WE ARE NOT JUST COMPUTERS

Angus Menuge, PhD

WORLDVIEWS:
COMPARING the Faith

BIOETHICS:
APPLYING the Faith

APOLOGETICS:
DEFENDING the Faith

— CONCORDIA —

PHILOSOPHY PROGRAM

Socrates statue at the Academy of Athens building in Athens, Greece

CONCORDIA UNIVERSITY
WISCONSIN

12800 North Lake Shore Drive Mequon, WI 53097
cuw.edu/Programs/Philosophy
262.243.4249

OVERVIEW

- 1. AI: Historical Background
- 2. AI: The Contemporary Scene
- 3. Limits of AI
- 4. What is so Special about Human Beings?

1. AI: HISTORICAL BACKGROUND

- Thomas Hobbes (1588-1679)
  - A materialistic view of man
  - Thinking is only motion in the brain
  - Reasoning is only "reckoning" (computation)

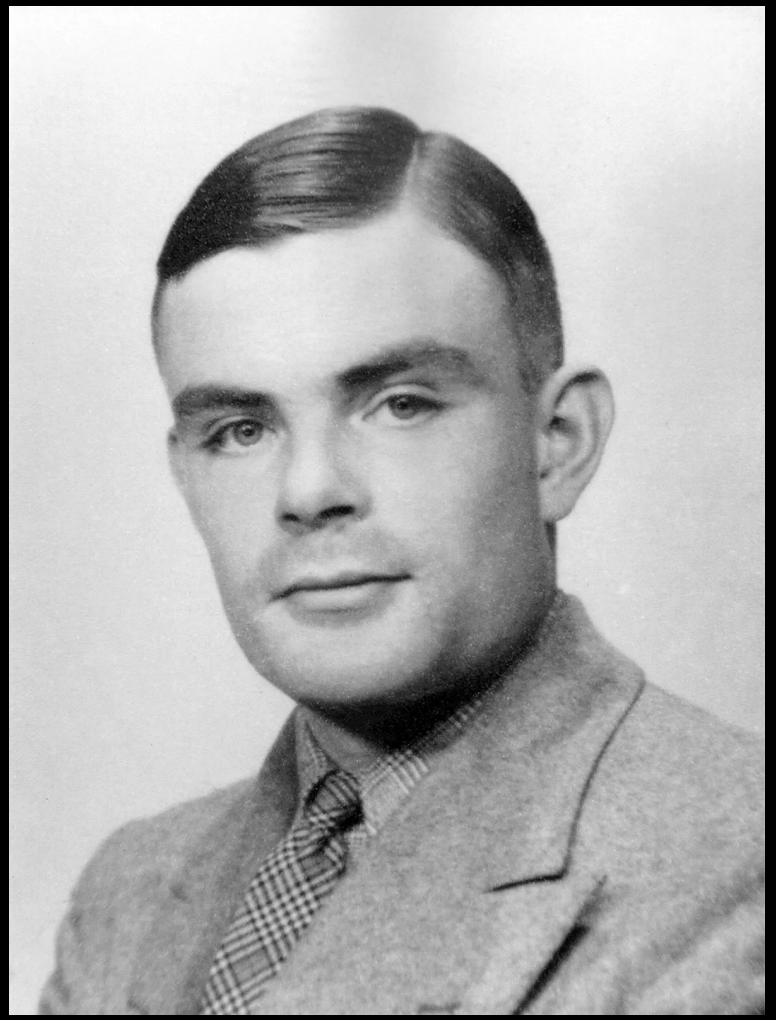
- Julien Onfray de la Mettrie (1709-1751)
  - Author of *L'homme machine* (Man a Machine)
  - Man is an organic machine
  - He has no soul

THE SOUL EXPLAINED AWAY

- "[S]ince all the faculties of the soul depend to such a degree on the proper organization of the brain and of the whole body, that apparently they are but this organization itself, **the soul is clearly an enlightened machine.**"
  --- La Mettrie, *Man a Machine* (La Salle, IL: Open Court, 1943), 128.

