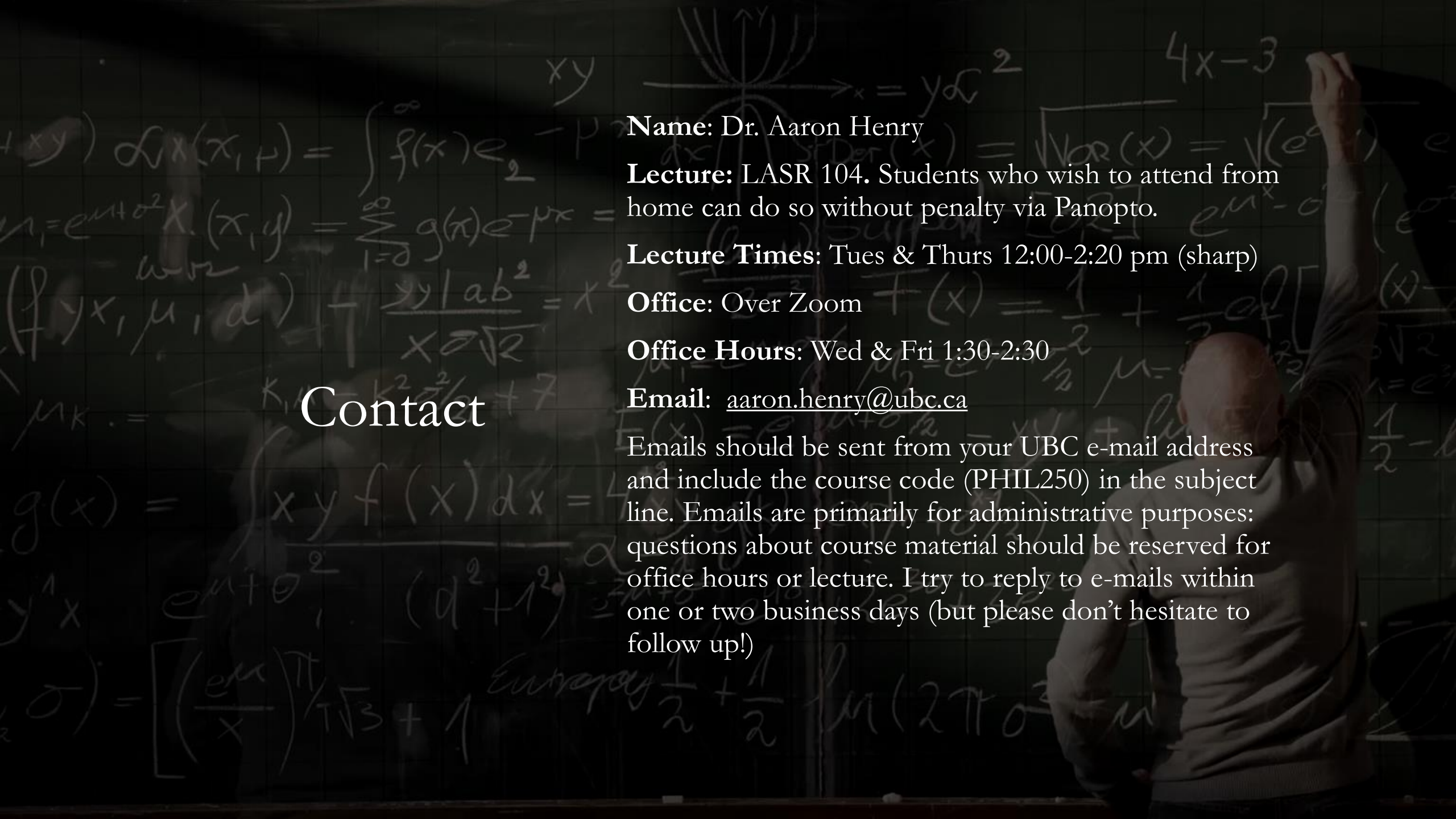
PHIL351: Philosophical Perspectives on Cognitive Systems Research

Dr. Aaron Henry
University of British Columbia
Term 1 2023

Plan

Syllabus

Short lecture on the philosophical history of cognitive science



Contact

Name: Dr. Aaron Henry

Lecture: LASR 104**.** Students who wish to attend from home can do so without penalty via Panopto.

