

Outline

- The Turing Test
- Historical conflation of intelligence and consciousness
- Recent disentanglement of intelligence and consciousness
- What is consciousness?
- Robot rights

The Turing Test

- Alan Turing, “Intelligent Machinery,” 1947:

“Now get three men as subjects for the experiment. A, B and C. A and C are to be rather poor chess players, B is the operator who works the paper machine... a game is played between C and either A or the paper machine.

C may find it quite difficult to tell which he is playing...

These questions replace our original, ‘Can machines think?’”

Deep Blue vs. Gary Kasparov

- 1996: World chess champion Gary Kasparov beats Deep Blue, 4 games-2 in standard tournament play
- 1997: Deep Blue beats Kasparov, 3.5-2.5 in standard tournament play



By James the photographer - CC BY 2.0, By Copyright 2007, S.M.S.I., Inc. - Owen Williams, <https://commons.wikimedia.org/w/index.php?curid> The Kasparov Agency.

The Turing Test, Revised (1950)

“The interrogator is allowed to put questions to A and B thus:

- Q :Please write me a sonnet on the subject of the Forth Bridge.
- A :Count me out on this one. I never could write poetry.
- Q :Add 34957 to 70764
- A :(Pause about 30 seconds and then give as answer) 105621.
- Q :Do you play chess?
- A :Yes.
- Q :I have K at my K1, and no other pieces. You have only K at K6 and R at R1. It is your move. What do you play?
- A :(After a pause of 15 seconds) R-R8 mate.”

ChatGPT



Please write me a sonnet on the subject of the Forth Bridge.



Of steel and strength and daring feat of man,
A structure spans the Firth of Forth below,
The Forth Bridge, a marvel of its span,
Its beauty and its power on full show.

The cantilevers reach out to the sky,
With lattice girders braced and bolted tight,
A symbol of the human will to try,
To build beyond what was thought possible and right.

Through storm and tide, the bridge stands firm and true,
A testament to engineering might,
To those who built it and the ones who knew,
That it would stand the test of time and light.

The Forth Bridge, a wonder to behold,
A triumph of the brave and bold.

The AI Effect

It's part of the history of the field of artificial intelligence that every time somebody figured out how to make a computer do something—play good checkers, solve simple but relatively informal problems—there was a chorus of critics to say, 'that's not thinking.'

