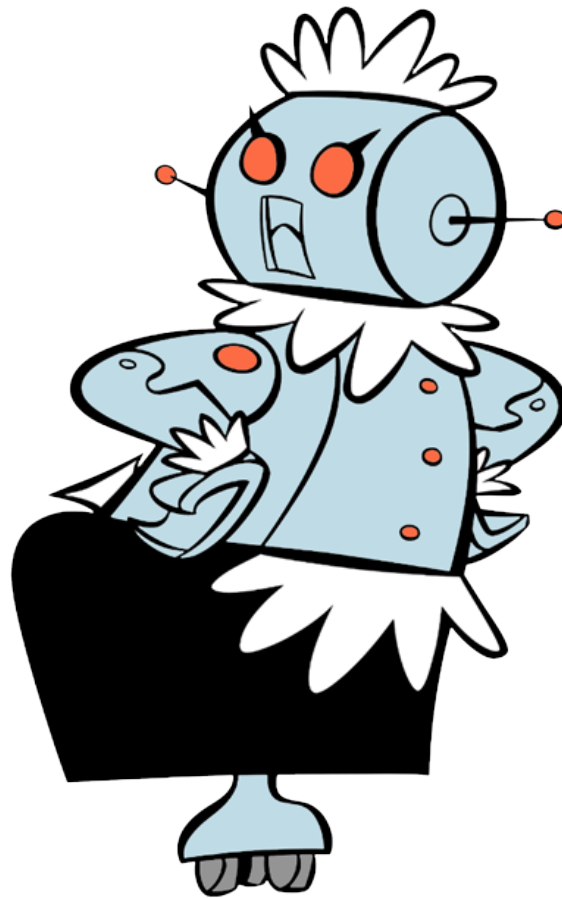


Mental Machines

Summer 2020

Gabe Dupre

April 8, 2020



Instructor: Gabe Dupre

E-mail: g.g.dupre@reading.ac.uk

Lecture: Pre-recorded, and posted to Blackboard on Wednesdays

Seminar: 11am-12noon Thursday, via MS Teams

Course Overview

The metaphor of the mind as a machine has been guiding naturalistic investigation of the mind for centuries. However, developments in the theory of computation in the 20th Century have made possible the idea that this claim may be literally true, not merely metaphorically suggestive. That is, the mind may be a machine. More precisely, the mind may be a computer: a device for processing and manipulating the flow of environmental information. We will look at the classical approach to cognitive science, *computationalism*, which aims to understand the mind precisely in these terms. We will then turn to perhaps the central problem for computational theories of mind: *the frame problem*. This is, roughly, the problem of figuring out the relevance of novel information to previously stored knowledge. We will then turn to a radical alternative to the classical approach to computationalism, based on recent developments in computer science under the banner of *Deep Learning* and *Big Data*. We will close with some skeptical consideration of this novel approach.

In this course, we will be concerned with questions about what the human mind is, how it relates to the operation of various kinds of machines, and what this tells us about our place in nature, including questions about the possibility of artificial intelligence. Along the way, we will draw on classical and contemporary work in philosophy, cognitive science and psychology, and computer science. section*Course Difficulty While there are no prerequisites for this course, it will be intellectually demanding— requiring you to master novel theoretical concepts and critically engage with complex arguments in your own writing. As with many other courses, to do well in PP3SPA you will need to:

- Attend all lectures and seminars and take (mental or physical) notes;
- Complete all the course readings, slowly and carefully, typically multiple times;
- Prepare your presentation in advance, reflecting carefully on reading and lectures;
- Work hard on your papers, including proof-reading and re-drafting multiple times, and submit them on time.
- **Optional, but highly recommended:** Meet with me to discuss questions regarding the course material or your paper topics.

This class is not about memorizing facts or correctly reproducing slogans; this class is about developing your own opinions and critical perspective on the topics discussed and enriching your ability to articulate and defend those opinions in a written form. Students who are prepared to work hard, challenge themselves, and attend lecture and section regularly will do well.

For pro-tips on reading and writing philosophy:

<https://sites.google.com/a/wellesley.edu/pinkguidetophilosophy/>

Course Materials

There will be no official textbook for the course. All of the required readings will be available digitally through the course webpage. Please check the course website regularly for updates.

If you are interested in reading further into a topic than the assigned readings, please let me know and I can suggest some material. Another good place to look is the Stanford Encyclopedia of Philosophy.

Course Requirements

1. **Final Paper:** 90% of final grade.
2. **Discussion Board Contribution** 10% of final grade.
3. **Be familiar with and abide by the University of Reading's policy on Academic Integrity:** This policy can be found at:
<https://libguides.reading.ac.uk/academicintegrity/about>
 - Students needing an academic accommodation based on a disability should notify the Disability Advisory Service (disability@reading.ac.uk, 0118 378 4202). When possible, students should contact the DAS within the first two weeks of the quarter, as reasonable notice is needed to coordinate accommodations.

