CSC 720: Artificial Intelligence II

Readings

Jon Doyle
Department of Computer Science
North Carolina State University
Spring 2011

Reading sets

The following lists the contents of each of the fourteen reading sets. The main readings are listed first, followed by supplementary and reference readings. Please read the main readings, and read supplementary readings as your interests lead you. The reference readings provide additional background.

Reference texts:

Primary
Knowledge Representation and Reasoning
Ronald Brachman and Hector Levesque
San Francisco: Morgan Kaufmann, 2004, ISBN 1558609326.

Secondary
Artificial Intelligence: A Modern Approach (3rd ed.)
Stuart J. Russell and Peter Norvig
New York: Prentice Hall, 2010, ISBN 0-13-604259-7.

Contents

1 Set 1: The big problems of artificial intelligence

- Doyle and Dean, eds., Strategic directions in artificial intelligence [42]

- McDermott, Kurzweil's argument for the success of AI [109]

- Goertzel, Human-level artificial general intelligence and the possibility of a technological singularity: A reaction to Ray Kurzweil's *The Singularity is Near*, and McDermott's critique of Kurzweil [64]

- William Swartout, Jonathan Gratch, Randall W. Hill, Eduard Hovy, Stacy Marsella, Jeff Rickel, and David Traum, Toward virtual humans [153]

- Penn, Holyoak, and Povinelli, Darwin's mistake: explaining the discontinuity nonhuman minds [128] (The attached commentaries and response are optional.)

- **Supplementary**

    - Doyle, Big problems for artificial intelligence [33]
    - Doyle, Rationality and its roles in reasoning [36]
    - Minsky, Steps toward artificial intelligence [114]
    - Kurzweil, The law of accelerating returns [88]
    - Goertzel, Artificial general intelligence: now is the time [63]
    - McCarthy, From here to human-level AI [106]
    - McDermott, Level-headed [108]
    - Nilsson, Human-level artificial intelligence? Be serious! [126]
    - Karen Myers, Pauline Berry, Jim Blythe, Ken Conley, Melinda Gervasio, Deborah McGuinness, David Morley, Avi Pfeffer, Martha Pollack, and Milind Tambe, An intelligent personal assistant for task and time management [123]

- **Reference**

    - Brachman & Levesque, Chapter 1. Introduction
    - Russell & Norvig, Chapter 2. Intelligent agents
    - Russell & Norvig, Chapter 27. AI: present and future

2   Set 2: Logical rationality

- McCarthy, Programs with common sense [105]

- Halpern, Reasoning about knowledge: a survey [74]

- Doyle, Rational belief revision [35]

- **Supplementary**

  – McCarthy & Hayes, Some philosophical problems from the standpoint of artificial intelligence [104]

  – Minsky, A framework for representing knowledge (appendix) [118]

  – McDermott, Critique of pure reason [107]

  – Scott, Domains for denotational semantics [140]

  – Hayes, Computation and deduction [78]

  – Hayes, Some problems and non-problems in representation theory [79]

  – Hayes, In defence of logic [80]

  – Doyle, How to frame it [31]

- **Reference**

  – Brachman & Levesque, Chapter 2. The language of first-order logic

  – Russell & Norvig, Chapter 7. Logical agents

  – Kalish and Montague, *Logic: Techniques of Formal Reasoning* [85]

  – Barwise, An introduction to first-order logic [3].

  – Barwise, Model-theoretic logics: background and aims [4]

3   Set 3: Economic rationality

- Doyle & Wellman, Impediments to universal preference-based default theories [46]

- Machina, Choice under uncertainty: problems solved and unsolved [100]

- Van Fraassen, Values and the heart's command [159]

- Stigler & Becker, De gustibus non est disputandum [151] (This paper is tough going for the non-economist, but please look it over.)

- **Supplementary**

    - Horvitz, Cooper & Heckerman, Reflection and action under scarce resources: theoretical principles and empirical study [82]
    - Gärdenfors & Sahlin, Bayesian decision theory—foundations and problems [58]
    - Simon, A behavioral model of rational choice [146]
    - Simon, Rational choice and the structure of the environment [147]
    - I. J. Good, A five-year plan for automatic chess [66]
    - I. J. Good, Twenty-seven principles of rationality [67]
    - Simon & Kadane, Optimal problem-solving search: all-or-none solutions [148]
    - Feldman & Sproull, Decision theory and artificial intelligence II: the hungry monkey [52]
    - Machina, Dynamic consistency and non-expected utility models of choice under uncertainty [101]
    - Russell & Wefald, Principles of metareasoning [134]
    - Thomason, The context-sensitivity of belief and desire [156]

- **Reference**

    - Russell & Norvig, Chapter 13. Uncertainty
    - Russell & Norvig, Chapter 16. Making simple decisions
    - von Neumann & Morgenstern, *Theory of Games and Economic Behavior* [160]
    - Luce & Raiffa, *Games and Decisions* [98]
    - Savage, *The Foundations of Statistics* [135]
    - Arrow, Values and collective decision-making [1]
    - Debreu, *Theory of Value: an axiomatic analysis of economic equilibrium* [26]
    - Sen, The impossibility of a Paretian liberal [142]
    - Tullock, The general irrelevance of the general impossibility theorem [158]
    - Gärdenfors & Sahlin, *Decision, Probability, and Utility: Selected Readings* [57]
    - Roberts, *Discrete Mathematical Models* [130, Sections 7.1 and 7.2]

4   Set 4: Psychological rationality

- Todd and Gigerenzer, Bounding rationality to the world [157]

- Newell, Précis of *Unified Theories of Cognition* [125]

- Minsky, The society theory of thinking [116]

- Cassimatis, A cognitive substrate for achieving human-level intelligence [15]

- Doyle, Reason maintenance and belief revision [37]

- **Supplementary**

  - Minsky, K-lines: a theory of memory [117]
  - Doyle, Reasoned assumptions and rational psychology [38]
  - Baron, Introduction to *Rationality and Intelligence* [2]
  - Doyle, A truth maintenance system [27]
  - Doyle, Some theories of reasoned assumptions [30]
  - Minsky, Problems of formulation for artificial intelligence [113]