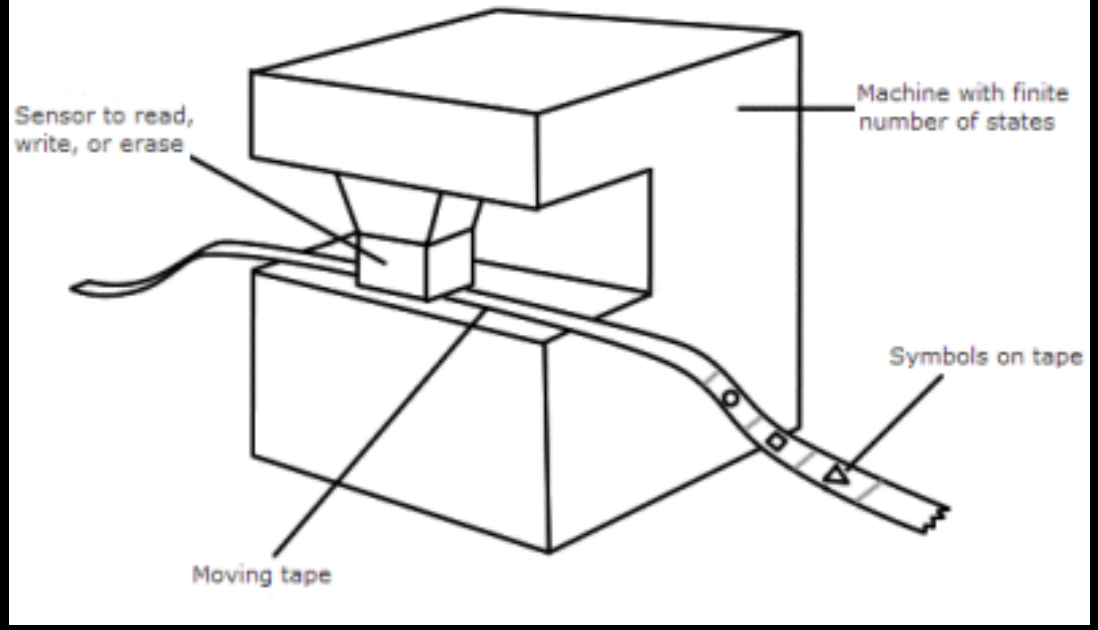
- "[T]he soul is but…the mainspring of the whole machine."
  --- La Mettrie, *Man a Machine*, 135.

ALAN TURING (1912-1954)