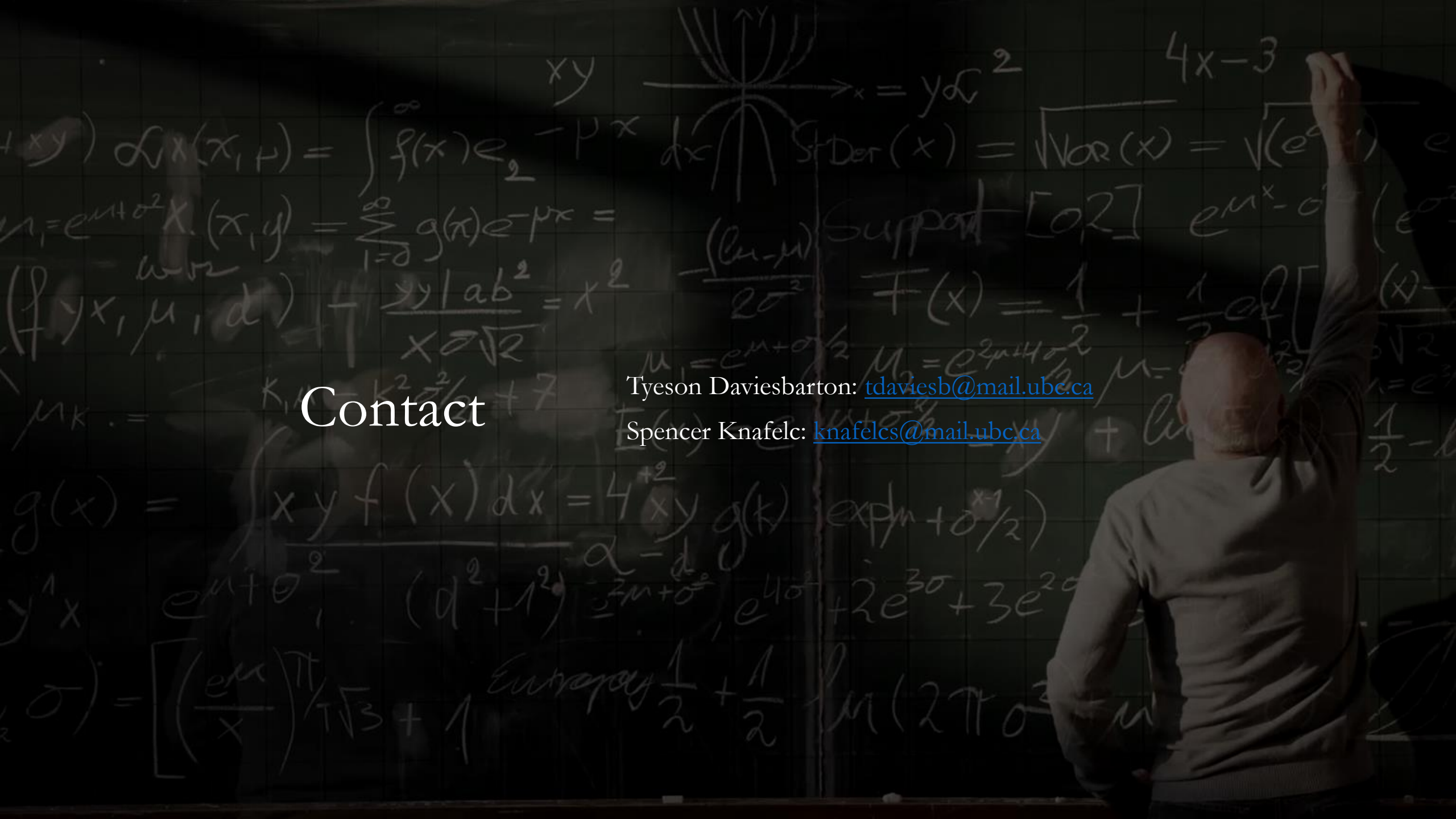
Lecture Times: Tues & Thurs 12:00-2:20 pm (sharp)

Office: Over Zoom

Office Hours: Wed & Fri 1:30-2:30

Email: aaron.henry@ubc.ca

Emails should be sent from your UBC e-mail address and include the course code (PHIL250) in the subject line. Emails are primarily for administrative purposes: questions about course material should be reserved for office hours or lecture. I try to reply to e-mails within one or two business days (but please don't hesitate to follow up!)



Contact

Tyeson Daviesbarton: tdaviesb@mail.ubc.ca

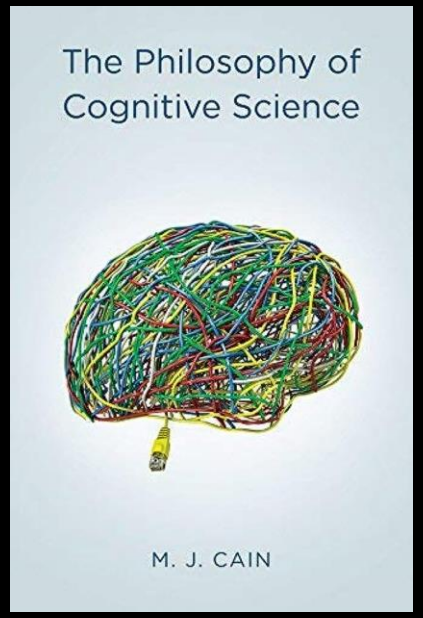
Spencer Knafelc: knafelcs@mail.ubc.ca

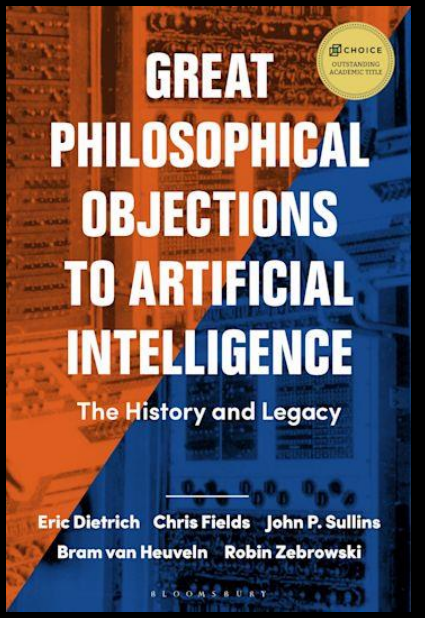
READINGS

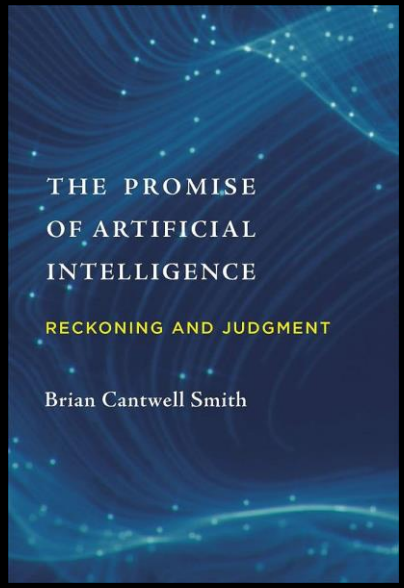
All course readings will be available through the course website or electronically through the UBC library. We will be drawing frequently on the following:

- Cain, M. J. (2016). *The philosophy of cognitive science*. Polity.
- Dietrich, E., Fields, C., Sullins, J. P., Van Heuveln, B., & Zebrowski, R. (2021). *Great Philosophical Objections to Artificial Intelligence: The History and Legacy of the AI Wars*. Bloomsbury Academic.
- Smith, B. C. (2019). *The Promise of Artificial Intelligence: Reckoning and Judgment*. MIT Press.

If you cannot find one of the assigned readings on Canvas or through the UBC library, please don't hesitate to email me letting me know.