- **Reference**

  - Minsky, *The Society of Mind* [119]
  - Minsky, *The Emotion Machine* [120]

5 Set 5: Reflection, consciousness, personhood, and self-government

- Minsky, Matter, mind and models [115]

- Frankfurt, Freedom of the will and the concept of a person [55]

- Cox, Metacognition in computation: A selected research review [19]

- Thaler & Schefrin, An economic theory of self-control [155]

- **Supplementary**

  - Searle, How to study consciousness scientifically [141]
  - Lipman, How to decide how to decide how to . . .: limited rationality in decisions and games [95]
  - Kydland & Prescott, Rules rather than discretion: the inconsistency of optimal plans [89]
  - Cox, Perpetual self-aware cognitive agents [20]
  - Doyle, The foundations of psychology [29]
  - Doyle, Artificial intelligence and rational self-government [32]
  - Doyle, Reasoning, representation, and rational self-Government [34]
  - Ginsberg, The computational value of nonmonotonic reasoning [60]
  - Ginsberg, Is there any need for domain-dependent control information? [62]
  - Shrobe and Doyle, Active trust management for autonomous adaptive survivable systems [145]
  - Schelling, The intimate contest for self-command [137]
  - Schelling, The mind as a consuming organ [138]
  - Becker, *The Economic Approach to Human Behavior* [5]
  - Doyle, A reasoning economy for planning and replanning [39]
  - Doyle, A model for deliberation, action, and introspection [28]

6 Set 6: Representing knowledge I: logic, descriptions, and propositional constraints

- Davis, Shrobe, & Szolovits, What is a knowledge representation? [23]

- McAllester and Givan, Natural language syntax and first order inference [103]

- Gomes and Selman, Can get satisfaction [65]

- Williams, Gomes, and Selman, Backdoors to typical case complexity [173]

- **Supplementary**

  - Brachman & Levesque, Chapter 9. Structured descriptions
  - Doyle and Patil, Two theses of knowledge representation: language restrictions, taxonomic classification, and the utility of representation services [43]
  - Lenat, CYC: a large-scale investment in knowledge infrastructure [91]
  - Genesereth and Fikes, Knowledge interchange format, version 3.0 reference manual [59]
  - Ginsberg, Knowledge interchange format: the KIF of death [61]
  - Neches, Fikes, Finin, Gruber, Patil, Senator, and Swartout, Enabling technology for knowledge sharing [124]
  - Kautz and Selman, The state of SAT [86]

- **Reference**

  - Brachman & Levesque, Chapter 3. Expressing knowledge
  - Brachman & Levesque, Chapter 8. Object-oriented representation
  - Russell & Norvig, Chapter 10. Knowledge representation
  - Lenat and Guha, *Building Large Knowledge-Based Systems: Representation and Inference in the Cyc Project* [94]
  - Davis, *Representations of Commonsense Knowledge* [22]

7 Set 7: Representing knowledge II: probabilistic, nonmonotonic, temporal, and other logics

- Brachman & Levesque, Chapter 11. Defaults

- Ferraris, Lee, and Lifschitz, A new perspective on stable models [53]

- Brachman & Levesque, Chapter 12. Vagueness, uncertainty, and degrees of belief

- Koller and Pfeffer, Object-oriented Bayesian networks [87]

- Elkan, The paradoxical success of fuzzy logic [48]

- **Supplementary**

  - Horty, Moral dilemmas and nonmonotonic logic [81]
  - Friedman, Getoor, Koller, and Pfeffer, Learning probabilistic relational models [56]
  - Wellman, Fundamental concepts of qualitative probabilistic networks [167]
  - Halpern, A counterexample to theorems of Cox and Fine [72]
  - Halpern, Cox's theorem revisited [73]
  - Elkan, The paradoxical controversy over fuzzy logic [47]
  - Elkan, Paradoxes of fuzzy logic, revisited [49]
  - Ruspini, On the semantics of fuzzy logic [133]
  - Ferraris and Lifschitz, Mathematical foundations of answer set programming [54]

- **Reference**

  - Brachman & Levesque, Chapter 14. Actions
  - Russell & Norvig, Chapter 14. Probabilistic reasoning
  - Russell & Norvig, Chapter 15. Probabilistic reasoning over time
  - Harel, Dynamic logic [75]

8   Set 8: Planning I: Individual action

- Rosenschein and Kaelbling, A Situated View of Representation and Control [132]

- Srivastava and Kambhampati, Synthesizing customized planners from specifications [150]

- Boutilier, Dean, and Hanks, Decision Theoretic Planning: Structural Assumptions and Computational Leverage [9].

- **Supplementary**

  - Rosenschein and Kaelbling, The synthesis of machines with provable epistemic properties [131]

  - Wellman, *Formulation of Tradeoffs in Planning Under Uncertainty* [166]

  - Brooks, Intelligence without representation [13]

  - Etzioni, Intelligence without robots: a reply to Brooks [51]

  - Tate, Towards a plan ontology [154]

  - Smith, Parra, and Westfold, Synthesis of Planning and Scheduling Software [149]

- **Reference**

  - Russell & Norvig, Chapter 11. Planning

  - Russell & Norvig, Chapter 12. Planning and acting in the real world

  - Bratman, *Intention, Plans, and Practical Reason* [12]

9 Set 9: Planning II: Group efforts

- Grosz and Kraus, Collaborative plans for complex group action [70]

- Shoham & Tennenholtz, On the synthesis of useful social laws for artificial agent societies [144]

- Davis & Smith, Negotiation as a metaphor for distributed problem solving [24]

- Wellman, A market-oriented programming environment and its application to distributed multicommodity flow problems [168]

- **Supplementary**

  - Grosz, Collaborative systems, [69]
  - Malone, Fikes, Grant & Howard, Enterprise: a market-like task scheduler for distributed computing environments [102]
  - Miller & Drexler, Markets and computation: agoric open systems [112]
  - Waldspurger et al., Spawn: a distributed computational economy [164]
  - Goldman and Lynch, Hierarchical correctness proofs for distributed algorithms [99]
  - Harel, Statecharts: a visual formalism for complex systems [76]