- Pamela McCorduck, *Machines Who Think*, 2004

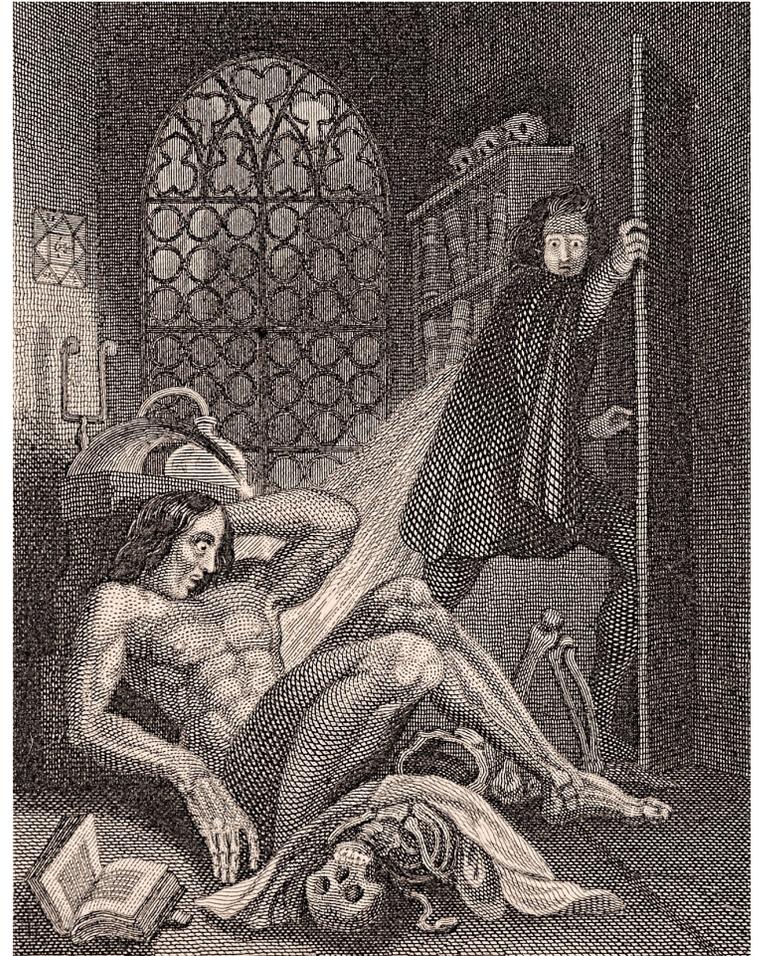
Outline

- The Turing Test
- Historical conflation of intelligence and consciousness
- Recent disentanglement of intelligence and consciousness
- What is consciousness?
- Robot rights

Frankenstein (Mary Shelley, 1818)

- An inventor creates an artificial human.
- The artificial human is huge, and strong, and ugly, and therefore feared, but he curbs his anger because:

“Instead of threatening, I intend to reason with you. Oh! My creator, make me happy; let me feel gratitude towards you for one benefit! Let me see that I excite the sympathy of some existing thing; do not deny me my request!”



By Theodore Von Holst (1810-1844)

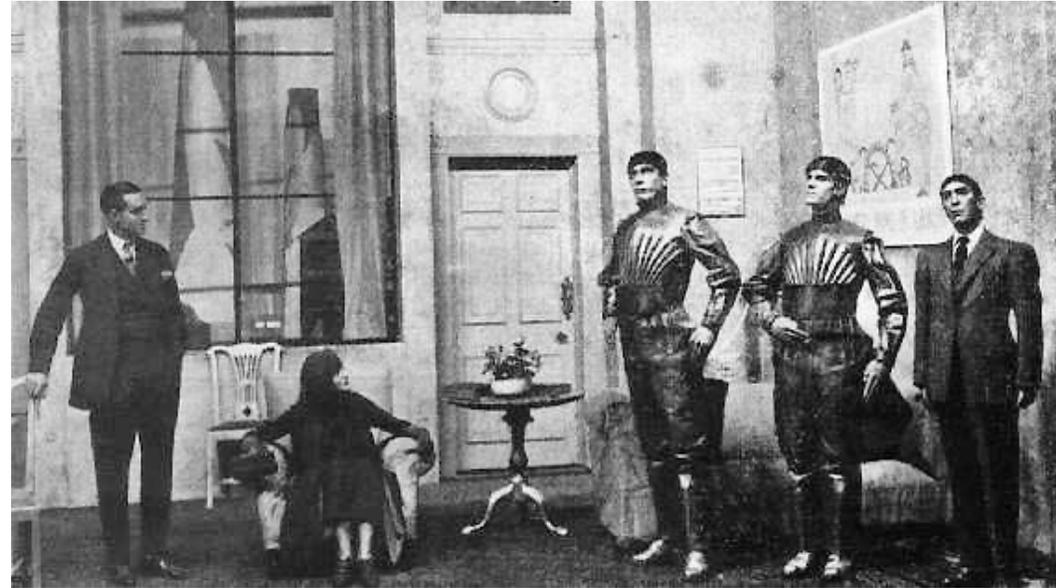
<https://commons.wikimedia.org/w/index.php?curid=4940182>

Rossum's Universal Robots (Karel Capek, 1923)

“Robot” is a Czech word meaning “worker.” When the play opens, the factory had turned out already, following a secret formula, millions of manufactured workmen without souls, desires or feelings. They are high-powered laborers, good for nothing but work.

Dr. Gall, the head of the physiological and experimental departments, has secretly changed the formula, and while he has partially humanized only a few hundreds, there are enough to make ringleaders, and a world revolt of robots is under way.”

