SYLLABUS FOR 1997 WINTER

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

1997 Feb 13, 12:17 p.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 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: [McC58], 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. **Preread:** (1) The section *Formalism* (on situation calculus (that name came later)) from [MH69] (2) [McC58]

4. Frame problem and monotonic frame axioms. We will emphasize the blocks world. Relevant formulas are in [**?**] . **Preread:** the sections *Introduction* and *The frame problem(s)* [SS94]

5. Continuation of blocks world formalization. Frames as objects. Exercises on situation calculus.

6. Philosophical issues. Semantics of "can". The practice of AI requires taking sides in certain longstanding philsophical 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]

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], [MA96], [Ami97] 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 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. Reading to come.(***)

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. Circumscription (1). Circumscription is a form of non-monotonic reasoning based on minimizing predicates and formulas. **Preread:** [McC80],[McC86]

14. Circumscription (2) **Preread:** [Lif93]

15. Default Logic is the second major surviving form of non-monotonic reasoning. **Preread:** [Rei80]

16. Nonmonotonic theories of action **Preread:** sections 3,4,6 in [SS94]

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

References

[Ami97] Eyal Amir. Machinery for elaborating action[1]. Unpublished manuscript, 1997.

[Hal95] Joseph Y. Halpern. Reasoning about Knowledge[2]. In D. Gabbai, editor, *Handbook of Logic in Artificial Intelligence and Logic Programming, Vol.4*. Oxford University Press, 1995.

[1] http://www-formal.stanford.edu/eyal/nmr/machinery-ea.ps
[2] http://www.cs.cornell.edu/home/halpern/papers/knowledge_survey.ps

[Lif87] Vladimir Lifschitz. Formalizing Common Sense Knowledge: an Introduction[3]. Unpublished manuscript, the ps file has only pages 1-7 (among 12), 1987.

[Lif93] Vladimir Lifschitz. Circumscription[4]. In *Handbook of Logic in Artificial Intelligence and Logic Programming, Volume 3: Nonmonotonic Reasoning and Uncertain Reasoning*. Oxford University Press, 1993.

[Lif96a] Vladimir Lifschitz. First and Second-Order Logic[5]. Unpublished manuscript, 1996.

[Lif96b] Vladimir Lifschitz. Formalizing Common Sense Knowledge and Reasoning: an Example[6]. Unpublished manuscript, 1996.

[MA96] John McCarthy and Eyal Amir. Missionaries and Cannibals: Making it Elaboration Tolerant[7]. *Amir Web page http://www-formal.stanford.edu/eyal/*, 1996. Available as http://www-formal.stanford.edu/eyal/nmr/m4.ps.

[McC58] John McCarthy. Programs with Common Sense[8]. In *Mechanisation of Thought Processes, Proceedings of the Symposium of the National Physics Laboratory*, pages 77–84, London, U.K., 1958. Her Majesty's Stationery Office. Reprinted in McC90.

[McC79a] John McCarthy. Ascribing mental qualities to machines[9]. In Martin Ringle, editor, *Philosophical Perspectives in Artificial Intelligence*. Harvester Press, 1979. Reprinted in [McC90].

[McC79b] John McCarthy. First Order Theories of Individual Concepts and Propositions[10]. In Donald Michie, editor, *Machine Intelligence*, volume 9. Edinburgh University Press, Edinburgh, 1979. Reprinted in [McC90].

[3]http://www-formal.stanford.edu/pub2/papers/form.ps
[4]http://www.cs.utexas.edu/users/vl/circumscription.ps
[5]http://www-formal.stanford.edu/pub2/papers/vl-class-paper1.ps
[6]http://www-formal.stanford.edu/pub2/papers/vl-class-paper2.ps
[7]http://www-formal.stanford.edu/eyal/nmr/m4.ps
[8]http://www-formal.stanford.edu/jmc/mcchay69.html
[9]http://www-formal.stanford.edu/jmc/ascribing.html
[10]http://www-formal.stanford.edu/jmc/concepts.html

[McC80] John McCarthy. Circumscription—A Form of Non-Monotonic Reasoning[11]. *Artificial Intelligence*, 13:27–39, 1980. Reprinted in [McC90].

[McC86] John McCarthy. Applications of Circumscription to Formalizing Common Sense Knowledge[12]. *Artificial Intelligence*, 28:89–116, 1986. Reprinted in [McC90].

[McC90] John McCarthy. *Formalization of common sense, papers by John McCarthy edited by V. Lifschitz*. Ablex, 1990.

[McC93] John McCarthy. Notes on Formalizing Context[13]. In *IJCAI-93*, 1993. Available on http://www-formal.stanford.edu/jmc/.

[McC96a] John McCarthy. From Here to Human-Level AI[14]. In *KR96 Proceedings*, 1996. Available as http://www-formal.stanford.edu/jmc/human.html.

[McC96b] John McCarthy. Making Robots Conscious of their Mental States[15]. In Stephen Muggleton, editor, *Machine Intelligence 15*. Oxford University Press, 1996. to appear, available on http://www-formal.stanford.edu/jmc/.

[McC96c] John McCarthy. Missionaries and Cannibals: Making it Elaboration Tolerant[16]. In *McCarthy's web page*, 1996. Available as http://www-formal.stanford.edu/jmc/cs323/missionaries.ps.

[MH69] John McCarthy and Patrick J. Hayes. Some Philosophical Problems from the Standpoint of Artificial Intelligence[17]. In B. Meltzer and D. Michie, editors, *Machine Intelligence 4*, pages 463–502. Edinburgh University Press, 1969.

[Rei80] Raymond Reiter. A Logic for Default Reasoning[18]. *Artificial Intelligence*, 13 (1–2):81–132, 1980.

[11] http://www-formal.stanford.edu/jmc/circumscription.html
[12] http://www-formal.stanford.edu/jmc/applications.html
[13] http://www-formal.stanford.edu/jmc/context.html
[14] http://www-formal.stanford.edu/jmc/human.html
[15] http://www-formal.stanford.edu/jmc/consciousness.html
[16] http://www-formal.stanford.edu/jmc/cs323/missionaries.ps
[17] http://www-formal.stanford.edu/jmc/mcchay69.html
[18] Sent a request to Ray for the paper

[SS94] Erik Sandewall and Yoav Shoham. Nonmonotonic Temporal Reasoning[19]. In *Handbook of Logic in Artificial Intelligence and Logic Programming, Volume 4: Epistemic and Temporal Reasoning.* Oxford University Press, 1994.

/@steam.stanford.edu:/u/jmc/f96/cs323.tex: begun Thu Dec 12 16:32:43 1996, latexed February 13, 1997 at 12:17 p.m.

[19]Reference requested from Yoav 1/2/97