10 Set 10: Decision-model construction and qualitative decision theory

- Wellman, Breese & Goldman, From knowledge bases to decision models [169]

- Doyle and Thomason, Background to qualitative decision theory [45]

- Pennock and Wellman, Representing aggregate belief through the competitive equilibrium of a securities market [129]

- Wellman & Doyle, Preferential semantics for goals [170]

- Boutlier, Brafman, Hoos and Poole, Reasoning with conditional ceteris paribus preference statements [8]

- Haddawy & Hanks, Utility Models for Goal-Directed Decision-Theoretic Planners [71]

- **Supplementary**

  - Wellman, Eckman, Fleming, Marshall, & Sonnenberg, Automated critiquing of medical decision trees [172]
  - Howard, An assessment of decision analysis [83]
  - Doyle, Shoham & Wellman, A logic of relative desire [44]
  - Wellman & Doyle, Modular utility representation for decision-theoretic planning [171]
  - Doyle, Prospects for preferences [41]
  - McGeachie and Doyle, Utility Functions for Ceteris Paribus Preferences [110]
  - La Mura and Shoham, Expected utility networks [90]
  - Loui, Defeasible specification of utilities [96]
  - Dean & Wellman, On the Value of Goals [25]
  - von Wright, *The Logic of Preference: An Essay* [161]
  - von Wright, The logic of preference reconsidered [162]

- **Reference**

  - Russell & Norvig, Chapter 16. Making simple decisions
  - Russell & Norvig, Chapter 17. Making complex decisions

11 Set 11: Deliberation and argumentation

- Dung, Kowalski, and Toni, Dialectic proof procedures for assumption-based, admissible argumentation

- Mikhail, Universal moral grammar: theory, evidence, and the future [111]

- **Supplementary**

  - Doyle, A model for deliberation, action, and introspection [28]
  - Gordon and Karacapilidis, The Zeno argumentation framework [68]
  - Loui, Process and policy: resource-bounded nondemonstrative reasoning [97]
  - Bench-Capon and Dunne: Argumentation in artificial intelligence [6]

12   Set 12: Knowledge acquisition

- Blythe, Kim, Ramachandran, and Gil, An integrated environment for knowledge acquisition [7]

- Eriksson, Fergerson, Shahar, and Musen. Automatic generation of ontology editors [50]

- Schreiber, Wielinga, Hoog, Akkermans, and Van de Velde, CommonKADS: A Comprehensive Methodology for KBS Development [139]

13   Set 13: Learning I: Conceptual and inferential methods

- Lenat, The ubiquity of discovery [92]

- Haussler, Probably approximately correct learning [77]

- Muggleton and De Raedt, Inductive Logic Programming: Theory and Methods [121]

- Choueiry, Iwasaki, and McIlraith: Towards a practical theory of reformulation for reasoning about physical systems [18]

- **Supplementary**

  - Sussman, A computer model of skill acquisition [152]
  - Lenat, EURISKO: a program that learns new heuristics and domain concepts [93]
  - Cucker and Smale, On the mathematical foundations of learning [21]
  - Page, ILP: Just do it [127]

- **Reference**

  - Russell & Norvig, Chapter 18. Learning from observations
  - Russell & Norvig, Chapter 19. Knowledge in learning

14 Set 14: Learning II: Statistical and reinforcement methods

- Müller, Mika, Rätsch, Tsuda, and Schölkopf, An introduction to kernel-based learning algorithms [122]

- Kaelbling, Littman, and Moore, Reinforcement Learning: A Survey [84]

- Shoham, Powers, and Grenager: If multi-agent learning is the answer, what is the question? [143]

- Wolpert, The supervised learning no-free-lunch theorems [174]

- **Supplementary**

    - Burges, A tutorial on support vector machines for pattern recognition [14]
    - Branting, Learning feature weights from customer return-set selections [11]
    - Chajewska and Koller, Utilities as random variables: Density estimation and structure discovery [16]
    - Chajewska, Koller, and Parr, Making rational decisions using adaptive utility elicitation [17]
    - Schaffer, A conservation law for generalization performance [136]
    - Wolpert and Macready, No free lunch theorems for optimization [175]

- **Reference**

    - Russell & Norvig, Chapter 20. Statistical learning methods
    - Russell & Norvig, Chapter 21. Reinforcement learning

References

[1] Kenneth J. Arrow. Values and collective decision-making. In Peter Laslett and W. G. Runciman, editors, *Philosophy, Politics and Society*, Third, pages 215–232. Basil Blackwell Oxford, 1967.

[2] Jonathan Baron. *Rationality and Intelligence*. Cambridge University Press, Cambridge, 1985.

[3] Jon Barwise. An introduction to first-order logic. In J. Barwise, editor, *Handbook of Mathematical Logic*, pages 5–46. North Holland, Amsterdam, 1977.

[4] Jon Barwise. Model-theoretic logics: background and aims. In Jon Barwise and Solomon Feferman, editors, *Model-Theoretic Logics*, pages 3–23. Springer-Verlag, New York, 1985.

[5] G. S. Becker. *The Economic Approach to Human Behavior*. University of Chicago Press, Chicago, 1976.

[6] Trevor J. M. Bench-Capon and Paul E. Dunne. Argumentation in artificial intelligence. *Artif. Intell.*, 171(10-15):619–641, 2007.

[7] Jim Blythe, Jihie Kim, Surya Ramachandran, and Yolanda Gil. An integrated environment for knowledge acquisition. In *Proceedings of the 2001 International Conference on Intelligent User Interfaces (IUI-2001)*, Santa Fe, New Mexico, January 2001.

[8] Craig Boutilier, Ronen I. Brafman, Holger H. Hoos, and David Poole. Reasoning with conditional ceteris paribus preference statements. In *Proceedings of the Fifteenth Annual Conference on Uncertainty in Artificial Intelligence (UAI-99)*, pages 71–80, 1999.