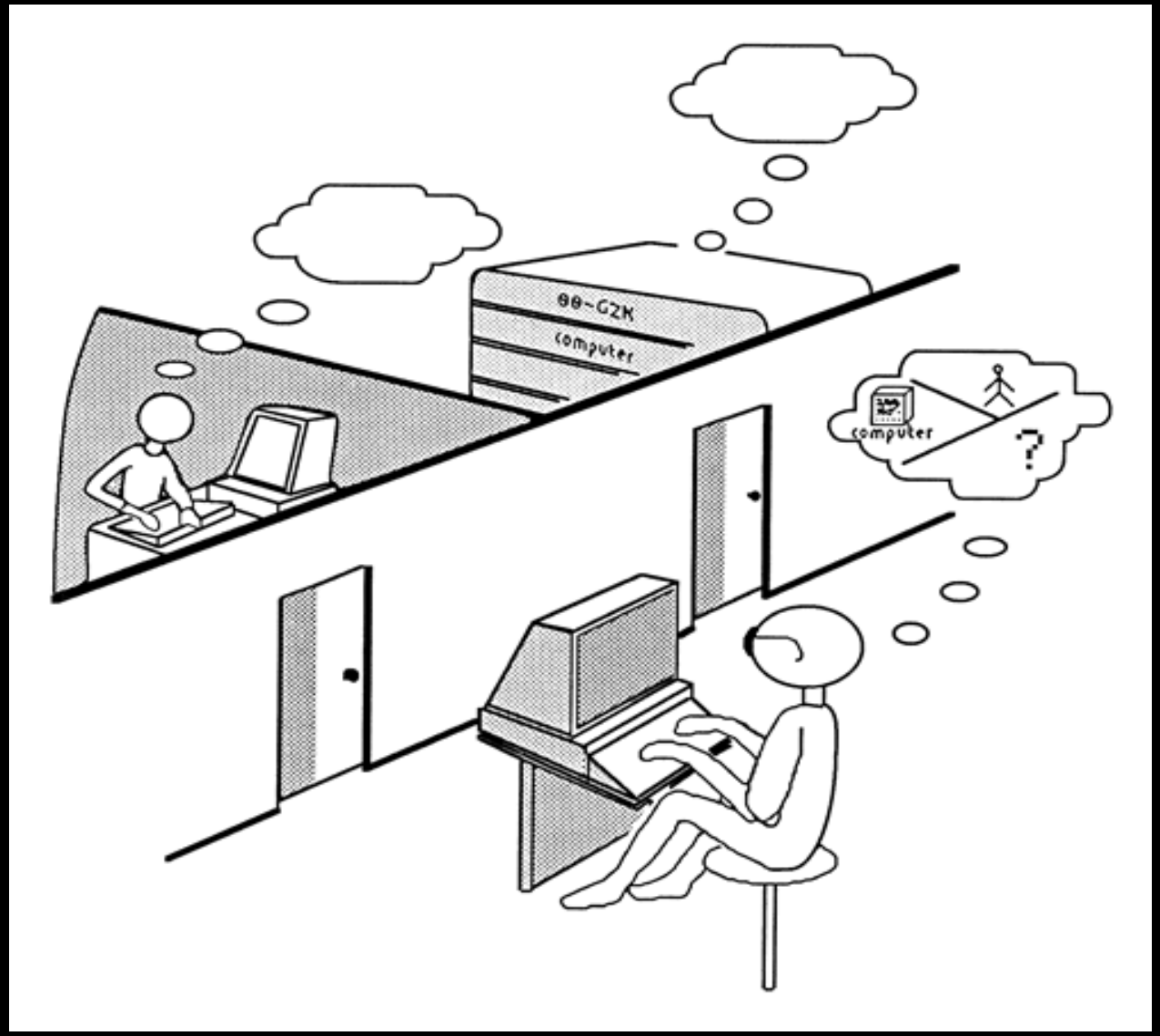


THE IDEA OF THE MODERN COMPUTER

A Turing machine is a theoretical generalized computer, composed of a tape on which symbols representing instructions are imprinted. The tape can move backwards and forwards in the machine, which can read the intructions and write the result-ant output back onto the tape.



THE TURING TEST



ACTING LIKE A HUMAN

- If a computer passes as a human as often as a human does, the computer is "intelligent"
- Can be adapted to handle multiple domains
- Treats intelligence as a behavioral concept:
  - A system is called intelligent if it produces behavior like ours

2. AI: THE CONTEMPORARY SCENE

- AI Paradigms:
  - 1. Weak AI
  - 2. Strong AI
  - 3. Artificial General Intelligence (AGI)

WEAK VS STRONG AI

- Weak AI
- AI is a "tool" for studying the mind because it simulates mental operations
  - A simulated tornado is not a tornado

- Strong AI
- "the appropriately programmed computer *is* a mind…computers given the right programs can be literally said to *understand* and have other cognitive states."

- John Searle, "Minds, Brains, and Programs," *Behavioral and Brain Sciences*, vol. 3, no. 3 (1980): 417-457.

ARTIFICIAL GENERAL INTELLIGENCE (AGI)

- Standard AI
  - Domain specific
    - e.g. chess, expert systems for medicine, law, geology
- AGI
  - Has **general purpose** problem-solving capacities that can be applied to many domains (even new ones)
- Many (wrongly) identify AGI with Strong AI
  - AGI may still only be a simulation (Weak AI)

What Would Count as AGI?

