Philosophy of AI (Fall 2001)

Selmer Bringsjord
Department of Cognitive Science
Department of Computer Science
Rensselaer Polytechnic Institute Troy NY 12180
Office Hours: Sage 3210, M Th 2–3 & by appt.
(518) 276-8105 (P) ● (518) 276-4871 (F)
selmer@rpi.edu ● `http://www.rpi.edu/∼brings`

August 26, 2001

1   Course Description

This course may be roughly divided into three general areas:

- Philosophical problems *in* AI and Cognitive Science;
- Philosophical debates which arise *because* of AI & Cog Sci;
- AI/Cog Sci research that is at least partly philosophical in nature.

An example from 1. is the problem of adjudicating between two competing camps within AI, the "connectionists" (who focus on sub-symbolic aspects of cognition and computing architectures (such as artificial neural networks) that are "brain-like") versus the "symbolicists" or "logicists" (who focus on symbolic aspects of cognition and use logic).[1] An example from 2. is John Searle's famous attack on so-called "Strong" AI by way of his Chinese Room Argument, in which he claims that computers can't genuinely understand anything because they simply move meaningless symbols around. An example from 3. is John Pollock's seminal attempt to design and build a person, the first stage of which has included building a computer system — OSCAR — which reasons defeasibly.[2]

Because the backgrounds of students in this course are so varied, the first three weeks of the course will be an introduction to relevant parts of logic, computability theory and the philosophy of mind. For the logic introduction, students are encouraged to use their laptops in class to follow along in FITCH and other such systems. *Introduction to Logic* is a formal prerequisite, but those who haven't taken this course, or can't remember it well if they took it, shouldn't panic: the review of basic logic at the start of the course should be sufficiently helpful. This is as good a place as any to note that there are so many problems and debates waiting for us, and so many of them involve advanced logic, philosophy, and mathematics, that we, in the space of this undergraduate course, can only scratch the surface.

Students are expected to try to *do* philosophy; the emphasis is on precise argumentation. Understanding, memorizing, etc. the material to be covered isn't sufficient to achieve a passing grade: philosophical arguments must be produced. Class discussion, both in class and out (and out, often electronically: email addresses will be collected and a class list assembled), is mandatory. In fact, our fourth contact hour is constituted by "Scholarly Skywriting" (to use Stevan Harnad's expression) which will explained a bit later.

In keeping with tradition, the movie *Blade Runner* will be seen at the end of the course. Well, actually, I'm giving thought to showing either *The Matrix* or (if available in time) *AI* instead; we'll see. (Along the way we'll be studying various relevant video, not only from these three movies, but others, and probably from *Star Trek* as well.)

---

[1]The connectionist-logicist clash won't be covered in this rendition of the course. For my position, see Bringsjord, S. (1991) "Is the Connectionist-Logicist Clash One of AI's Wonderful Red Herrings?" *Journal of Experimental & Theoretical Artificial Intelligence* **3.4**: 319–349.

[2]Whether an issue is placed within 1. or 3. is sometimes a matter of taste.

2 Objectives

The objectives of the course are the following five:

1. Students will understand and remember some prominent philosophical problems in AI, and so will be prepared to tackle these problems in the future;

2. Students will gain appreciable skill at articulating philosophical arguments, counter-examples, etc., a skill which will serve them well in the marketplace;

3. Students will understand and remember the chief philosophical issues, arguments, etc. which arise because of AI;

4. Students will understand and remember the approach to "person building" taken by John Pollock, in part because of their hands-on exposure to OSCAR.

5. Some students taking the course will see fit to enroll in the *Minds & Machines Laboratory & Program*,[3] which is designed to give students marketable skills in that part of the information economy that relates to AI, cognitive systems engineering, human-computer interface and so on; and

6. A good time will be had by all!

3 Texts

Much of the Bringsjordian thought covered in this class can be found in detail in

- *What Robots Can & Can't Be* (WRC&CB) by Selmer Bringsjord

A précis of WRC&CB is on the Web in *Psycoloquy*, as are a number of reviews of the book (and, in some cases, responses to the reviews). Yes, this book is expensive, so it's recommended only.

There are three required books, viz.,

- *Artificial Intelligence and Literary Creativity: Inside the Mind of Brutus, a Storytelling Machine* (Br)

- *Cognitive Carpentry* by John Pollock (CC)

- *Shadows of the Mind* by Roger Penrose (SOTM)

Required readings will also include some handouts and papers, both to be placed on reserve (and often our web page). Handouts are largely synopses of positions worked out in detail elsewhere, often in *WRC&CB*, sometimes in the forthcoming book *Super-Minds: A Defense of Uncomputable Cognition* (by Bringsjord & Zenzen; Kluwer). Handouts are:

- (H1) "Logic Tools"
- (H2) "Necessary Rudiments from Computability Theory"
- (H3) "Computationalism"
- (H4) "Main Theses at Issue"
- (H7) "The Dilemma"
- (H8) "Proof of Incompatibilism"
- (H9) "The Free Will Disproof of Computationalism"
- (H10) "Selmer's Searlean Argument"

Here are the papers we'll be reading:

(T) "Computing Machinery & Intelligence," Alan Turing

(B1) "Could, How Could We Tell If, and Why Should—Androids Have Inner Lives?", Selmer Bringsjord

(S) "Minds, Brains and Programs," John Searle

(B2) "Searle on the Brink," Selmer Bringsjord (in the on-line journal *Psyche*)

---

[3]Information about the program can be obtained via a link at the top-level of Bringsjord's web site.