[9] Craig Boutilier, Thomas Dean, and Steve Hanks. Decision theoretic planning: Structural assumptions and computational leverage. *Journal of Artificial Intelligence Research*, 11:1–94, 1999.

[10] Ronald J. Brachman and Hector J. Levesque, editors. *Readings in Knowledge Representation*. Morgan Kaufmann Publishers, Inc., Los Altos, California, 1985.

[11] L. Karl Branting. Learning feature weights from customer return-set selections. *Knowledge and Information Systems*, 2003. To appear.

[12] Michael Bratman. *Intention, Plans, and Practical Reason*. MIT Press, Cambridge, 1987.

[13] Rodney A. Brooks. Intelligence without representation. *Artificial Intelligence*, 47:139–159, 1991.

[14] C. J. C. Burges. A tutorial on support vector machines for pattern recognition. *Knowledge Discovery and Data Mining*, 2(2):121–167, 1998.

[15] Nicholas L. Cassimatis. A cognitive substrate for achieving human-level intelligence. *AI Magazine*, 27(2):45–56, Summer 2006.

[16] Urszula Chajewska and Daphne Koller. Utilities as random variables: density estimation and structure discovery. In *Proceedings of the Sixteenth Conference on Uncertainty in Artificial Intelligence (UAI-2000)*. AUAI, August 2000.

[17] Urszula Chajewska, Daphne Koller, and Ronald Parr. Making rational decisions using adaptive utility elicitation. In *Proceedings of the Seventeenth National Conference on Artificial Intelligence (AAAI-00)*, pages 363–369. AAAI, AAAI Press, August 2000.

[18] Berthe Y. Choueiry, Yumi Iwasaki, and Sheila A. McIlraith. Towards a practical theory of reformulation for reasoning about physical systems. *Artif. Intell.*, 162(1-2):145–204, 2005.

[19] Michael T. Cox. Metacognition in computation: A selected research review. *Artif. Intell.*, 169(2):104–141, 2005.

[20] Michael T. Cox. Perpetual self-aware cognitive agents. *AI Magazine*, 28(1):32–46, Spring 2007.

[21] Felipe Cucker and Steve Smale. On the mathematical foundations of learning. *Bulletin of the American Mathematical Society*, 39:1–49, 2002.

[22] Ernest Davis. *Representations of Commonsense Knowledge*. Morgan Kaufmann, 1990.

[23] Randall Davis, Howard Shrobe, and Peter Szolovits. What is a knowledge representation. *AI Magazine*, 14(1):17–33, 1993.

[24] Randall Davis and Reid G. Smith. Negotiation as a metaphor for distributed problem solving. *Artificial Intelligence*, 20:63–109, 1983.

[25] Thomas Dean and Michael P. Wellman. On the value of goals. In *Proceedings from the Rochester Planning Workshop*, April 1989. Published as University of Rochester TR 284.

[26] G. Debreu. *Theory of Value: an axiomatic analysis of economic equilibrium*. Wiley, New York, 1959.

[27] Jon Doyle. A truth maintenance system. *Artificial Intelligence*, 12(2):231–272, 1979.

[28] Jon Doyle. A model for deliberation, action, and introspection. AI-TR 581, Massachusetts Institute of Technology, Artificial Intelligence Laboratory, 1980.

[29] Jon Doyle. What is Church's Thesis? An outline. Written at Carnegie Mellon University and privately circulated thereafter, eventually published as [40], September 1982.

[30] Jon Doyle. Some theories of reasoned assumptions: An essay in rational psychology. Technical Report 83-125, Department of Computer Science, Carnegie Mellon University, Pittsburgh, PA, 1983.

[31] Jon Doyle. How to frame it: Modern applied logic from the top down, or secrets of the metamathematicians. Technical report, Carnegie Mellon University Computer Science Department, October 1986.

[32] Jon Doyle. Artificial intelligence and rational self-government. Technical Report CS-88-124, Carnegie-Mellon University Computer Science Department, 1988.

[33] Jon Doyle. Big problems for artificial intelligence. *AI Magazine*, 9(1):19–22, 1988.

[34] Jon Doyle. Reasoning, representation, and rational self-government. In Zbigniew W. Ras, editor, *Methodologies for Intelligent Systems, 4*, pages 367–380, New York, 1989. North-Holland.

[35] Jon Doyle. Rational belief revision (preliminary report). In Richard E. Fikes and Erik Sandewall, editors, *Proceedings of the Second Conference on Principles of Knowledge Representation and Reasoning*, pages 163–174, San Mateo, CA, 1991. Morgan Kaufmann.

[36] Jon Doyle. Rationality and its roles in reasoning. *Computational Intelligence*, 8(2):376–409, 1992.

[37] Jon Doyle. Reason maintenance and belief revision: Foundations vs. coherence theories. In Peter Gärdenfors, editor, *Belief Revision*, pages 29–51. Cambridge University Press, Cambridge, 1992.

[38] Jon Doyle. Reasoned assumptions and rational psychology. *Fundamenta Informaticae*, 20(1-3):35–73, 1994.

[39] Jon Doyle. A reasoning economy for planning and replanning. In Mark H. Burstein, editor, *Proceedings of the 1994 ARPA/Rome Laboratory Knowledge-Based Planning and Scheduling Initiative Workshop*, pages 35–43, San Francisco, CA, 1994. Morgan Kaufmann.

[40] Jon Doyle. What is Church's thesis? an outline. *Minds and Machines*, 12(4):519–520, November 2002.

[41] Jon Doyle. Prospects for preferences. *Computational Intelligence*, 20(2):111–136, May 2004.

[42] Jon Doyle and Thomas L. Dean. Strategic directions in artificial intelligence. *ACM Computing Surveys*, 28(4):653–670, December 1996.

[43] Jon Doyle and Ramesh S. Patil. Two theses of knowledge representation: Language restrictions, taxonomic classification, and the utility of representation services. *Artificial Intelligence*, 48(3):261–297, April 1991.

[44] Jon Doyle, Yoav Shoham, and Michael P. Wellman. A logic of relative desire (preliminary report). In Zbigniew Ras, editor, *Proceedings of the Sixth International Symposium on Methodologies for Intelligent Systems*, Lecture Notes in Computer Science, Berlin, 1991. Springer-Verlag.