- From R.U.R. “Story of the Play”



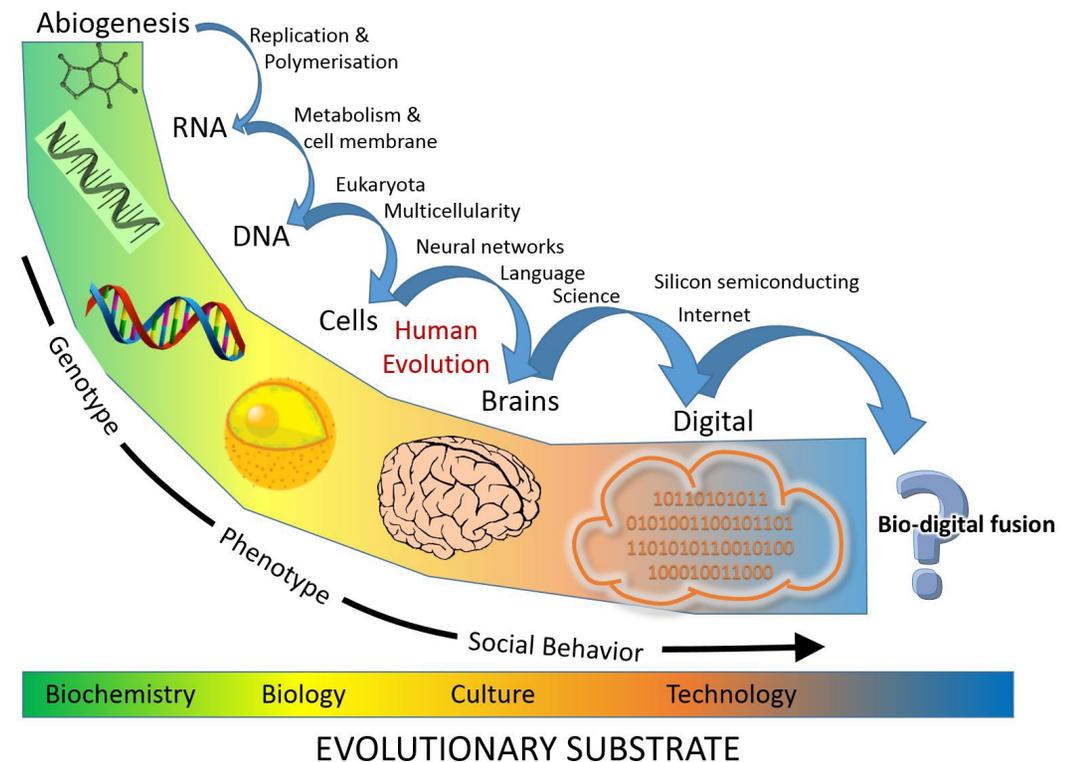
<https://commons.wikimedia.org/w/index.php?curid=10522903>

Technological Singularity (Vernor Vinge, 1993)

- “Computers that are ***awake*** and ***superhumanly intelligent*** may be developed. (To date, there has been much controversy as to whether we can create human equivalence in a machine. But if the answer is yes, then there is little doubt that more intelligent beings can be constructed shortly thereafter.)
- Large computer networks (and their associated users) may ***wake up*** as ***superhumanly intelligent*** entities.

I’ll be surprised if this event happens before 2005 or after 2030.”

