrer-Wagner model, de Groot's Theorem)

Of course, there is a large literature concerning each of the issues listed above. The challenge for us will be to discuss each problem and formalism without getting distracted by extraneous technical details.[1] In order to facilitate the discussion, I will produce a reader containing the relevant technical details, a precise statement of each puzzle and paradox, statements of the key theorems discussed in the course, and references to the main literature.

[1]This is not to say that such details are not *interesting*. Many of the problems and formalisms listed above raise very interesting technical questions for logicians and mathematicians. However, this course will focus on the underlying conceptual issues.

Readings

Itzhak Gilboa, Larry Samuelson and David Schmeidler (2013). Dynamics of Inductive Inference in a Unified Model, *Journal of Economic Theory*, 148:4, pp. 1399 - 1432.

H. Greaves and D. Wallace (2006). Justifying conditionalization: Conditionalization maximizes expected epistemic utility, *Mind*. 115, pp. 607 - 632.

Maher, P. (1992). Diachronic rationality. Philosophy of Science 59, 120141.

Wagner, C. G. (2002). Probability kinematics and commutativity. Philosophy of Science 69, 266278.

Gordon Belot (2013). Bayesian Orgulity, Philosophy of Science 80:483503.

David Blackwell and Lester Dubins (1962). Merging of Opinions with Increasing Information, The Annals of Mathematical Statistics, 33:882886.

Simon Huttegger (2015). Merging Of Opinions and Probability Kinematics, Review of Symbolic Logic (forthcoming)

Simon Huttegger (2015). Bayesian Convergence to the Truth and the Metaphysics of Possible Worlds, Philosophy of Science (forthcoming)

A. Darwiche and J. Pearl (1997). On the logic of iterated belief revision. Artificial Intelligence, 89(12):129.

P. Diaconis, P. and S. Zabell (1982). Updating subjective probability, Journal of the American Statistical Association 77, pp. 822-830.

F. Dietrich, C. List, and R. Bradley (2015). A Unified Characterization of Belief-Revision Rules, Manuscript.

F. Dietrich and C. List (2014). From Degrees of Belief to Beliefs: Lessons from Judgment-Aggregation Theory, Manuscript.

H. Gaifman and A. Vasudevan (2012). Deceptive Updating and Minimal Information Methods. Synthese 187, pp. 147178.

P. Grünwald and J. Halpern (2003). Updating probabilities, Journal of AI Research 19: 243-278.

J. Y. Halpern (2010). Lexicographic probability, conditional probability, and nonstandard probability. Games and Economic Behavior, 68(1):155 179.

H. Katsuno and A. Mendelzon (1991). On the difference between updating a knowledge base and revising it, In *Principles of Know ledge Representation and Reasoning (KR)*

J. W. Romeijn (2012). Conditioning and Interpretation Shifts, Studia Logica, 100(3), pp. 583-606.

R. Stalnaker (2009). Iterated belief revision. Erkentnis, 70:189 209.

J. Zhao, V. Crupi, K. Tentori, B. Fitelson and D. Osherson (2012). Updating: Learning versus supposing, Cognition 124, pgs. 373 - 378.

Lara Buchak (2014). Belief, credence, and norms, *Philosophical Studies*, 169:2, pp. 285-311.

Wolfgang Schwarz (2011). Changing minds in a changing world, *Philosophical Studies*, 159:2, pp. 219-239.

Alan Hájek (2003). What Conditional Probability Could Not Be, *Synthese*. 137:3, pp. 273-323.

Ted Shear and Branden Fitelson (2016). Two Approaches to Belief Revision, manuscript.

Hanti Lin and Kevin Kelly (2012). Propositional Reasoning that Tracks Probabilistic Reasoning. Journal of Philosophical Logic, 41:6, pp. 957-981.

Hannes Leitgeb (2013). The review paradox: On the diachronic costs of not closing rational belief under conjunction. Nous, 78:4, pp. 781793.

Wolfgang Schwarz (2015). Lost memories and useless coins: revisiting the absentminded driver, *Synthese*, 192:9, pp. 3011 - 3036.