[45] Jon Doyle and Richmond H. Thomason. Background to qualitative decision theory. *AI Magazine*, 20(2):55–68, Summer 1999.

[46] Jon Doyle and Michael P. Wellman. Impediments to universal preference-based default theories. *Artificial Intelligence*, 49(1-3):97–128, May 1991.

[47] C. Elkan. The paradoxical controversy over fuzzy logic. *IEEE Expert*, pages 47–49, August 1994.

[48] C. Elkan. The paradoxical success of fuzzy logic. *IEEE Expert*, pages 3–8, August 1994. With fifteen responses on pp. 9-46.

[49] C. Elkan. Paradoxes of fuzzy logic, revisited. *International Journal of Approximate Reasoning*, 26(2):153–155, 2001.

[50] H. Eriksson, R. W. Fergerson, Y. Shahar, and M. A. Musen. Automatic generation of ontology editors. In *Twelfth Banff Knowledge Acquisition for Knowledge-based systems Workshop*, Banff, Alberta, Canada, 1999.

[51] Oren Etzioni. Intelligence without robots: A reply to brooks. *AI Magazine*, 14(4):7–13, 1993.

[52] Jerome A. Feldman and Robert F. Sproull. Decision Theory and Artificial Intelligence II: The hungry monkey. *Cognitive Science*, 1:158–192, 1977.

[53] Paolo Ferraris, Joohyung Lee, and Vladimir Lifschitz. A new perspective on stable models. In Manuela M. Veloso, editor, *IJCAI*, pages 372–379, 2007.

[54] Paolo Ferraris and Vladimir Lifschitz. Mathematical foundations of answer set programming. In Sergei N. Artëmov, Howard Barringer, Artur S. d'Avila Garcez, Luís C. Lamb, and John Woods, editors, *We Will Show Them! (1)*, pages 615–664. College Publications, 2005.

[55] Harry Frankfurt. Freedom of the will and the concept of a person. *Journal of Philosophy*, 68:5–20, 1971.

[56] Nir Friedman, Lise Getoor, Daphne Koller, and Avi Pfeffer. Learning probabilistic relational models. In Thomas Dean, editor, *IJCAI*, pages 1300–1309. Morgan Kaufmann, 1999.

[57] Peter Gärdenfors and Nils-Eric Sahlin, editors. *Decision, Probability, and Utility: Selected Readings*. Cambridge University Press, Cambridge, 1988.

[58] Peter Gärdenfors and Nils-Eric Sahlin. Introduction: Bayesian decision theory—foundations and problems. In Peter Gärdenfors and Nils-Eric Sahlin, editors, *Decision, Probability, and Utility: Selected Readings*, pages 1–15. Cambridge University Press, Cambridge, 1988.

[59] M. R. Genesereth and R. E. Fikes. Knowledge interchange format, version 3.0 reference manual. Technical Report Logic-92-1, Computer Science Department, Stanford University, March 1992.

[60] Matthew L. Ginsberg. The computational value of nonmonotonic reasoning. In James Allen, Richard Fikes, and Erik Sandewall, editors, *Proceedings of the Second International Conference on Principles of Knowledge Representation and Reasoning*, pages 262–268, San Mateo, CA, April 1991. Morgan Kaufmann.

[61] Matthew L. Ginsberg. Knowledge interchange format: the KIF of death. *AI Magazine*, 12(3):57–63, 1991.

[62] Matthew L. Ginsberg and Donald F. Geddis. Is there any need for domain-dependent control information? In Thomas Dean and Kathleen McKeown, editors, *Proceedings of the Ninth National Conference on Artificial Intelligence*, pages 452–457, Cambridge, MA, July 1991. MIT Press.

[63] Ben Goertzel. Artificial general intelligence: Now is the time. Published on KurzweilAI.net April 9, 2007, 2007.

[64] Ben Goertzel. Goertzel, human-level artificial general intelligence and the possibility of a technological singularity: A reaction to Ray Kurzweil's *The Singularity is Near*, and McDermott's critique of Kurzweil. *Artificial Intelligence*, 171:1161–1173, 2007.

[65] Carla P. Gomes and Bart Selman. Can get satisfaction. *Nature*, 435:751–752, June 2005.

[66] I. J. Good. A five-year plan for automatic chess. In E. Dale and D. Michie, editors, *Machine Intelligence 2*, volume 2, pages 89–118. Oliver and Boyd, London, 1968.

[67] I. J. Good. Twenty-seven principles of rationality. In V. P. Godambe and D. A. Sprott, editors, *Foundations of Statistical Inference*, pages 124–127. Holt, Rinehart and Winston, Toronto, 1971. Appendix to "The probabilistic explication of information, evidence, surprise, causality, explanation, and utility".

[68] Thomas F. Gordon and Nikos Karacapilidis. The zeno argumentation framework. In *Proceedings of the Sixth International Conference on Artificial Intelligence and Law*, pages 10–18, Melbourne, Austrailia, 1997.

[69] B. Grosz. Collaborative systems. *AI Magazine*, 17(2):67–85, 1996.

[70] Barbara J. Grosz and Sarit Kraus. Collaborative plans for complex group action. *Artificial Intelligence*, 86(2):269–357, 1996.

[71] Peter Haddawy and Steve Hanks. Utility models for goal-directed decision-theoretic planners. *Computational Intelligence*, 14(3):392–429, 1998.

[72] J. Y. Halpern. A counterexample to theorems of cox and fine. *Journal of Artificial Intelligence Research*, 10:67–85, 1999.

[73] J. Y. Halpern. Cox's theorem revisited. *Journal of Artificial Intelligence Research*, 11:429–435, 1999.

[74] Joseph Y. Halpern. Reasoning about knowledge: a survey. In D. Gabbay, C. J. Hogger, and J. A. Robinson, editors, *Handbook of Logic in Artificial Intelligence and Logic Programming*, volume 4, pages 1–34. Oxford University Press, 1995. http://www.cs.cornell.edu/home/halpern/papers/knowledge_survey.pdf.