Assessment

Reflection paper	(15% of final)	Oct 6th, 11:59 PM
Midterm	(25% of final)	Oct 19
Final paper	(25% of final)	Dec 8th, 11:59 PM
Final exam	(35% of final)	TBD

Policies

Course Website

All announcements and course documents will be posted on Canvas. To access this site, go to https://canvas.ubc.ca/and login with your CWLid and password. PHIL351 will appear under the "courses" portion of the welcome page, on the left hand side. Click on the link to access our site. You should check this site regularly for updates.

All papers are to be submitted (and will be returned) via Canvas.

Late work

Assignments will be penalized 1/3 a letter grade for each day that they are late. Extensions may be granted under special circumstances, but students should contact me to request an extension before the due date. Assignments more than 5 days late will not be accepted.

Academic Integrity

Academic integrity is essential to the pursuit of learning and scholarship in a university, and to ensuring that a degree from the University of British Columbia is a strong signal of each student's individual academic achievement. Accordingly, the University treats cases of cheating and plagiarism very seriously. Plagiarism, which is intellectual theft, occurs where an individual submits or presents the oral or written work of another person as his or her own. Scholarship quite properly rests upon examining and referring to the thoughts and writings of others. However, when another person's words (i.e. phrases, sentences, or paragraphs), ideas, or entire works are used, the author must be acknowledged in the text, in footnotes, in endnotes, or in another accepted form of academic citation. Where direct quotations are made, they must be clearly delineated (for example, within quotation marks or separately indented). Failure to provide proper attribution is plagiarism because it represents someone else's work as one's own. Plagiarism should not occur in submitted drafts or final works. A student who seeks assistance from a tutor or other scholastic aids must ensure that the work submitted is the student's own. Students are responsible for ensuring that any work submitted does not constitute plagiarism. Students who are in any doubt as to what constitutes plagiarism should consult their instructor before handing in any assignments. A link about Academic Misconduct:

http://www.calendar.ubc.ca/Vancouver/index.cfm?tree=3,54,111,959

Use of generative AI tools (e.g., ChatGPT)

- If you make use of generative artificial intelligence tools (e.g., ChatGPT) to complete any course-related work, the generated material must be clearly and correctly indicated, and cited/referenced using an accepted referencing style. Failure to clearly indicate and reference AI-generated material will be reported as academic misconduct. You should consult your instructor if you have any questions about the use of generative AI tools.

Accessibility needs

The University of British Columbia is committed to accessibility. If you have a disability that may interfere with your ability to successfully take this course, then please email me in the first few weeks. You must also register with Access and Diversity, so that they can help provide support
https://students.ubc.ca/about-student-services/access-diversity

Support

See syllabus for various internet and university resources.

Reading for today

Cantwell-Smith, ‘Introduction’ (8 pages)

Cain, Chapter 1, §§1-3 (6 pages)

Optional:

Dietrich et al., Introduction (pp. 1-5)

Haugeland, The Sage of the Modern Mind (Chapter 1 of *Artificial Intelligence: The Very Idea*)

What is a ‘philosophical perspective on cognitive systems research’?

Natural interpretation: the philosophy of cognitive science.

… But what is that?

What is a ‘philosophical perspective on cognitive systems research’?

We can distinguish two recognizably philosophical projects:

- The foundations project: debates about the theoretical foundations for a science of mind.
  - E.g., what are the basic explanatory principles and constructs?
- The reconciliation project: debates about how (if, at all) a scientific perspective on the mind can be reconciled with our non-scientific (e.g., common-sense) perspectives on the mind.
  - Integrating the ‘scientific’ and ‘manifest’ image of ourselves.
- We should not assume that these two intellectual projects will be wholly independent.

What is a ‘philosophical perspective on cognitive systems research’?

There is a third recognizably philosophical project that is increasingly central in recent discussions of cognitive science: the project of ensuring that any cognitive systems we build – i.e., artificial intelligences (AIs) – are built *ethically*.

Thus, we might subdivide the philosophical questions we will be exploring in this course into those of:

i. the foundations project

ii. the reconciliation project

iii. the ethical AI project

‘Cognitive science’

Question: What do you take this mean?

‘Cognitive science’

Potential answers:

- The scientific study of mind.
  - Isn’t that just ‘psychology’?
- The *interdisciplinary* scientific study of mind (classically: psychology, linguistics, computer science, artificial intelligence, and – yes – philosophy).
  - This is better … But it suggests a *mere diversity* of approaches, whereas the pioneers of cognitive science envisioned something theoretically integrated.
- The *theoretically integrated* interdisciplinary study of mind.
  - Pretty good. But now we should ask what this ‘theoretical integration’ consists in.

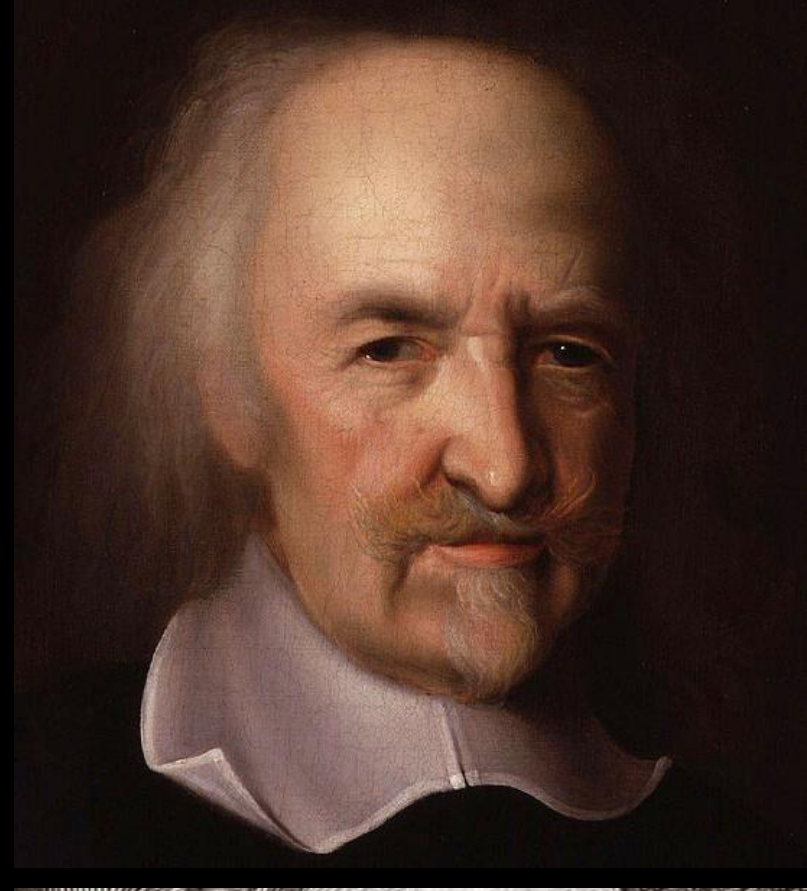
‘Cognitive science’