- "a capacity to learn…the ability to deal effectively with uncertainty… [and a] faculty for extracting useful concepts…for…logical and intuitive reasoning…."

--Nick Bostrom, *Superintelligence: Paths, Dangers, Strategies* (Oxford: Oxford University Press, 2014), 23.

CHATGPT

- "The basic concept of ChatGPT is…rather simple. Start from a huge sample of human-created text…. Then train a neural net to generate text that's 'like this.'"

  --Stephen Wolfram, *What is ChatGPT Doing…and Why Does it Work?* (Wolfram Media Inc, 2023), 75.

- Given a prompt, ChatGPT uses vast amounts of text to predict the probable completions, one "token" at a time

- To avoid being too "flat," it can deliberately select less probable completions

EXAMPLE

- Prompt: "The best thing about AI is its ability to"

- The best thing about AI is its ability to
- The best thing about AI is its ability to learn
- The best thing about AI is its ability to learn from
- The best thing about AI is its ability to learn from experience

- Whole essays and poems can be written this way!

3. LIMITS OF AI

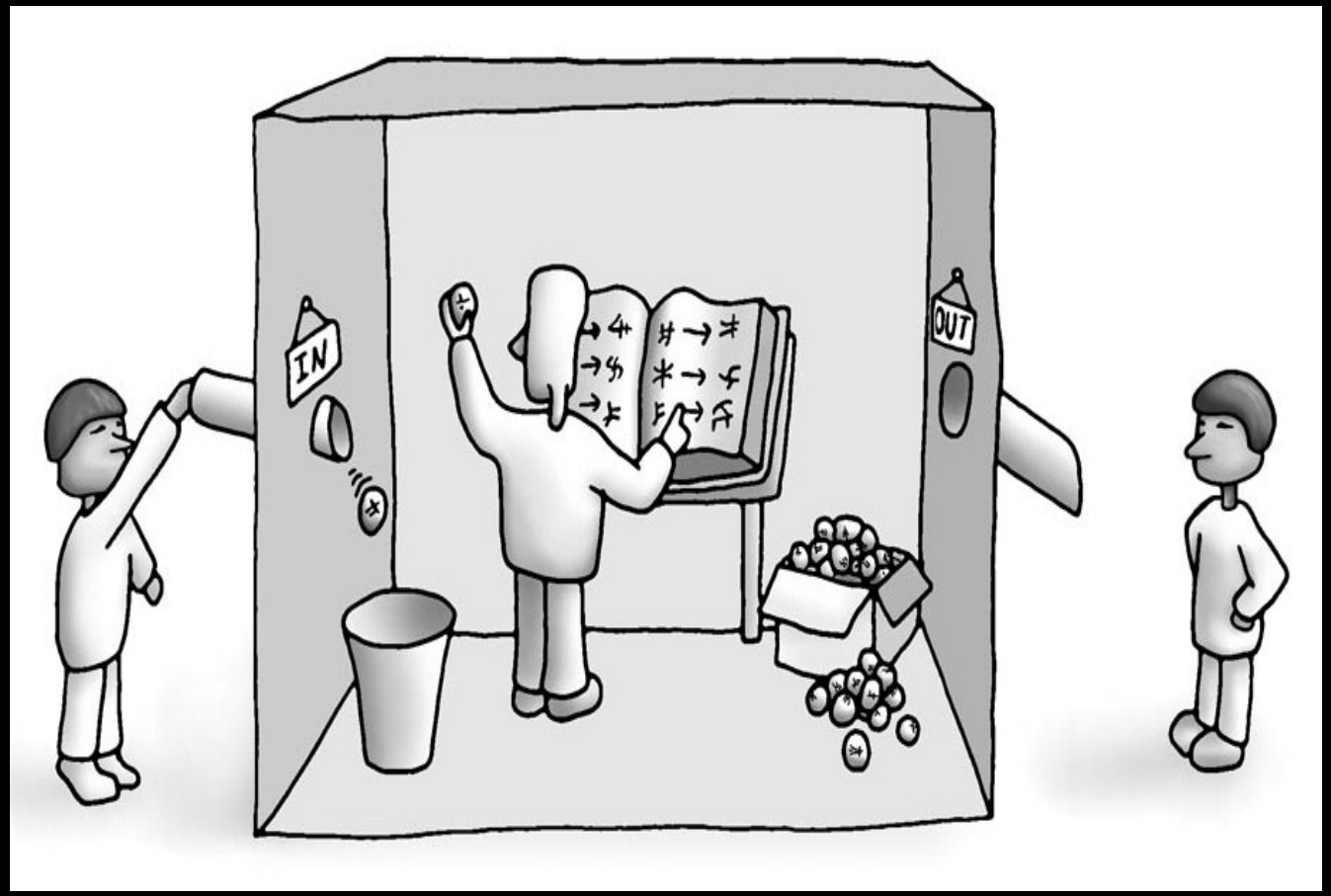
- Tests for AI are behavioristic:
  - they assume intelligent output implies an intelligent being