[75] David Harel. Dynamic logic. In D. Gabbay and F. Guenther, editors, *Handbook of Philosophical Logic*, volume II, pages 497–604. Reidel, Dordrecht, 1984.

[76] David Harel. Statecharts: A visual formalism for complex systems. *Science of Computer Programming*, 8(3):231–274, June 1987.

[77] David Haussler. Probably approximately correct learning. In *National Conference on Artificial Intelligence*, pages 1101–1108, 1990.

[78] Patrick J. Hayes. The frame problem and related problems in artificial intelligence. In A. Elithorn and D. Jones, editors, *Artificial and Human Thinking*, pages 45–59. Jossey-Bass, San Francisco, 1973. Reprinted in [165].

[79] Patrick J. Hayes. Some problems and non-problems in representation theory. In *Proceedings of the AISB Summer Conference*, pages 63–79. Univ. of Sussex, 1974. Reprinted in [10].

[80] Patrick J. Hayes. In defence of logic. In *Proceedings of the Fifth International Joint Conference on Artificial Intelligence*, pages 559–565, 1977.

[81] John F. Horty. Moral dilemmas and nonmonotonic logic. *Journal of Philosophical Logic*, 1992. To appear.

[82] Eric J. Horvitz, Gregory F. Cooper, and David E. Heckerman. Reflection and action under scarce resources: Theoretical principles and empirical study. In N. S. Sridharan, editor, *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence*, volume 2, pages 1121–1127, San Mateo, CA, August 1989. International Joint Conferences on Artificial Intelligence, Inc., Morgan Kaufmann.

[83] Ron A. Howard. An assessment of decision analysis. *Operations Research*, 28(1):4–27, 1980.

[84] Leslie Pack Kaelbling, Michael L. Littman, and Andrew W. Moore. Reinforcement learning: A survey. *Journal of Artificial Intelligence Research*, 4, 1996.

[85] Donald Kalish and Richard Montague. *Logic: Techniques of Formal Reasoning*. Harcourt Brace Jovanovich, New York, 1964.

[86] Henry A. Kautz and Bart Selman. The state of sat. *Discrete Applied Mathematics*, 155(12):1514–1524, 2007.

[87] Daphne Koller and Avi Pfeffer. Object-oriented bayesian networks. In *Proceedings of the 13th Annual Conference on Uncertainty in AI (UAI)*, pages 302–313), 1997.

[88] Ray Kurzweil. The law of accelerating returns. Published on KurzweilAI.net, March 7 2001. http://www.kurzweilai.net/articles/art0134.html?printable=1.

[89] F. E. Kydland and E. C. Prescott. Rules rather than discretion: the inconsistency of optimal plans. *J. Political Economy*, 85(3):473–491, 1977.

[90] Piero La Mura and Yoav Shoham. Expected utility networks. In *Proceedings of the Fifteenth Annual Conference on Uncertainty in Artificial Intelligence (UAI-99)*, pages 366–373, 1999.

[91] D. B. Lenat. CYC: a large-scale investment in knowledge infrastructure. *Communications of the ACM*, 38(11):33–38, November 1995.

[92] Douglas B. Lenat. Computers and thought lecture: The ubiquity of discovery. In *IJCAI*, page 1093, 1977.

[93] Douglas B. Lenat. Eurisko: A program that learns new heuristics and domain concepts. *Artif. Intell.*, 21(1-2):61–98, 1983.

[94] Douglas B. Lenat and R. V. Guha. *Building Large Knowledge-Based Systems: Representation and Inference in the Cyc Project*. Addison-Wesley, Reading, MA, 1990.

[95] Barton Lipman. How to decide how to decide how to . . .: Limited rationality in decisions and games. In *Working Notes of the AAAI Symposium on AI and Limited Rationality*, pages 77–80. AAAI, March 1989.

[96] Ronald Loui. Defeasible specification of utilities. In Henry E. Kyburg, Jr., Ronald P. Loui, and Greg N. Carlson, editors, *Knowledge Representation and Defeasible Reasoning*, pages 345–359. Kluwer Academic Publishers, 1990.

[97] Ronald Prescott Loui. Process and policy: Resource-bounded nondemonstrative reasoning. *Computational Intelligence*, 14(1):1–38, January 1998.

[98] R. D. Luce and H. Raiffa. *Games and Decisions*. Wiley, New York, 1957.

[99] N. Lynch and M. Tuttle. Hierarchical correctness proofs for distributed algorithms. In *Proceedings of the Sixth ACM Symposium on Principles of Distributed Computing*, pages 137–151, 1987.

[100] Mark J. Machina. Choice under uncertainty: Problems solved and unsolved. *Journal of Economic Perspectives*, 1(1):121–154, Summer 1987.

[101] Mark J. Machina. Dynamic consistency and non-expected utility models of choice under uncertainty. *Journal of Economic Literature*, XXVII(4):1622–1668, December 1989.

[102] T. W. Malone, R. E. Fikes, K. R. Grant, and M. T. Howard. Enterprise: A market-like task scheduler for distributed computing environments. In B. A. Huberman, editor, *The Ecology of Computation*, pages 177–205. North-Holland, Amsterdam, 1988.

[103] David A. McAllester and Robert Givan. Natural language syntax and first order inference. *Artificial Intelligence*, 1992.

[104] J. McCarthy and P. J. Hayes. Some philosophical problems from the standpoint of artificial intelligence. In B. Meltzer and D. Michie, editors, *Machine Intelligence 4*, pages 463–502. Edinburgh University Press, 1969.

[105] John McCarthy. Programs with common sense. In Ronald J. Brachman and Hector J. Levesque, editors, *Readings in Knowledge Representation*, pages 300–307. Morgan Kaufmann, Los Altos, 1985. Originally published 1958.

[106] John McCarthy. From here to human-level AI. *Artificial Intelligence*, 171:1174–1182, 2007.

[107] Drew McDermott. A critique of pure reason. *Computational Intelligence*, 3:151–160, 1987.

[108] Drew McDermott. Level-headed. *Artificial Intelligence*, 171:1183–1186, 2007.