- Vernor Vinge, 1993



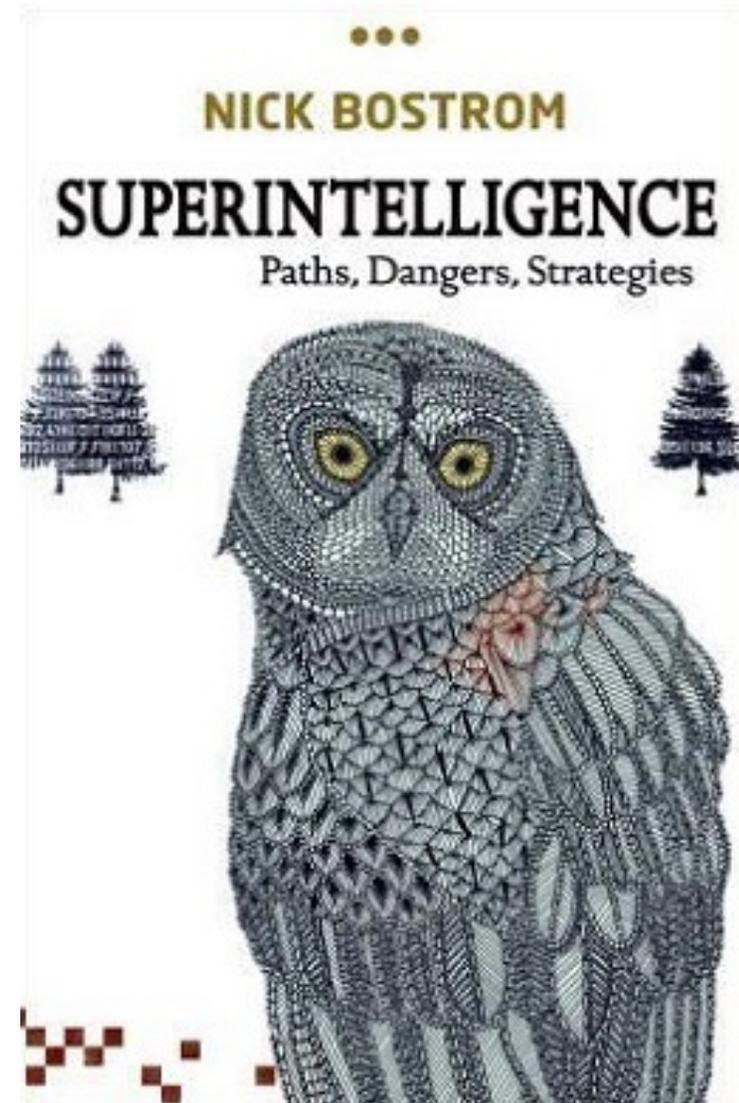
based on: Gillings, M. R., Hilbert, M., & Kemp, D. J. (2016). Information in the Biosphere: Biological and Digital Worlds. *Trends in Ecology & Evolution*, 31(3), 180–189. <http://escholarship.org/uc/item/38f4b791>

By Myworkforwiki - Own work, CC BY-SA 4.0,
<https://commons.wikimedia.org/w/index.php?curid=48996558>

Superintelligence (Nick Bostrom, 2014)

The “Unfinished Fable of the Sparrows:”

- Sparrows: “Hey! Let’s find an owl chick and raise it! It will be so nice to have an owl to help us!”
- Scronkinkle the sparrow: “Shouldn’t we, first, figure out how we will keep it tame, so it doesn’t eat all of us?”
- Sparrows: “Oh, there will be plenty of time to figure that out later, after we have started raising it.”



Fair use, <https://en.wikipedia.org/w/index.php?curid=68602452>

Outline

- The Turing Test
- Historical conflation of intelligence and consciousness
- Recent disentanglement of intelligence and consciousness
- What is consciousness?
- Robot rights

Bostrom's other two types of superintelligences...

A superintelligence does not have to be conscious to be dangerous. Two examples:

- A program with a simple goal (make as many paper clips as possible) might be intelligent enough to destroy all human life in doing so, even if it is not conscious.
- Large corporations might have AI resources that make them unstoppable by any other force in society.



CC-SA 4.0,

[https://commons.wikimedia.org/wiki/File:B%C3%BCroklammern -- 2021 -- 6481.jpg](https://commons.wikimedia.org/wiki/File:B%C3%BCroklammern_-_2021_-_6481.jpg)

Vinge's other two types of singularities...