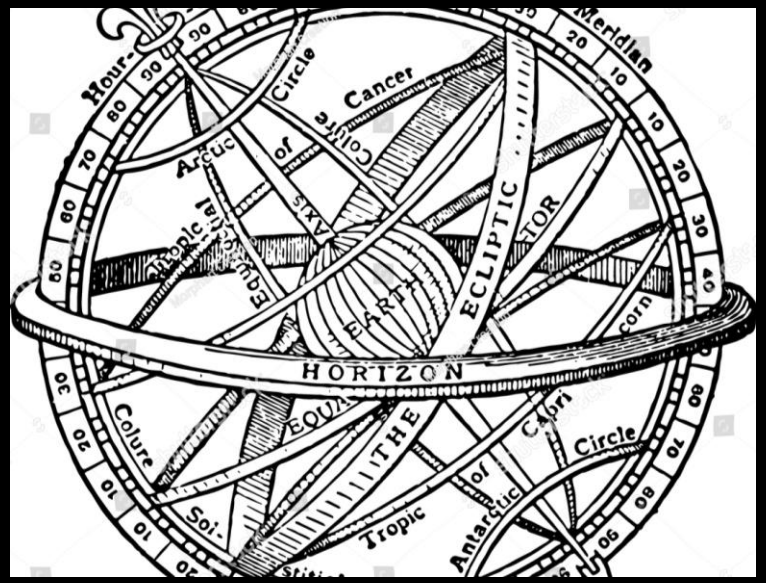
According to what we might call ‘classical’ cognitive science, behind the seeming diversity of approaches to the scientific study of mind there lies a coherently unified scientific understanding of cognition as *computation* or *information-processing*.

To fully understand the significance of this proposal, it helps to situate it historically.

A (pre-)pre-history of cognitive science

- The idea of cognitive science can be traced back to the scientific revolution of the 17th century when people began to explain seemingly heterogeneous phenomena as manifestations of the same observation-based, mathematically formulated, laws – e.g., Newton's laws of motion and universal gravitation.





The Aristotelian-Ptolemaic cosmology

- Earth at the center, surrounded by the ‘heavens’ above.
- A ‘great chain of being’ with different principles characterizing different kinds of entity.
- Everything seeking its natural place according its respective internal principle (teleology or goal-directedness)

The scientific revolution

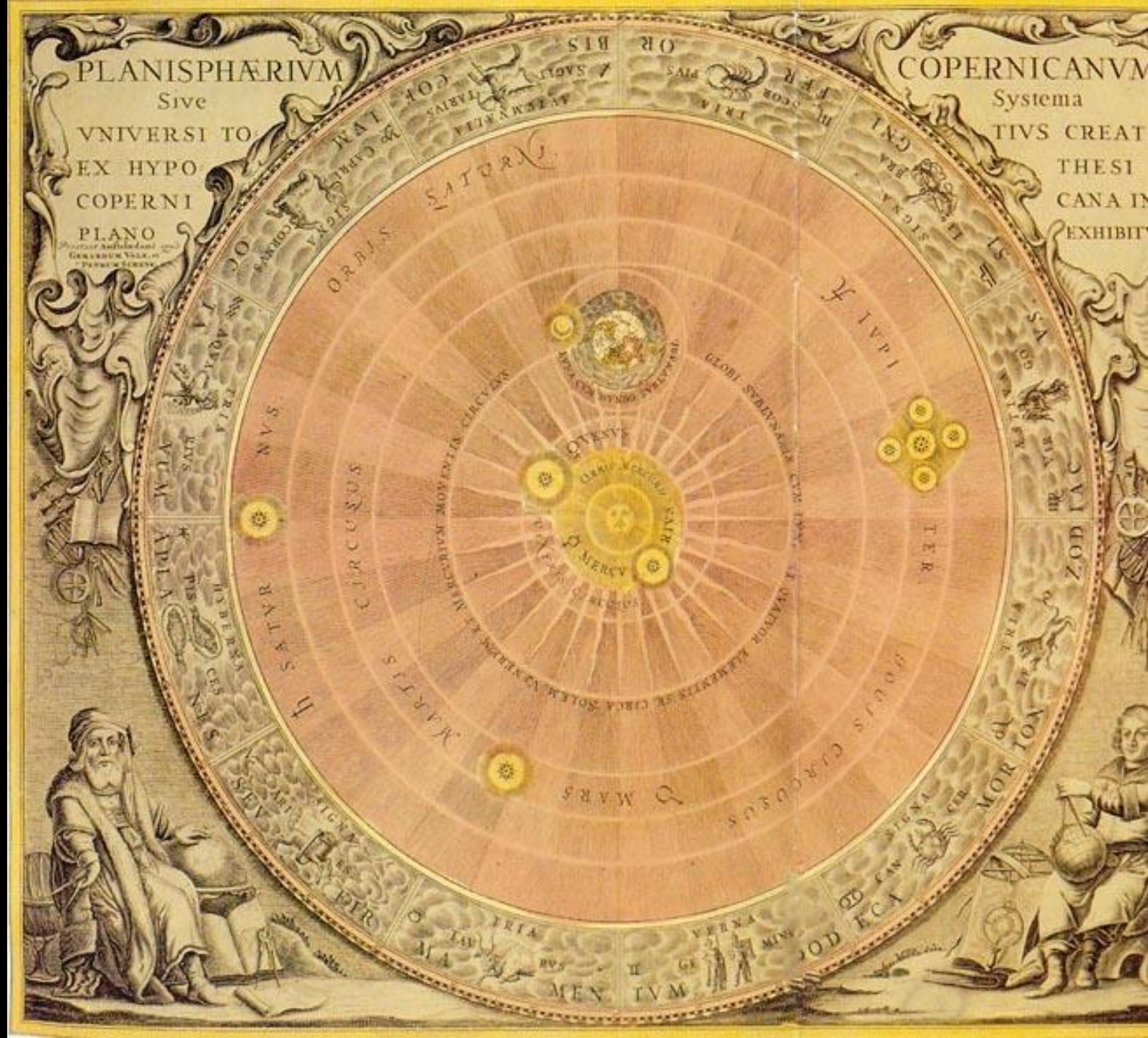
Copernicus (1473-1543)