[109] Drew V. McDermott. Kurzweil's argument for the success of ai. *Artif. Intell.*, 170(18):1227–1233, 2006.

[110] Michael McGeachie and Jon Doyle. Utility functions for ceteris paribus preferences. *Computational Intelligence*, 20(2):158–217, May 2004.

[111] John Mikhail. Universal moral grammar: Theory, evidence, and the future. *Trends in Cognitive Sciences*, 11(4):143–152, April 2007.

[112] M. S. Miller and K. E. Drexler. Markets and computation: Agoric open systems. In B. A. Huberman, editor, *The Ecology of Computation*, pages 133–176. North-Holland, Amsterdam, 1988.

[113] M. Minsky. Problems of formulation for artificial intelligence. In *Proceedings of a Symposium on Mathematical Problems in Biology*, pages 35–46. American Mathematical Society, Providence, 1962.

[114] Marvin Minsky. Steps toward artificial intelligence. In Edward A. Feigenbaum and Jerome Feldman, editors, *Computers and Thought*, pages 406–450. McGraw-Hill, 1963.

[115] Marvin Minsky. Matter, mind, and models. In *Information Processing: Proceedings of the IFIP Congress*, pages 45–49, Amsterdam, 1965. North-Holland.

[116] Marvin Minsky. The society theory of thinking. In P. H. Winston and R. Brown, editors, *Artificial Intelligence: An MIT perspective*, pages 421–452. MIT Press, Cambridge, MA, 1979.

[117] Marvin Minsky. K-lines: a theory of memory. *Cognitive Science*, 4:117–133, 1980.

[118] Marvin Minsky. A framework for representing knowledge. In John Haugeland, editor, *Mind Design*, pages 95–128. MIT Press, Cambridge, MA, 1981.

[119] Marvin Minsky. *The Society of Mind*. Simon and Schuster, New York, 1986.

[120] Marvin Minsky. *The Emotion Machine*. Simon and Schuster, November 2006.

[121] Stephen Muggleton and Luc De Raedt. Inductive logic programming: Theory and methods. *Journal of Logic Programming*, 19/20:629–679, 1994.

[122] K.-R. Müller, S. Mika, G. Rätsch, K. Tsuda, and B. Schölkopf. An introduction to kernel-based learning algorithms. *IEEE Neural Networks*, 12(2):181–201, May 2001.

[123] Karen Myers, Pauline Berry, Jim Blythe, Ken Conley, Melinda Gervasio, Deborah McGuinness, David Morley, Avi Pfeffer, Martha Pollack, and Milind Tambe. An intelligent personal assistant for task and time management. *AI Magazine*, 28(2):47–61, Summer 2007.

[124] R. Neches, R. E. Fikes, T. Finin, T. R. Gruber, R. Patil, T. Senator, and W. R. Swartout. Enabling technology for knowledge sharing. *AI Magazine*, 12(3):16–36, 1991.

[125] Allen Newell. Précis of Unified Theories of Cognition. *Behavioral and Brain Sciences*, 15(3):425–492, 1992.

[126] Nils J. Nilsson. Human-level artificial intelligence? Be serious! *AI Magazine*, 26(4):68–75, Winter 2005.

[127] David Page. Ilp: Just do it. In J. Cussens and A. Frisch, editors, *Proceedings of ILP'2000*, number 1866 in Lecture Notes in Artificial Intelligence, pages 3–18, Berlin, 2000. Springer.

[128] Derek C. Penn, Keith J. Holyoak, and Daniel J. Povinelli. Darwin's mistake: Explaining the discontinuity between human and nonhuman minds. *Behavioral and Brain Sciences*, 31(02):109–130, 2008.

[129] David M. Pennock and Michael P. Wellman. Representing aggregate belief through the competitive equilibrium of a securities market. In Dan Geiger and Prakash P. Shenoy, editors, *UAI*, pages 392–400. Morgan Kaufmann, 1997.

[130] Fred S. Roberts. *Discrete Mathematical Models*. Prentice-Hall Inc., Englewood Cliffs, New Jersey, 1976.

[131] S. J. Rosenschein and L. P. Kaelbling. The synthesis of machines with provable epistemic properties. In J. Halpern, editor, *Proceedings of the Conference on Theoretical Aspects of Reasoning About Knowledge*, pages 83–98, Los Altos, CA, 1986. Morgan Kaufmann.

[132] Stanley J. Rosenschein and Leslie Pack Kaelbling. A situated view of representation and control. *Artificial Intelligence*, 73, 1995.

[133] Enrique H. Ruspini. On the semantics of fuzzy logic. *International Journal of Approximate Reasoning*, 5(1):45–88, 1991.

[134] Stuart Russell and Eric Wefald. Principles of metareasoning. *Artificial Intelligence*, 49(1-3):361–395, May 1991.

[135] Leonard J. Savage. *The Foundations of Statistics*. Dover Publications, New York, second edition, 1972.

[136] Cullen Schaffer. A conservation law for generalization performance. In *ICML*, pages 259–265, 1994.

[137] T. C. Schelling. The intimate contest for self-command. In *Choice and Consequence: Perspectives of an errant economist*, pages 57–82. Harvard University Press, Cambridge, 1984.

[138] T. C. Schelling. The mind as a consuming organ. In *Choice and Consequence: Perspectives of an errant economist*, pages 328–346. Harvard University Press, Cambridge, 1984.

[139] Guus Schreiber, Bob Wielinga, Robert de Hoog, Hans Akkermans, and Walter Van de Velde. Commonkads: A comprehensive methodology for kbs development. *IEEE Expert: Intelligent Systems and Their Applications*, 9(6):28–37, December 1994.

[140] Dana S. Scott. Domains for denotational semantics. In M. Nielsen and E. M. Schmidt, editors, *Automata, Languages, and Programming: Ninth Colloquium*, volume 140 of *Lecture Notes in Computer Science*, pages 577–613, Berlin, 1982. Springer-Verlag.

[141] John R. Searle. How to study consciousness scientifically. *Philosophical Transactions: Biological Sciences*, 1998.