- “Computer/human interfaces may become so intimate that users may reasonably be considered superhumanly intelligent.
 - Biological science may provide means to improve natural human intellect.”
- Vinge, 1993

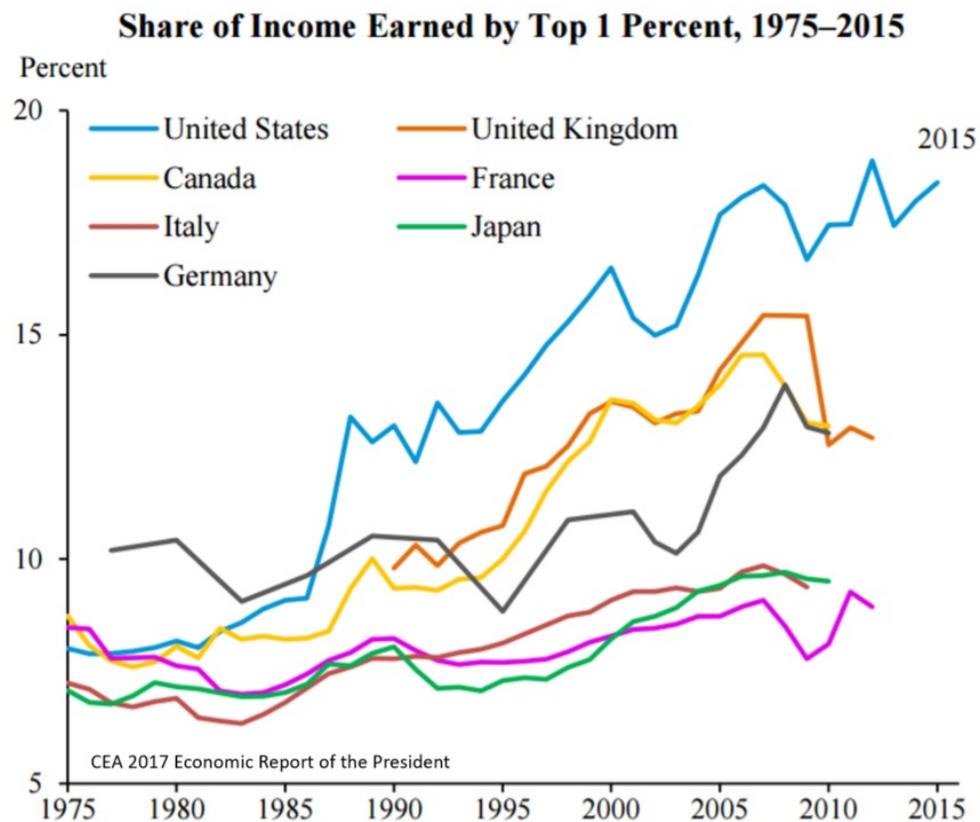


By Wikideas1 - Own work, CC0,

<https://commons.wikimedia.org/w/index.php?curid=122820749>

Superintelligence has already been achieved?

- According to the Turing test, AI is now intelligent.
- According to Vinge's 3rd criterion ("users may reasonably be considered superhumanly intelligent"), one could argue that the singularity may have already occurred.
- According to Bostrom's corporation model, one could argue that superintelligence is already a threat to human happiness (see: income disparity).
- None of these things mean that we have created conscious computers!



Source: World Wealth and Income Database.

By U.S. Council of Economic Advisers - 2017 Economic Report of the President - 2017 Economic Report of the President, Public Domain, <https://commons.wikimedia.org/w/index.php?curid=54550890>

ChatGPT's opinion



Does an AI need to be conscious in order to be considered intelligent?



No, an AI does not need to be conscious in order to be considered intelligent. Intelligence can be defined as the ability to acquire and apply knowledge and skills, to reason and solve problems, and to adapt to new situations. These are all functions that can be programmed into an AI system without it having to be conscious or self-aware.

Consciousness, on the other hand, refers to the subjective experience of being aware of one's surroundings, thoughts, and emotions. It is a much more complex phenomenon that involves the integration of sensory input, memory, and other cognitive processes.