Brahe (1546-1601) and Kepler (1571-1630)

Galileo (1564-1642)

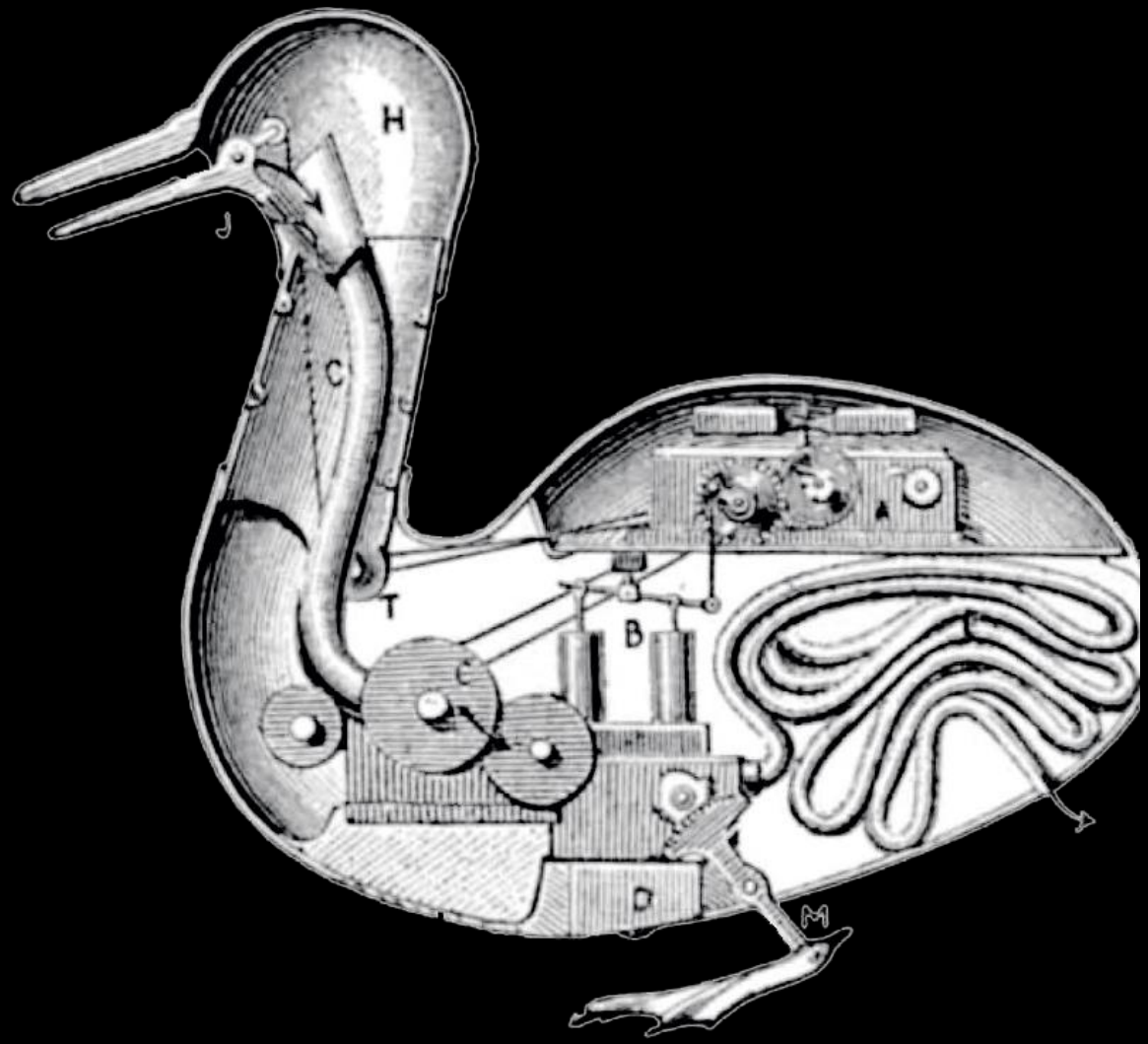
Newton (1642-1726)

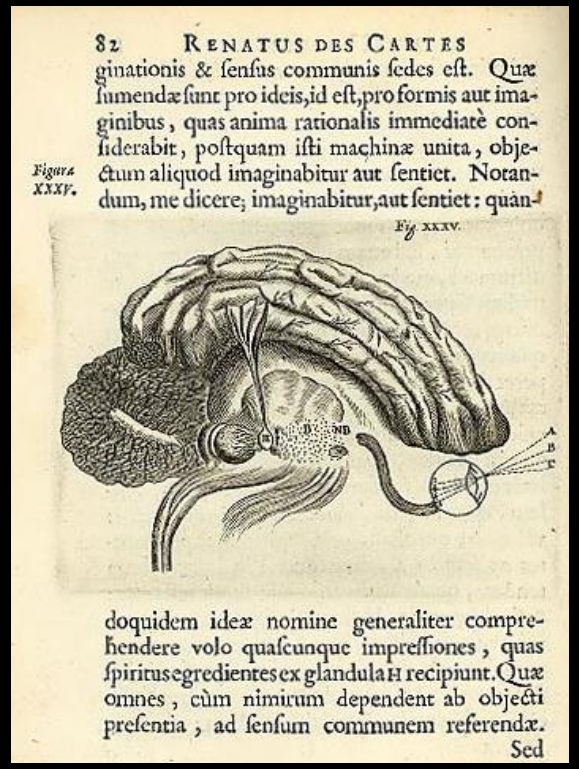
- Continuities:
  - Explaining diverse phenomena using general principles
  - Partly non-empirical inspiration
- Discontinuities:
  - Replacement of teleology/goal-directedness with external forces
  - The use algebra to represent abstract geometric relationships
  - "The book of nature is written in the language of mathematics"
  - Unparalleled degree of predictive accuracy and precision *and* explanatory unification – e.g., heavenly bodies obey the same laws as terrestrial bodies (despite appearances to the contrary).