[142] Amartya Sen. The impossibility of a Paretian liberal. *Journal of Political Economy*, 78:152–157, 1970.

[143] Yoav Shoham, Rob Powers, and Trond Grenager. If multi-agent learning is the answer, what is the question? *Artif. Intell.*, 171(7):365–377, 2007.

[144] Yoav Shoham and Moshe Tennenholtz. On the synthesis of useful social laws for artificial agent societies. In *Proceedings of the Tenth National Conference on Artificial Intelligence*, pages 276–281, Cambridge, MA, 1992. MIT Press.

[145] Howard Shrobe and Jon Doyle. Active trust management for autonomous adaptive survivable systems. In P. Robertson, H. Shrobe, and R. Laddaga, editors, *Self-Adaptive Software*, pages 40–49. Springer Verlag, Berlin, Revised papers from the First International Workshop on Self-Adaptive Software (IWSAS 2000) 2001.

[146] Herbert A. Simon. A behavioral model of rational choice. *Quarterly Journal of Economics*, 69:99–118, 1955.

[147] Herbert A. Simon. Rational choice and the structure of the environment. *Psychological Review*, 63:129–138, 1956.

[148] Herbert A. Simon and Joseph B. Kadane. Optimal problem-solving search: All-or-none solutions. *Artificial Intelligence*, 6:235–247, 1975.

[149] D. R. Smith, E. A. Parra, and S. J. Westfold. Synthesis of planning and scheduling software. In Austin Tate, editor, *Advanced Planning Technology*. AAAI Press, 1996.

[150] B. Srivastava and S. Kambhampati. Synthesizing customized planners from specifications. *Journal of Artificial Intelligence Research*, 8:93–128, 1998.

[151] George J. Stigler and Gary S. Becker. De gustibus non est disputandum. *American Economic Review*, 67:76–90, 1977.

[152] Gerald Jay Sussman. *A Computer Model of Skill Acquisition*. American Elsevier, 1975.

[153] William Swartout, Jonathan Gratch, Randall W. Hill, Eduard Hovy, Stacy Marsella, Jeff Rickel, and David Traum. Toward virtual humans. *AI Magazine*, 27(2):96–108, Summer 2006.

[154] A. Tate. Towards a plan ontology. *AI-IA Notizie (Quarterly Publication of the Associazione Italiana per l'Intelligenza Artificiale)*, 9(1):19–26, March 1996. Special Issue on Aspects of Planning Research.

[155] Richard H. Thaler and H. M. Shefrin. An economic theory of self-control. *Journal of Political Economy*, 89(2):392–406, 1981.

[156] R. H. Thomason. The context-sensitivity of belief and desire. In Michael P. Georgeff and Amy L. Lansky, editors, *Reasoning about Actions and Plans: Proceedings of the 1986 Workshop*, pages 341–360. Morgan Kaufmann, 1986.

[157] P. M. Todd and G. Gigerenzer. Bounding rationality to the world. *Journal of Economic Psychology*, 24(2):143–165, 2003.

[158] Gordon Tullock. The general irrelevance of the general impossibility theorem. *Quarterly Journal of Economics*, 81:256–270, May 1967.

[159] Bas C. Van Fraassen. Values and the heart's command. *Journal of Philosophy*, LXX(1):5–19, January 11 1973.

[160] John von Neumann and Oskar Morgenstern. *Theory of Games and Economic Behavior*. Princeton University Press, Princeton, third edition, 1953.

[161] Georg Henrik von Wright. *The Logic of Preference: An Essay*. Edinburgh University Press, Edinburgh, 1963.

[162] Georg Henrik von Wright. The logic of preference reconsidered. *Theory and Decision*, 3:140–167, 1972. Reprinted in [163].

[163] Georg Henrik von Wright. *Philosophical Logic: Philosophical Papers, Volume II*. Cornell University Press, Ithaca, NY, 1984.

[164] Carl A. Waldspurger, Tad Hogg, Bernardo A. Huberman, Jeffrey O. Kephart, and Scott Stornetta. Spawn: a distributed computational economy. *IEEE Transactions on Software Engineering*, 1992.

[165] Bonnie L. Webber and Nils J. Nilsson, editors. *Readings in Artificial Intelligence*. Morgan-Kaufmann, Los Altos, 1981.

[166] Michael P. Wellman. *Formulation of Tradeoffs in Planning Under Uncertainty*. Pitman and Morgan Kaufmann, 1990.

[167] Michael P. Wellman. Fundamental concepts of qualitative probabilistic networks. *Artificial Intelligence*, 44(3):257–304, 1990.

[168] Michael P. Wellman. A market-oriented programming environment and its application to distributed multicommodity flow problems. *J. Artif. Intell. Res. (JAIR)*, 1:1–23, 1993.

[169] Michael P. Wellman, John S. Breese, and Robert P. Goldman. From knowledge bases to decision models. *The Knowledge Engineerting Review*, 7(1):35–53, 1992.

[170] Michael P. Wellman and Jon Doyle. Preferential semantics for goals. In *Proceedings of the National Conference on Artificial Intelligence*, pages 698–703, 1991.

[171] Michael P. Wellman and Jon Doyle. Modular utility representation for decision-theoretic planning. In *Proceedings of the First International Conference on AI Planning Systems*, pages 236–242, June 1992.

[172] Michael P. Wellman, Mark H. Eckman, Craig Fleming, Sharon L. Marshall, Frank A. Sonnenberg, and Stephen G. Pauker. Automated critiquing of medical decision trees. *Medical Decision Making*, 9(4):272–284, 1989.

[173] Ryan Williams, Carla P. Gomes, and Bart Selman. Backdoors to typical case complexity. In Georg Gottlob and Toby Walsh, editors, *IJCAI*, pages 1173–1178. Morgan Kaufmann, 2003.

[174] David H. Wolpert. The supervised learning no-free-lunch theorems. In *Proceedings of the 6th Online World Conference on Soft Computing in Industrial Applications*, 2001.

[175] David H. Wolpert and William G. Macready. No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation*, 1(1):67–82, April 1997.