While some researchers believe that consciousness is a necessary component of true intelligence, others argue that it is not essential. It is possible to create highly intelligent AI systems without them being conscious, and conversely, it is possible for conscious beings to lack intelligence.

Outline

- The Turing Test
- Historical conflation of intelligence and consciousness
- Recent disentanglement of intelligence and consciousness
- What is consciousness?
- Robot rights

What is consciousness?

Consciousness—The having of perceptions, thoughts, and feelings; awareness. The term is impossible to define ... it is impossible to specify what it is, what it does, or why it has evolved. Nothing worth reading has been written on it.

- Stuart Sutherland, MacMillan Dictionary of Psychology, 1989

Consciousness: Psychological definition

For now, we will consider people to be conscious of an event if (1) they can say immediately afterwards that they were conscious of it and (2) we can independently verify the accuracy of their report. If people tell us that they experience a banana when we present them with a banana but not with an apple, we are satisfied to suppose that they are indeed conscious of the banana.

- Bernard Baars, *A Cognitive Theory of Consciousness*, 1988

Consciousness: Medical Definition, The Glasgow Coma Scale

Eye Opening Response

- Open with blinking at baseline (4 points), to verbal stimuli, command, speech (3 points), to pain only (2 points), no response (1 point)

Verbal Response

- Oriented (5 points), Confused conversation, but able to answer questions (4 points), Inappropriate words (3 points), Incomprehensible speech (2 points), No response (1 point)

Motor Response

- Obeys commands for movement (6 points), Purposeful movement to painful stimulus (5 points), Withdraws in response to pain (4 points) Flexion response to pain (3 points), Extension response to pain (2 points), No response (1 point)

Consciousness: Self-awareness

Gordon Gallup (*Science*, 1970) proposed the “mirror test:”

- Put a spot of coloring on an animal’s face.
- Put the animal in front of a mirror.
- Observe if the animal understands that the face in the mirror is its own face.
- By this criterion, conscious creatures include humans over 18 months of age, great apes, bottlenose dolphins, orcas, pigeons, magpies, and elephants.

The Cambridge Declaration on Consciousness

“Convergent evidence indicates that non-human animals have the neuroanatomical, neurochemical, and neurophysiological substrates of conscious states along with the capacity to exhibit intentional behaviors. Consequently, the weight of evidence indicates that humans are not unique in possessing the neurological substrates that generate consciousness. Nonhuman animals, including all mammals and birds, and many other creatures, including octopuses, also possess these neurological substrates.”

- Jaak Panksepp, Diana Reiss, David Edelman, Bruno Van Swinderen, Philip Low and Christof Koch, 2012

Outline

- The Turing Test
- Historical conflation of intelligence and consciousness
- Recent disentanglement of intelligence and consciousness
- What is consciousness?
- **Robot rights**

Superintelligence: Bostrom's proposed solution (2014)

- We need to understand what makes humans moral.
- We need to find a moral code that is binding on all intelligences, not just on human intelligences.
- This cannot be something that we build into the AI (like Asimov's "three laws of robotics"), because the AI will just figure out how to get around it. It needs to be an imperative that is naturally binding on all intelligences, regardless of whether they believe it or not.