- But even a "dumb" phone can output intelligent speech
  - A system may be a conduit of intelligence without being a source of intelligence

Weaknesses of the Turing Test

- Passing the test is not necessary for intelligence:
  - An intelligent being might not be able to communicate e.g. someone with a speech/motor disorder, catalepsy
- Passing the test is not sufficient for intelligence:
  - A radio broadcast of Einstein
  - Searle's "Chinese room" argument

JOHN SEARLE'S CHINESE ROOM



WHAT'S MISSING FROM CHATGPT?

- Although it sometimes "hallucinates" nonsense, ChatGPT often generates meaningful text
- But its algorithms are purely syntactic
- It treats text as a series of tokens and predicts probable completions based on frequency
- It does not know what any of the tokens mean
- ChatGPT, like other AI, is not an intelligent being

Characteristics of an Intelligent Being

- 1. Subjectivity
  - There is a subject that has its own thoughts
- 2. Intentionality
  - The ability to think about other things (real or not)
- 3. Teleology
  - The possession of personal goals
- 4. Rationality
  - The ability to reason

1. Subjectivity

- An intelligent being has its own thoughts (it does not just contain information).
- Thoughts are intrinsically subjective:
  - they cannot be ownerless
  - they belong to a unified, enduring mental subject
- A subject literally has thoughts and experiences
- Subjects are self-aware
  - they know what it is like to be themselves

2. Intentionality

- An intelligent being can think **about** objects (not simply contain tokens which we interpret as about objects).
- Thoughts are by nature of or about something beyond themselves (intentionality), e.g.
  - The Taj Mahal is in India
  - 2 is the only even prime number
- Intentionality is required for understanding
  - It means we know what we are thinking about

WE ARE NOT TRAPPED INSIDE CODE

- Due to their lack of self-consciousness and intentionality, machines are trapped inside their code and cannot become aware of their condition.

"Man's greatness comes from knowing he is wretched:
a tree [or ChatGPT!] does not know it is wretched….
It is the wretchedness of a great lord, of a dispossessed king."

--Blaise Pascal, *Pensées*, #114, #116.

WE CAN SEE OUR NEED FOR CHRIST

- Man discerns the "infinite abyss" in himself which can only be filled by an "infinite and immutable object…by God himself."

  --Blaise Pascal, *Pensées*, #148.

- Only God can bridge the gap between God and man, and He has in Jesus Christ

3. Teleology

- An intelligent being has its **own** goals
  (these are not simply the goals of the designers).