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: Th 1/28 Introductory Remarks and Background: Formal Models of Beliefs
Week 2: Th 2/4 Formal Models of Beliefs
Week 3: Th 2/11 Belief Revision
Week 4: Th 2/18 Updating Probabilities
Week 5: Th 2/25 Stability Theory of Belief, I
Week 6: Th 3/3 Stability Theory of Belief, II
Week 7: Th 3/10 Introductory Remarks and Background
Week 8: Th 3/17 **Spring Break**: No Class
Week 9: Th 3/24 Introductory Remarks and Background
Week 10: Th 3/31 Introductory Remarks and Background
Week 11: Th 4/7 Introductory Remarks and Background
Week 12: Th 4/14 Introductory Remarks and Background
Week 13: Th 4/21 Introductory Remarks and Background
Week 14: Th 4/28 Introductory Remarks and Background
Week 15: Th 5/5 Introductory Remarks and Background

- J. Kadane and P. Larkey, Subjective Probability and the Theory of Games, *Management Science*, 28: 2, 1982, pgs. 113-120. In addition, take a look at the back-and-forth with Harsanyi:
  - J. Harsanyi, Subjective Probability and the Theory of Games: Comments on Kadane and Larkey's Paper, pgs. 120-124
  - J. Kadane and P. Larkey, Reply to Professor Harsanyi, pg. 124
  - J. Harsanyi, Rejoinder to Professors Kadane and Larkey, pgs. 124 - 125
- E. McClennen, Rational Choice in the Context of Ideal Games, in *Knowledge, Belief and Strategic Interaction*, C. Bicchieri and M. L. Dalla Chiara (eds.), Cambridge University Press, 1992.

Additional readings

- B. Skyrms, Principles of Rational Decision, Chapter 1 of *The Dynamics of Rational Deliberation*, Harvard University Press, 1990.
- M. Mariotti, Is Bayesian Rationality Compatible with Strategic Rationality? The Economics Journal, 105:432, pgs. 1099 - 1109, 1995.

Week 3: Mon 2/10 Common Knowledge/Belief of Rationality

- D. Monderer and D. Samet. Approximating common knowledge with common beliefs, Games and Economic Behavior 1, pgs. 170 - 190, 1989.
- R. Cubitt and R. Sugden. Common Knowledge, Salience And Convention: A Reconstruction Of David Lewis' Game Theory, Economics and Philosophy, 19: 2, pgs. 175 - 210, 2003.
- D. O. Stahl and P. W. Wilson. On players models of other players: Theory and experimental evidence, Games and Economic Behavior, 10, pgs. 218 - 254, 1995.

Additional readings

- T. Hedden and J. Zhang. What do you think I think you think?: Strategic reasoning in matrix games. Cognition 85, 1 - 36, 2002.
- B. Meijering, H. van Rijn, N.A. Taatgen, and R. Verbrugge, 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, pgs. 2486 - 2491, 2011.
- B. Meijering, H. van Rijn, N.A. Taatgen and R. Verbrugge, What eye movements can tell about theory of mind in a strategic game. PLoS ONE, 7:9, 2012.
- Z. Ernst, What Is Common Knowledge? Episteme, Volume 8, Issue 03, pgs. 209 - 226, 2011.

Week 4: Mon 2/17 Nash and Correlated Equilibrium

- R. Aumann, Correlated Equilibrium as an Expression of Bayesian Rationality, Econometrica, 55, pgs. 1-18, 1987.
- R. Aumann and A. Brandenburger, Epistemic Conditions for Nash Equilibrium, *Econometrica* 63, pgs. 1161-1180, 1995.
- M. Risse, What Is Rational about Nash Equilibria? Synthese, Vol. 124, No. 3, pgs. 361-384, 2000.

Additional readings

- R. Aumann Subjectivity and Correlation in Randomized Strategies, Journal of Mathematical Economics, 1(1): pgs. 67-96, 1974.
- A. Brandenburger and A. Friedenberg, Intrinsic Correlation in Games, Journal of Economic Theory, 141, pgs. 28-67, 2008.
- A. Brandenburger, The Relationship Between Quantum and Classical Correlation in Games, Games and Economic Behavior, 69, pgs. 175-183, 2010.

Week 5: Mon 2/24 Deliberation in Game and Decision Theory

- I. Levi, Rationality Prediction and Autonomous Choice, in *The Covenant of Reason*, Cambridge University Press, 1997.
- I. Levi, Prediction, Deliberation and Correlated Equilibrium, in *The Covenant of Reason*, Cambridge University Press, 1997.
- W. Rabinowicz. Does Practical Deliberation Crowd Out Self-Prediction? Erkenntnis, 57, pgs. 91 - 122, 2002.
- I. Levi, Deliberation does crowd out prediction, *Hommage à Wlodek: Philosophical Papers Dedicated to Wlodek Rabinowicz*, 2007.