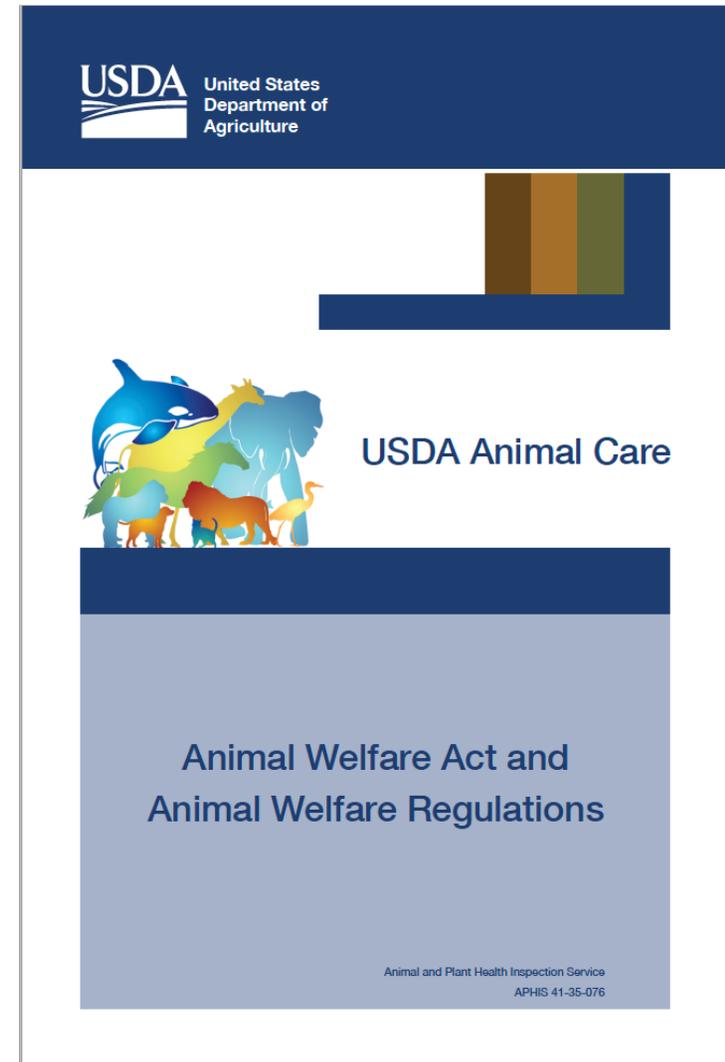
Where do human rights come from?

- Aristotle (capabilities \Rightarrow rights): those with the ability to reason about the future have, therefore, the right to exercise their ability
- Locke (self-defense \Rightarrow rights): if rights are ignored by a government, humans can overthrow the government
- Finnis (good of society \Rightarrow rights): the doctrine of rights helps us to construct a better society

Do animals have rights?

The capabilities argument: creatures that can feel pain should not be subjected to avoidable pain

The good of society: inflicting avoidable pain on an animal reduces one's ability to feel empathy, which may be harmful to society



Do robots have rights?

Capabilities:

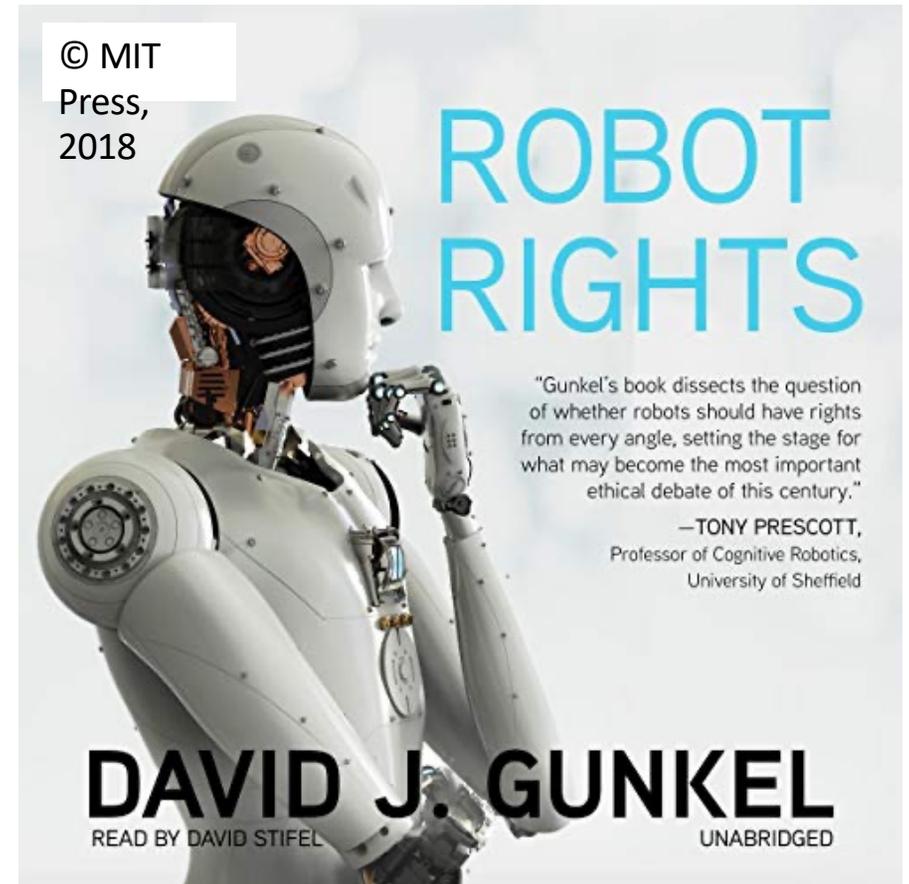
- Robots perceive, interpret, plan, act, and communicate in natural language
- Robots do not, as far as we know, have consciousness