The impact of mechanical artefacts (e.g., clocks): complex wholes can be analyzed into an organized set of interacting components (which can then be formalized using precise, quantitative mechanical laws).

Together with the popularity of animal automata, this fueled the development of 'clockwork' models of living systems and of the physical universe more generally.





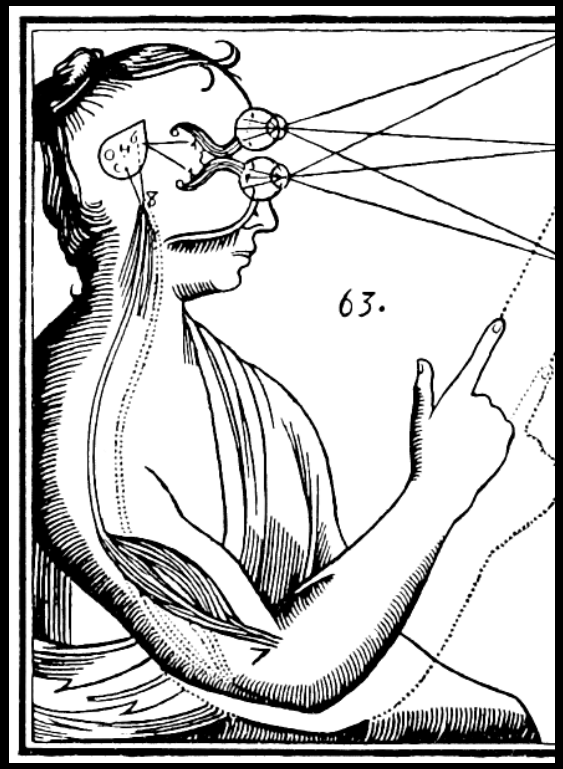
82 RENATUS DES CARTES

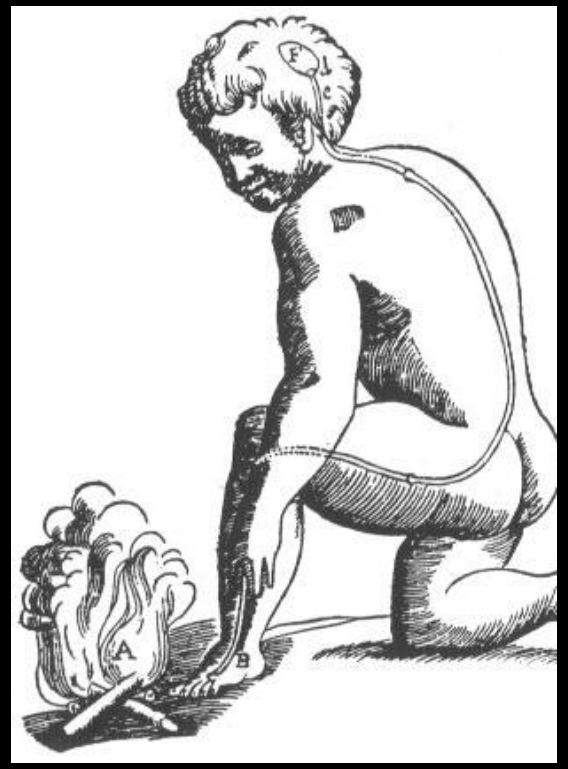
ginationis & ſenſus communis ſedes eſt. Quæ ſumendæ ſunt pro ideis, id eſt, pro formis aut imaginibus, quas anima rationalis immediatè conſiderabit, poſtquam iſti machinæ unita, objectum aliquod imaginabitur aut ſentiet. Notandum, me dicere; imaginabitur, aut ſentiet: quàn-