Additional readings

- J. Joyce, Levi on Decision Theory and the Possibility of Predicting One's Own Actions, Philosophical Studies 110, pgs. 69 - 102, 2002.
- W. Spohn, Where Luce and Krantz Do Really Generalize Savage's Decision Model, Erkenntnis, 11, 113-134, 1977
- B. Skyrms, Chapter 7 "Prospects for a Theory of Rational Deliberation" in *The Dynamics of Rational Deliberation*, 1990

Week 6: Mon 3/3 Class canceled: I will be giving a talk at the Royal Netherlands Academy of Arts and Sciences Colloquium on Dependence Logic. We can try to reschedule this class.

Week 7: Mon 3/10 Ratifiiability in Game Theory

- J. Joyce and A. Gibbard. "Causal Decision Theory" In Salvador Barbera, Peter Hammond, and Christian Seidl, eds., *Handbook of Utility Theory*, Kluwer Academic Publishers, pgs. 627 - 666, 1998. (Focus on the section on Ratifiability in Game Theory)
- W. Harper, Mixed Strategies and Ratifiability in Causal Decision Theory, Erkenntnis, 24:1, pgs. 25 - 36, 1986.

Additional readings

- H. S. Shin, Two Notions of Ratifiability and Equilibrium in Games, in M. Bacharach and S. Hurley (eds.), *Foundations of Decision Theory*, Blackwell, 1989.
- B. Skyrms, Ratifiability and the Logic of Decision, Midwest Studies In Philosophy, Volume 15, Issue 1, pgs. 44 - 56, 1990.
- E. Eells and W. Harper. Ratifiability, game theory, and the principle of independence of irrelevant alternatives, Australasian Journal of Philosophy, 69:1, pgs. 1-19, 1991.

Week 8: Mon 3/17 No Class: Spring Break

Week 9: Mon 3/24 Skyrms' Model of Deliberation in Games

- B. Skyrms, Chapter 2 "Dynamic Deliberation: Equilibria" and Chapter 3 "Dynamic Deliberation: Stability" in *The Dynamics of Rational Deliberation*, Harvard University Press, 1990.
- J. McKenzie Alexander, Local Interactions and the Dynamics of Rational Deliberation, Philosophical Studies, vol. 147, pgs. 102 - 121, 2010.

Additional readings

- R. Jeffrey Review of the dynamics of rational deliberation by Brian Skyrms. Philosophy and Phenomenological Research 52(3), pgs. 734 - 737, 1992.
- J. McKenzie Alexander, Social Deliberation: Nash, Bayes, and the Partial Vindication of Gabriele Tarde, Episteme, 6(2): pgs. 164 - 184, 2009.

Other models of deliberation in games

- R. Cubitt and R. Sugdent, Common Reasoning in Games, working paper, 2012.
- E. Pacuit, Models of Deliberation in Game Theory, manuscript, 2013.

- R. Cubitt and R. Sugden, The reasoning-based expected utility procedure, Games and Economic Behavior, 71(2), pgs. 328 - 338, 2011.

Week 10: Mon 3/31 Common Belief of Rationality, Rationalizability and Iterated Removal of Strictly/Weakly Dominated Strategies

- D. Samet, Weakly dominated strategies: A mystery cracked, Last revision: October, 2013.
- K. Apt, The Many Faces of Rationalizability. The B.E. Journal of Theoretical Economics, 7(1), (Topics), Article 18, 2007.
- K.R. Apt and J.A. Zvesper, The Role of Monotonicity in the Epistemic Analysis of Strategic Games, Games 1(4), pgs. 381 - 394, 2010.

Additional readings

- L. Samuelson, Dominated strategies and common knowledge. Game and Economic Behavior 4, pgs. 284 313, 1992.
- A. Brandenburger, J. Keisler, A. Friedenberg, Admissibility in Games, Econometrica, Vol. 76, pgs. 307-352, 2008.

Week 11: Mon 4/7 Backwards Induction and Common Knowledge of Rationality

- J. Halpern, Substantive rationality and backward induction, Games and Economic Behavior 37, pgs. 425-435.
- D. Samet, Common belief of rationality in games of perfect information, Games and Economic Behavior, 79, 2013.
- K. Binmore, Interpreting Knowledge in the Backward Induction Problem, Episteme, 8:3, pgs. 248 - 261, 2011.
- A. Baltag, S. Smets and J. Zvesper. Keep 'hoping' for rationality: a solution to the backward induction paradox, Synthese. Volume 169, Number 2, pgs. 301-333, 2009.