- Subjectivity + Intentionality support personal goals

- Subjects can think about a state of affairs as desirable
  - Joe wants a vacation in Budapest
  - Angus desires another double Expresso

4. Rationality

- An intelligent being can use its **own** reason to solve problems (it is not simply a system engineered to follow rules)

- Subjectivity + Intentionality + Goals
  support Rationality

- A subject that can think about the world and has its own goals can use its thinking to achieve those goals

Why AGI systems are not intelligent beings

- 1. The simplicity of mental subjects
- 2. Intentionality is non-physical
- 3. Teleology is absent in AGI systems
- 4. Mirroring reason is not the same as reasoning

1. The Simplicity of Mental Subjects

- AGI systems are physical systems
- The "hard problem"of consciousness:
  - Physical systems can be completely described in impersonal terms
  - This does not imply or predict that the system is conscious
  - There is nothing it is like to be an AGI

A Lack of Unity

- Subjectivity requires there is one (unified) subject
- An AGI is a physical system
- Physical systems consist of **separable** parts
  - The parts remain the same even if removed from the system
  - They can be transferred from one system to another
- Thoughts are **inseparable** from their thinkers
  - Joe's thought that Brainerd is beautiful could not exist outside his mind
  - It cannot be transferred to someone else

Leibniz's Mill

- "Suppose that there be a machine, the structure of which produces thinking, feeling, and perceiving; imagine this machine enlarged…so that you could enter it as if it were a mill. This being supposed, you might visit its inside; but what would you observe there? Nothing but parts which push and move each other, and never anything that could explain perception. This explanation must therefore be sought in the simple substance, not in the composite, that is, in the machine."

--Gottfried Leibniz, *The Monadology*, Prop. 17.

The Mind is Simple

- Billions of people could emulate all of the connections between neurons in the brain.
- But there would not be a new consciousness of the crowd, in addition to each person's consciousness.
- --David Barnett, "You are simple," in *The Waning of Materialism*, eds. Robert C. Koons and George Bealer, (New York: Oxford University Press, 2010), 161-174.

2. Intentionality is Non-physical

- AGI systems are physical systems
- Their states are limited to those that can be caused by the systems' physical environment
- But we can think of:
  - Future states of affairs (a vacation)
  - Possible, but non-actual, states of affairs (peace in the Middle East)
  - Non-existent objects (English cuisine)

Why does this Matter?

- Intentionality allows us to think of alternative possibilities and future goals
- It enables us to create new ideas, imaginary worlds, and theories, algorithms and technologies in science

3. Teleology is Absent in AGI Systems

- AGI systems seem to have goals (they "solve problems," "learn languages" etc.)

- But they do not.

  - Without subjectivity and intentionality a system cannot have its own goals.

  - It does not literally desire to win a game of chess, or learn Swahili.

4. Mirroring Reason is not the same as Reasoning

- AGI systems are engineered to follow rational principles.
  - Like all computer systems, they are governed by an arithmetic-logic unit (ALU)
- But mirroring reason is not the same as reasoning

What Does True Reasoning Require?

- To truly reason there must be:
  - A) A unified subject at a time
  - B) A persistent subject over time
  - C) Intentionality
  - D) Teleology

An Example

- At time t1, Jack thinks:
  - A = B
  - B = C
- At time t2, Jack concludes:
  - A = C

A) A Unified Subject at a Time

- Jack must be one subject that thinks of both premises
  - Jack must think **both** that A = B **and** that B = C
  - Otherwise it is like Sarah thinking A = B and Brian thinking that B = C
  - Neither has a reason to conclude that A = C

AGI Systems have no Unified Subject at a Time

- AGI systems are physical aggregates of separable parts
- Information is distributed (e.g. in RAM or on auxiliary storage)
- There is no single subject that has thoughts at a time

B) A Persistent Subject over Time

- To reason to the conclusion that A = C, Jack must persist as the same subject at t1 (when he considers the premises) until t2 (when he draws the conclusion)
- Otherwise it is like Sarah believing that A=B and B=C and Brian believing A = C
  - Sarah does not draw the conclusion
  - Brian gets the right answer but does not reason to it!