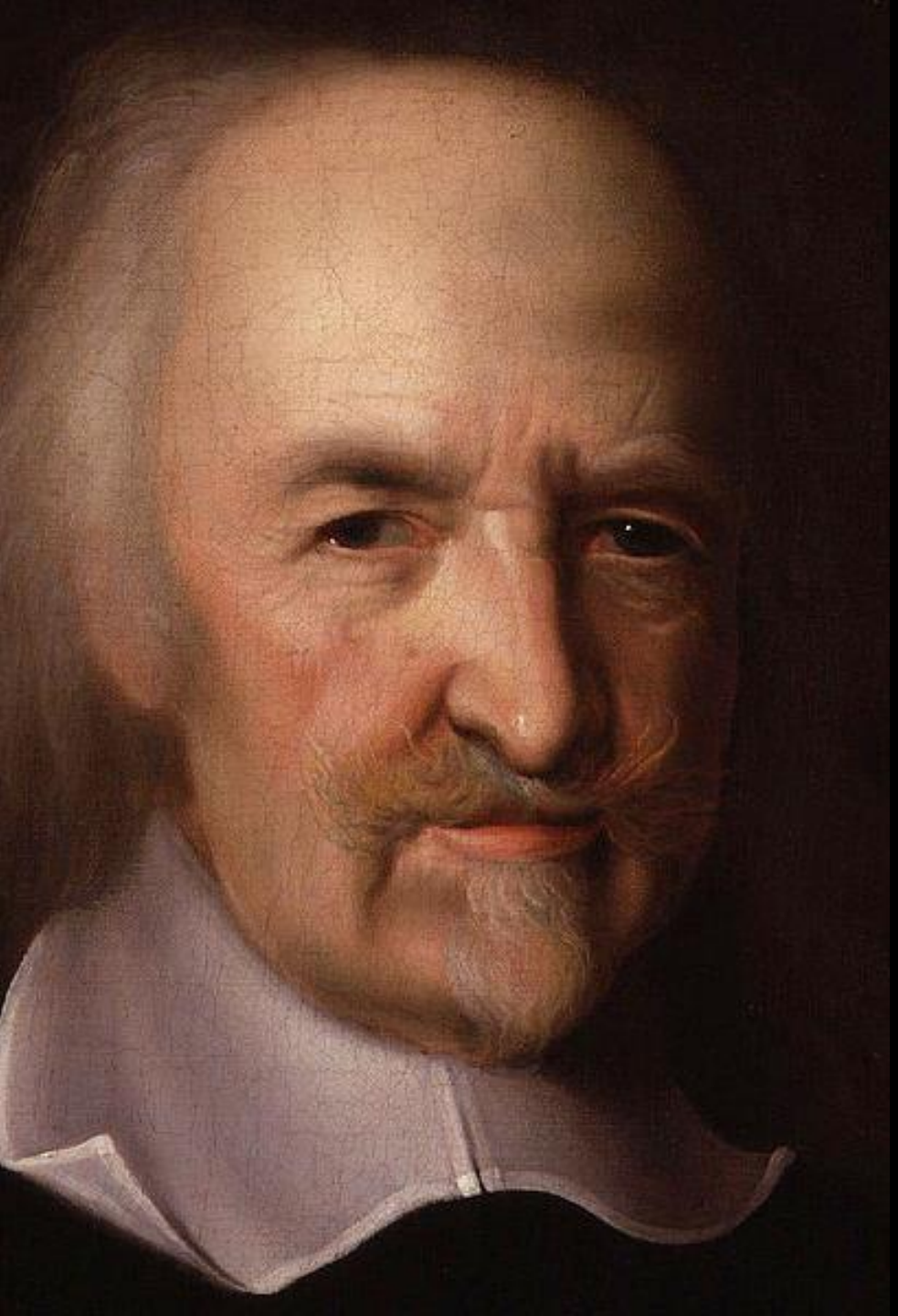
Figura XXXV.

Fig XXXV.

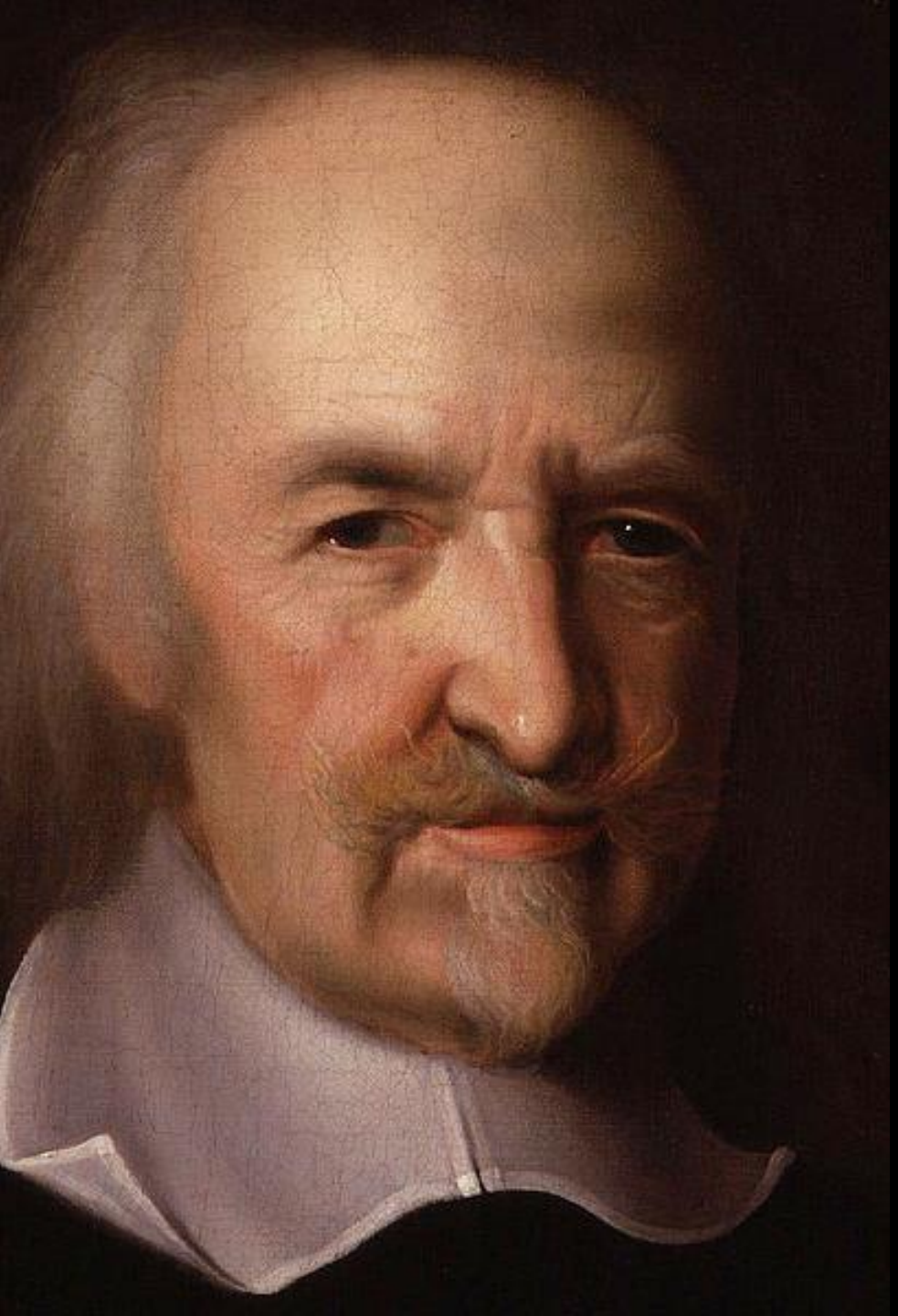
doquidem ideæ nomine generaliter comprehendere volo quaſcunque impreſſiones, quas ſpiritus egredientes ex glandula H recipiunt. Quæ omnes, cùm nimirum dependent ab objecti preſentia, ad ſenſum communem referendæ. Sed