(B3) "The Zombie Attack on the Computational Conception of Mind," *Philosophy and Phenomenological Research*, Selmer Bringsjord

(B5) "Cognition Is Not Computation: The Argument From Irreversibility," *Synthese*, Selmer Bringsjord & Michael Zenzen

(B6) "The Logical Minds Manifesto" (the genesis of the book *In Defense of Logical Minds*), Selmer Bringsjord, Yingrui Yang, & Kelsey Rinella

(B7) "Animals, Zombanimals, and the Total Turing Test: The Essence of AI," Selmer Bringsjord & Clarke Caporale

(B8) "A Modalized Gödelian Disproof of "Strong" AI," Selmer Bringsjord and Kostas Arkoudas

(B9) "Real Robots and the Missing Thought Experiment in the Chinese Room Dialectic," Selmer Bringsjord & Ron Noel, forthcoming in M. Bishop and J. Preston, eds., *The Chinese Room: New Essays on John Searle's Arguments against "Strong AI"*, Oxford University Press, Oxford, UK.

4 Grading

There will be three papers, balanced as follows. first paper 20% (3 pages in length); second 30% (4 pages); third paper 40% (5 pages). The first two papers will be on an assigned topic; the third paper can be on a topic of your choice, but must be cleared by me. (Some students will be allowed to do a software project involving OSCAR instead of a paper. This option will be explained in class.) Trend of grades is important over and above the percentage weighting. Class discussion will count for 10%. I reserve the right to call upon students at any time "out of the blue."

5 Schedule

Date	Topic	Reading
August 27	General Orientation	NA
August 30	Logic I	(H1), (WRC&CB)
September 3	No Class (Labor Day)	
September 6	Logic I con.	(H1), (WRC&CB)
September 10	Logic II	(H1), (H2), (WRC&CB)
September 13	Computability Theory I	(H2), (WRC&CB)
September 17	Computability Theory II	(H2), (WRC&CB)
September 20	What is "Strong" AI?	(H3), (WRC&CB)
September 24	Main Theses at Issue	(H4), (WRC&CB)
September 27	Turing Test	(T)
October 1	Attack on TT	(B1)
	Paper 1 Out	
October 4	The Chinese Room	(S), (H10)
October 8	No Class (Midterm Break)	
October 11	The Chinese Room	(B9)
October 15	Zombies & Zombanimals	(B3), (B7)
October 18	Rationality	(B6)
October 22	The Argument From Irreversibility	(S)
October 25	The Argument From Irreversibility	(S)
	Paper 1 Due	
October 29	Free Will I	(WRC&CB), (H7)
November 1	Free Will II	(WRC&CB), (H8), (H9)
November 5	AI, Creativity & Betrayal	(Br)
	Paper 2 Out	
November 8	AI, Creativity & Betrayal	(Br)
November 12	OSCAR	(CC), Chs. 1, 2
November 15	OSCAR	(CC), Ch. 3, 4
November 19	No Class (Thanksgiving)	
November 22	No Class (Thanksgiving)	
November 26	Modalized Gödelian Disproof	(SOTM), (B8)
	Paper 2 Due	
November 29	Modalized Gödelian Disproof	(SOTM), (B8)
December 3	*Blade Runner* ∨ *The Matrix* ∨ *AI* (AI & Emotion)	(B2)
December 6	*Blade Runner* ∨ *The Matrix* ∨ *AI*; Review	
December ?	Paper 3 Due by 5:00PM	

6 Academic Honesty

Student-teacher relationships are built on mutual respect and trust. Students must be able to trust that their teachers have made responsible decisions about the structure and content of the course, and that they're conscientiously making their best effort to help students learn. Teachers must be able to trust that students do their work conscientiously and honestly, making their best effort to learn. Acts that violate this mutual respect and trust undermine the educational process. They counteract and contradict our very reason for being at Rensselaer and will not be tolerated. Any student who engages in any form of academic dishonesty will receive an F in this course and will be reported to the Dean of Students for further disciplinary action. (The *Rensselaer Handbook* defines various forms of Academic Dishonesty and procedures for responding to them. All of these forms are violations of trust between students and teachers. Please familiarize yourself with this portion of the handbook.)