SYLLABUS FOR 2000 FALL

John McCarthy
Computer Science Department
Stanford University
Stanford, CA 94305
jmc@cs.stanford.edu
http://www-formal.stanford.edu/jmc/

2000 Sep 27, 10:54 a.m.

Abstract

CS323 covers primarily logical AI with emphasis on the epistemology (often called ontology these days) and nonmonotonic reasoning.

I'm not sure how rigorously I will follow this syllabus. There will definitely be some additional readings, and some topics may take longer to cover than planned.

The first lectures will follow the preface, introduction and first chapter of Shanahan.

The 19 lectures will cover topics approximately as follows:

1. Approaches to AI. Biological approach that imitates the human, computer science that looks at the problems the world presents. The logic approach that emphasizes facts more than programs. Epistemology and heuristics. Some history. Nonmonotonic reasoning, context and other extensions to mathematical logic. Importance of the common sense informatic situation. How far to human-level AI? Reading: preface and introduction of Shanahan book, [McC59], [McC99], Optional: [McC96a].

2. Logic for AI (Lifschitz's simple blocks world example). Logical languages, interpretations, models. Intended and unintended models. **Preread:** [Lif87],[Lif96a],[Lif96b].

3. Situation Calculus: [MH69] is the original reference, but this year we'll follow Shanahan.. **Preread:** Chapter 1 and 2 of Shanahan up to p.47.

4. Continuation of blocks world formalization. Frames as objects. (not in Shanahan) Exercises on situation calculus.

5. Logical Foundations including circumscription. The rest of chapter 2 and 3 and 4 of Shanahan.

6. Philosophical issues. Semantics of "can". The practice of AI requires taking sides in certain longstanding philosophical controversies. For example, a person designing a computer program to learn about the world has to regard the world as more than a construct in the "mind" of the computer program. Otherwise, how could he compare the beliefs the program will come to have with facts about the world. Lots more philosophy is involved, but maybe AI people will do all right without thinking much about the conundrums philosophers have created for themselves. **Preread:** (1) section on "can" from [MH69], (2) [McC79a], [McC99].

7. Contexts as formal objects. It is a truism that the meanings of sentences and terms depend on context. The meaning of a context is itself contextual, and you can carry this as far as seems useful. The innovation here is a formal theory of the relations of different contexts and how meanings depend on context. **Preread:** [McC93]

8. Elaboration tolerance. **Preread:** McCarthy-Amir drafts : [McC96c], This is a new topic, but it relates to the continual computer science desire for modularity. We are studying how to make logical formalisms that allow changes in axiomatizations of phenomena without having to start all over. Our *Drosophila* is the missionary and cannibals problem. The point is to see what variants can be made purely by adding sentences to the original statement.

9. Approximate concepts and approximate theories. Many important concepts, perhaps most, are intrinsically approximate in that they cannot be given if-and-only-if definitions. We study how statements involving approximate concepts can have definite truth values. The semantics of approximate concepts may be different from what is standard in mathematical logic. [McC00].

10. First order theories of individual concepts and propositions. We treat individual concepts and propositions as first class objects. This lets us say more about them than can be said using them just as terms and formulas of first order logic. **Preread:** [McC79b]

11. Heuristics: We would like to have a declarative theory of heuristics so programs can reason about them. Most likely we won't have more than illustrative examples of heuristics and hits at making them into objects. Reading to come, maybe.

12. Formalization of knowledge **Preread:** [Hal95]

13. Introspection for Robots. Like people, robots will need to think about their own mental states, e.g. about their own intentions. **Preread:** [McC96b]

References

[Hal95] Joseph Y. Halpern. Reasoning about Knowledge[1]. In D. Gabbay, editor, *Handbook of Logic in Artificial Intelligence and Logic Programming, Vol.4*. Oxford University Press, 1995.

[Lif87] Vladimir Lifschitz. Formalizing Common Sense Knowledge: an Introduction[2]. Unpublished manuscript, the ps file has only pages 1-7 (among 12), 1987.