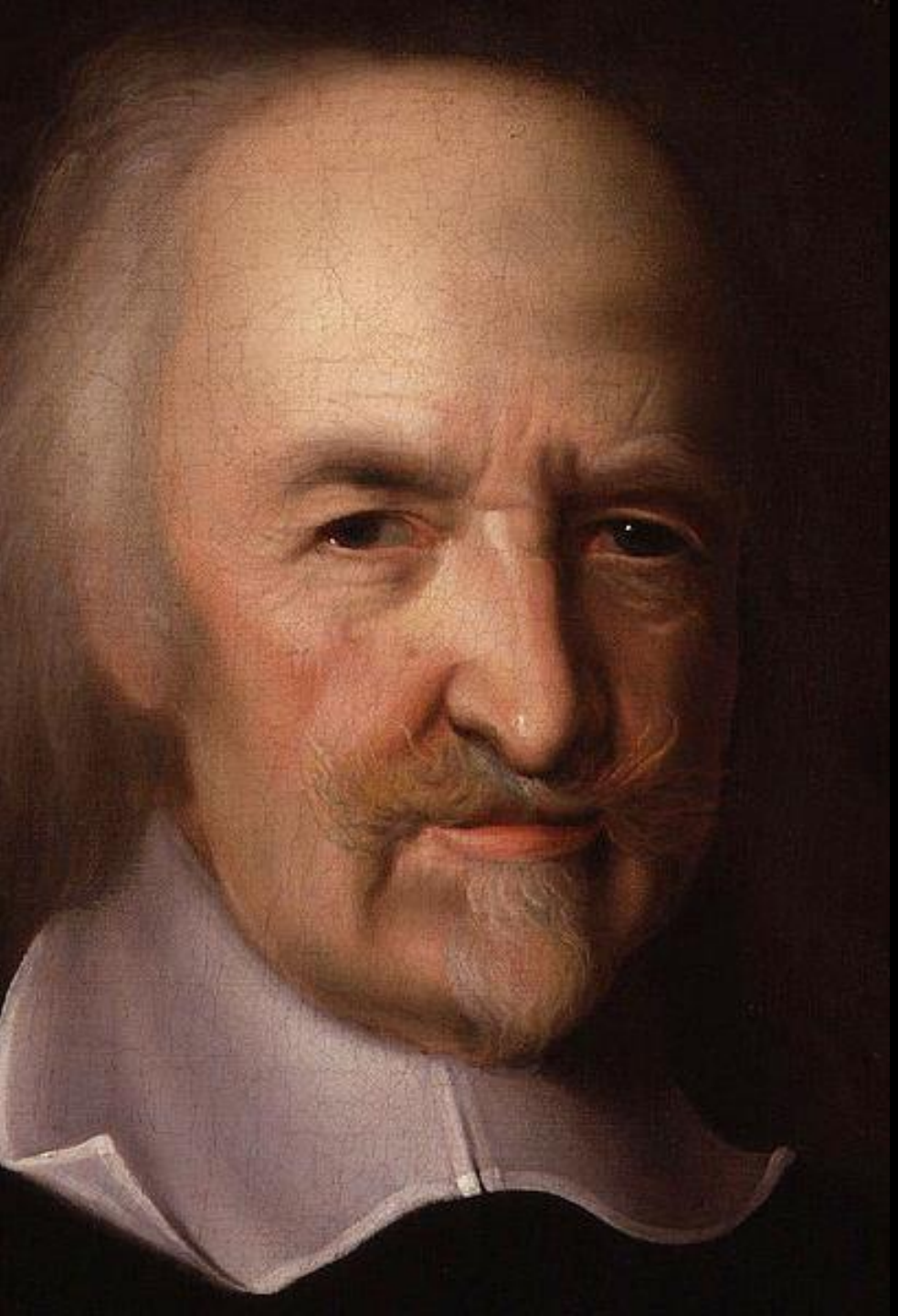
Thomas Hobbes (1588-1679)



Thomas Hobbes (1588-1679)

Reason Defined

Out of all which we may define […] what […] is meant by this word Reason […] Reason […] is nothing but Reckoning (that is, Adding and Substracting) of the Consequences of generall names agreed upon, for the Marking and Signifying of our thoughts; I say Marking them, when we reckon by our selves; and Signifying, when we demonstrate, or approve our reckonings to other men. (Hobbes, *Leviation*, Book I)



Thomas Hobbes (1588-1679)

"By ratiocination, I mean *computation*. Now to compute is either to collect the sum of many things that are added together, or to know what remains when one thing is taken out of another. Ratiocination, therefore, is the same with addition and subtraction … We must not think that computation, that is ratiocination, has a place only in numbers, as if man were distinguished from other living creatures ... by nothing but the faculty of numbering; for *magnitude, body, motion, time ... action, conception, ... speech* and *names* ... are capable of addition and subtraction. Now such things as we add or subtract, ... we are said to *consider*... to *compute, reason* or *reckon.* (Hobbes, *Elements of Philosophy Concerning Body,* 1656)

La Mettrie (1709-1751)

"Let us then conclude boldly that man is a machine and that there is in the whole universe only one diversely modified substance."

"[T]he human body is a clock … [S]ince all the soul's faculties depend so much on the specific organization of the brain and of the whole body that they are clearly nothing but that very organization, the machine is perfectly explained!"

René Descartes (1596-1650)

Despite being an enthusiastic participant in the mechanistic turn (including its applications to human behaviour) Descartes argued that mechanical principles would be insufficient to explain the human capacity *to judge* and *to reason*. Contra Hobbes, reasoning is *not* mere 'reckoning.'

Furthermore, the specific considerations that Descartes raised against the possibility of a fully mechanistic account of human behaviour cast significant light on why cognitive science developed the way that it did. This is where we will begin next time.