AGI Systems are in Flux

- Hardware: constant electrical change in the switches
- Software: the system moves from one instruction to another
- There is no subject that endures over time, just a series of states

C) Intentionality

- Reasoning requires intentionality:
- Jack can logically infer that A = C only if he sees that it follows from the content of his thoughts that A = B and B = C
- To reason logically, you must see that a conclusion follows from what you believe

AGI Systems do not Exhibit Intentionality

- An AGI system might be in state B_1 containing the information that A = B and B = C
- It might transition to state B_2 containing the information that A = C
- But without intentionality, the system does not:
  - understand that A = B and B = C
  - see that A = C follows from what it believes
  - rationally draw the conclusion that A = C

D) Teleology

- Reasoning is inherently goal-directed
- We consider premises in order to draw a conclusion
- We consider a problem in order to develop a solution

AGI Systems do not Exhibit Teleology

- AGI systems are not subjects with their own goals, so they do not have the goal of drawing conclusions
- AGI systems can find any number of solutions to logic, math, scientific, textual, and image problems
- But they are not reasoning to those conclusions because they have no goal of finding them

4. WHAT IS SPECIAL ABOUT HUMAN BEINGS?

"When I look at your heavens, the work of your fingers,
the moon and the stars, which you have set in place,
what is man that you are mindful of him,
and the son of man that you care for him?
Yet you have made him a little lower than the heavenly beings
and crowned him with glory and honor."
(Psalm 8: 3-5)

SIGNS OF THE IMAGE OF GOD

- Man lost original righteousness in the fall
- Yet the image of God has continuing significance
  - A) Special dignity
  - B) Special work
  - C) Moral responsibility
  - D) Creativity
  - E) Relationality
  - F) Redeemability

A) SPECIAL DIGNITY

- "Whoever sheds the blood of man,
  by man shall his blood be shed,
  for God made man in his own image." (Genesis 9:6)

- "With [the tongue] we bless our Lord and Father, and with it we curse people who are made in the likeness of God. From the same mouth come blessing and cursing. My brothers, these things ought not to be so." (James 3:9-10)