4. **Late assignments:** Late papers will be docked by 10% for every 24 hour period after the due date up to 5 days. After this, the paper will receive no credit. Extensions may be granted if you have a valid reason, For an extension, you must contact the School of Humanities. For details see the course website.

Course Structure

Due to the continuing spread of Covid19, this course will take place entirely online. Lectures will be recorded and uploaded to the course website no later than Wednesday afternoon. These lectures are *mandatory* and must be viewed before seminar. Seminar will be conducted via MS Teams, from 11am until Noon on Thursdays. Just like lecture, seminar attendance is *mandatory*.

Seminars will give us a chance to discuss the material together. We will aim to both clarify the week's material and evaluate it in our own terms.

I will not be holding regular office hours, but am more than happy to (digitally) meet and discuss any aspect of the course with you, either individually or in groups. If you do have any questions, please e-mail me and we can set up a meeting via Skype, Zoom, MS Teams, or any other medium which is easiest for you. I will aim to respond to any e-mails within 48 hours.

Obviously, moving our entire course online is not ideal. In-person interaction in lecture, seminar, and office hours is not a fully replaceable aspect of a philosophical education. But, we will do everything we can to minimize the negatives. To this end, I will be setting up weekly discussion boards in which students can post questions and thoughts they had about the readings, and respond to the posts of other students. The discussions on these boards will provide guidance for our weekly seminars. I would like to see all students contributing regularly to these discussions.

Classroom Etiquette

Participation is strongly encouraged. The best way to learn philosophy is to *do* philosophy, and this involves discussion and argumentation. Such discussion may get heated and this is OK. However, what is not OK is behavior that discourages other students from engaging. Finding the line between passionate debate and personal attack is a vital skill for anyone in or outside of academia. As you will likely have noticed, discussions that take place in digital forums are not necessarily more charitable and collaborative than those that take place in person. For this reason, we must be extra careful for this course that our discussions are congenial and inviting for all students.

All other students in the classroom must be treated respectfully, as peers engaged in a collective activity. Behavior that will not be tolerated includes, but is not limited to: dismissing another student's opinion, talking over another student, personal attacks etc. In short: Don't be a jerk. Class discussion is every student's opportunity to engage with the material, and behavior which impedes this will not be tolerated.

Reading Schedule

Readings marked with an '*' are recommended, but not required. Readings must be done before lecture.

- **Week 1: The Mind as Machine**

- Fodor, J. *The Mind/Body Problem*
- * Crane, T. *The Mechanical Mind* (Chapter 2)

The view of the mind as a machine, in its contemporary form, developed in response to the failure of two philosophical/psychological programs of the early-to-mid 20th Century: behaviorism and the identity theory. This week we will outline exactly how these programs failed, and why a mechanistic, functionalist approach to the mind is able to retain the strengths of each approach, without taking on their weaknesses as well.

- **Week 2: The Mind as Computer**

- Block, N. *The Mind as the Software of the Brain*
- *
- * Crane, T. *The Mechanical Mind* (Chapter 3)

Saying that the mind is a machine is not hugely informative, unless we specify what kind of machine it is. Block presents the outlines of the computationalist theory of mind, according to which the mind is an information processing engine. This is the dominant program in contemporary psychology and cognitive science today.

- **Week 3: The Frame Problem**

- Dennett, D. *Cognitive Wheels: The Frame Problem for AI*

- * Haugeland, J. *An Overview of the Frame Problem*

According to the computationalist hypothesis, the mind is a computer. This suggests a plausible method for testing this hypothesis: compare the capacities of human minds and actual computers. If the human mind *is* a computer, we might predict to find analogous successes and failures. However, work on the so-called ‘Frame Problem’ suggests that this is not the case. It has proved remarkably difficult to design artificial systems which can perform tasks that human beings find trivial, and for principled reasons. This poses a deep challenge to the computational view.

- **Week 4: Deep Learning**

- Domingos, P. *The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World* (Chapter 1)

- * Sejnowski, T. *The Deep Learning Revolution* (Chapter 1)

One prominent response to the Frame Problem stems from the idea that the model of computers on which classical computationalism is based is itself outdated. Much recent work in computer science and engineering, especially in the Deep Learning and Big Data programs, suggests an alternative computational model for the mind. These programs stress the surprising abilities of relatively simple models to ‘learn’ complex capacities when trained on massive amounts of data.

- **Week 5: The Perils of ‘Big Data’**

- Marcus, G. and Davis, E. *Rebooting AI: Building Artificial Intelligence We Can Trust* (Chapter 4)

- * Dupre, G. *Notes on Big Data and the Explanatory Economy*

We will close by looking again at characteristic problems that machines face, and which seem far less troublesome to human minds. Gary Marcus argues that the Big Data revolution is unlikely to provide the insight into human psychology promoted by some of its advocates, by focusing again on recurrent failures of machines to learn certain sorts of skills, including human language. This suggests that the classical picture of computationalism may still be needed.