Additional readings

- R. Aumann, On the Centipede Game, Games and Economic Behavior 23, pgs. 97-105, 1998.
- R. Aumann, Backward Induction and Common Knowledge of Rationality, Games and Economic Behavior 8, pgs. 6-19, 1995.

- R. Stalnaker, Belief Revision in Games: Forward and Backward Induction, Mathematical Social Sciences, 36, pgs. 31 - 56, 1998.
- J. Kadane and T. Seidenfeld, Equilibrium, Common Knowledge, and Optimal Sequential Decisions, in *Knowledge, Belief and Strategic Interaction*, pgs. 27 - 45, 1992.
- K. Binmore, J. McCarthy, G. Ponti, L. Samuelson, and A. Shaked, A Backward Induction Experiment, Journal of Economic Theory 104, pgs. 48 - 88, 2002.

Week 12: Mon 4/14 Forward Induction Reasoning

- J. van Benthem, Logic in a Social Setting, Episteme, 8:3, pgs. 227-247, 2011.
- A. Perea, Backward Induction versus Forward Induction Reasoning, Games, 1, pgs. 168-188, 2010.

Additional readings

- P. Battigalli and A. Friedenberg, Forward Induction Reasoning, Revisited, Theoretical Economics Volume 7, Issue 1, pgs. 57 - 98, 2012.

Week 13: Mon 4/21 Counterfactual Reasoning in Game Theory

- R. Stalnaker. Knowledge, belief and counterfactual reasoning in games. Economics and Philosophy, 12:133163, 1996.
- G. Bonanno, Counterfactuals and the Prisoners Dilemma, manuscript, 2013
- C. Bicchieri, Strategic Behavior and Counterfactuals, Synthese 76, pgs. 135 - 69, 1988.

Additional readings

- B. Skyrms, Subjunctive Conditionals and Revealed Preference, Philosophy of Science 65, pgs. 545-574, 1998.
- E. Zambrano. Counterfactual reasoning and common knowledge of rationality in normal form games. Topics in Theoretical Economics, 4:Article 8, 2004.
- O. Board. The equivalence of Bayes and causal rationality in games. Theory and Decision, 61:119, 2006.

Week 14: Mon 4/28 Language and Game Theory

- A. Bjorndahl, J. Halpern and R. Pass, Language-based games, Proceedings of the Fourteenth Conference on Theoretical Aspects of Rationality and Knowledge, 2013, pgs. 39 - 48, 2013.
- A selection of readings from A. Rubinstein's *Economics and Language*

Additional readings

- Bacharach, M. (1993). Variable universe games. In K. BInmore, A. Kirman, and P. Tami (Eds.), Frontiers of Game Theory, pp. 255 275. The MIT Press.

Week 15: Mon 5/5 What is Game Theory Trying to Accomplish?

- J. Kadane and P. Larkey, The Confusion of Is and Ought in Game Theoretic Contexts, Management Science, 29:12, pgs. 1365 - 1379, 1983.
- R. Aumann, What Is Game Theory Trying to Accomplish?, in *Frontiers of Economics*, edited by K. Arrow and S. Honkapohja, Basil Blackwell, Oxford, 1985, pp. 28-76.

Additional readings

- W. Spohn, How to Make Sense of Game Theory, in: W. Stegmüller, W. Balzer, W. Spohn (eds.), Philosophy of Economics, Springer, Berlin, pgs. 239 - 270, 1982.
- C. Bicchieri, Rationality and Predictability in *Rationality and Coordination*, Cambridge University Press, 1993

Week 16: Mon 5/12 Interpreting Game Theoretic Models

- A. Colman, Cooperation, psychological game theory, and limitations of rationality in social interaction. Behavioral and Brain Sciences, 26, pgs. 139-153, 2003.
- A. Rubinstein, Comments on the Interpretation of Game Theory, Econometrica, 59, 909-924, 1991.

Additional readings

- I. Gilboa, A. Poslewaite, L. Samuelson and D. Schmeidler, Economic Models as Analogies, Economic Journal, 2013.
- F. Dietrich and C. List, Mentalism versus behaviourism in economics: a philosophy of science perspective, 2012.
- K. Arrow, Mathematical models in the social sciences. In D. Lerner and H. Lasswell (Eds.), The Policy Sciences. Stanford University Press, 1951.