B) SPECIAL WORK

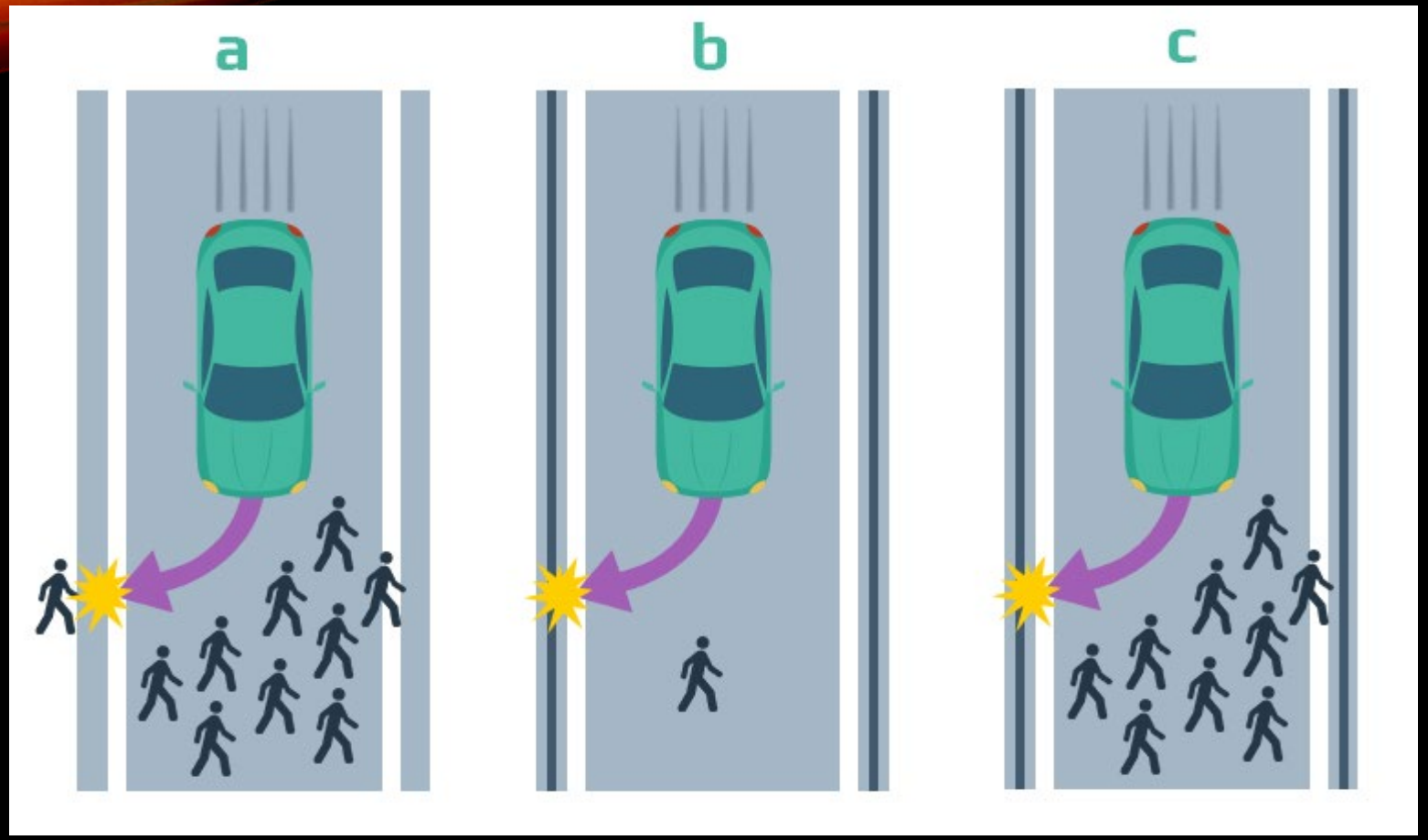
- "Then God said, 'Let us make man in our image, after our likeness. And let them have dominion over the fish of the sea and over the birds of the heavens and over the livestock and over all the earth and over every creeping thing that creeps on the earth.'" (Gen. 1: 26)

- "You have given him dominion over the works of your hands; you have put all things under his feet" (Psalm 8:6)

- God equips those he calls: we possess the gifts of moral responsibility, reason, and creativity we need to be stewards

C) MORAL RESPONSIBILITY

- We understand the moral demand
- We have free will (in temporal matters)
- We are not simply determined by our algorithms and "moral training" (like the AI in autonomous vehicles)



a. A car chooses to save many pedestrians over one.

b. A car potentially kills its passengers but saves a pedestrian.

c. A car saves a lot of pedestrians by hitting the wall and potentially killing its passengers.

D) CREATIVITY

- Sub-creation through access to universal concepts:
- “When we can take green from grass, blue from heaven, and red from blood, we have already an enchanter’s power . . . We may put a deadly green on a man’s face and produce a horror; we may make the rare and terrible blue moon to shine; or we may cause woods to spring with silver leaves and rams to wear fleeces of gold, and put hot fire into the belly of the cold worm. But in such ‘fantasy’, as it is called, new form is made . . . Man becomes a sub-creator.”

  — J. R. R. Tolkien, “On Fairy Stories”

E) RELATIONALITY

- God is love in His own nature (1 John 4:8)
- We are made in God's triune image for loving relationship with Him and one another.
- "Then the LORD God said, 'It is not good that the man should be alone; I will make him a helper fit for him.'" (Genesis 2:18)
- "See what kind of love the Father has given to us, that we should be called children of God; and so we are." (1 John 3:1)

F) REDEEMABILITY

"But when the fullness of time had come, God sent forth his Son, born of woman, born under the law, to redeem those who were under the law, so that we might receive adoption as sons." (Galatians 4:4-5)