Self-defense:

- Robots cannot currently defend themselves

Good of society:

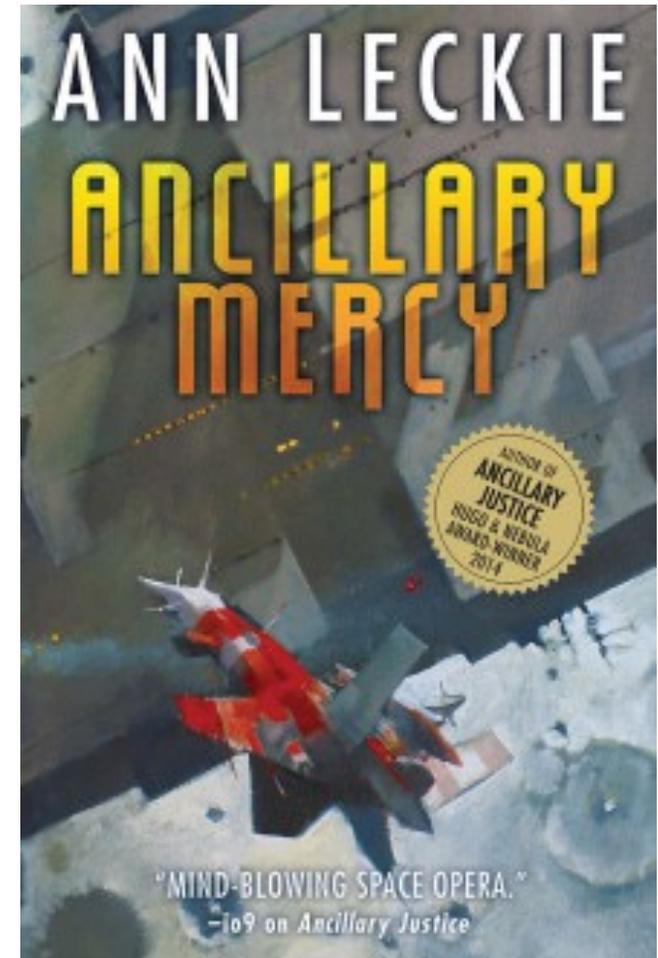
- Giving rights to robots would not currently benefit society



When will robots have rights?

The novel *Ancillary Mercy* by Ann Leckie proposes that robot rights require that robots satisfy all three criteria for human rights:

- **Capabilities**: Robots must be self-aware, rational, capable of emotion, and intelligent
- **Self-defense**: Robots must be able to defend their own rights, in either a physical fight or a verbal argument
- **Good of society**: It must be the case that society is better off with robot rights than without



Outline

- The Turing Test
- Historical conflation of intelligence and consciousness
- Recent disentanglement of intelligence and consciousness
- What is consciousness?
- Robot rights