[Lif96a] Vladimir Lifschitz. First and Second-Order Logic[3]. Unpublished manuscript, 1996.

[Lif96b] Vladimir Lifschitz. Formalizing Common Sense Knowledge and Reasoning: an Example[4]. Unpublished manuscript, 1996.

[McC59] John McCarthy. Programs with Common Sense[5]. In *Mechanisation of Thought Processes, Proceedings of the Symposium of the*

[1] http://www.cs.cornell.edu/home/halpern/papers/knowledge_survey.ps
[2] http://www-formal.stanford.edu/pub2/papers/form.ps
[3] http://www-formal.stanford.edu/pub2/papers/vl-class-paper1.ps
[4] http://www-formal.stanford.edu/pub2/papers/vl-class-paper2.ps
[5] http://www-formal.stanford.edu/jmc/mcc59.html

National Physics Laboratory, pages 77–84, London, U.K., 1959. Her Majesty's Stationery Office. Reprinted in [?].

[McC79a] John McCarthy. Ascribing mental qualities to machines[6]. In Martin Ringle, editor, *Philosophical Perspectives in Artificial Intelligence*. Harvester Press, 1979. Reprinted in [?].

[McC79b] John McCarthy. First Order Theories of Individual Concepts and Propositions[7]. In Donald Michie, editor, *Machine Intelligence*, volume 9. Edinburgh University Press, Edinburgh, 1979. Reprinted in [?].

[McC93] John McCarthy. Notes on Formalizing Context[8]. In *IJCAI-93*, 1993.

[McC96a] John McCarthy. From Here to Human-Level AI[9]. In *KR96 Proceedings*, 1996. Available as http://www-formal.stanford.edu/jmc/human.html.

[McC96b] John McCarthy. Making Robots Conscious of their Mental States[10]. In Stephen Muggleton, editor, *Machine Intelligence 15*. Oxford University Press, 1996. to appear in 2000. The web version is improved from that presented at Machine Intelligence 15 in 1995.

[McC96c] John McCarthy. Missionaries and Cannibals: Making it Elaboration Tolerant[11]. In *McCarthy's web page*, 1996. Available as http://www-formal.stanford.edu/jmc/cs323/missionaries.ps.

[McC99] John McCarthy. Philosophical and Scientific Presuppositions of Logical AI[12]. In H. J. Levesque and F. Pirri, editors, *Logical Foundations for Cognitive Agents: Contributions in Honor of Ray Reiter*, pages 72–78. Springer-Verlag, 1999.

[6]http://www-formal.stanford.edu/jmc/ascribing.html

[7]http://www-formal.stanford.edu/jmc/concepts.html

[8]http://www-formal.stanford.edu/jmc/context.html

[9]http://www-formal.stanford.edu/jmc/human.html

[10]http://www-formal.stanford.edu/jmc/consciousness.html

[11]http://www-formal.stanford.edu/jmc/cs323/missionaries.ps

[12]http://www-formal.stanford.edu/jmc/phil2.html

[McC00] John McCarthy. Approximate objects and approximate theories[13]. In Anthony G. Cohn, Fausto Giunchiglia, and Bart Selman, editors, *KR2000: Principles of Knowledge Representation and Reasoning,Proceedings of the Seventh International conference*, pages 519–526. Morgan-Kaufman, 2000.

[MH69] John McCarthy and Patrick J. Hayes. Some Philosophical Problems from the Standpoint of Artificial Intelligence[14]. In B. Meltzer and D. Michie, editors, *Machine Intelligence 4*, pages 463–502. Edinburgh University Press, 1969.

/@steam.stanford.edu:/u/ftp/jmc/cs323/cs323-00.tex: begun Thu Dec 12 16:32:43 1996, latexed September 27, 2000 at 10:54 a.m.

[13] http://www.formal.stanford.edu/jmc/approximate.html
[14] http://www-formal.stanford.edu/jmc/mcchay69.html