

**Can humans not-think?**

*or*

**Humanity Lost? Reverse Turing tests to  
examine the human-side of the Turing  
line**

*or just*

**We Frogs**

Brett Frischmann  
Cardozo Law School

# Plan

- Introduce project
- Discuss essay, *Can Humans Not-think?*

# Themes

- When does technology (automated systems) replace or diminish our humanity?
- Can we detect when this happens? how will we?
- What makes us human (vs machine)?

*We Frogs need to be able to detect changes in water temperature before we can contemplate whether it is getting too hot, much less who is controlling the heat.*

# Four essays (maybe book parts)

## 1. Human focused Turing type tests

- You'll see – this is what I'm going to discuss!

## 2. Series of thought experiments

- Adapted Black and White Mary thought experiment
- Temperature regulation
- Brains in vats and s(t)imulated happiness
- Fully extended minds

## 3. Normative analysis

- Happiness vs. capabilities; thingification of people; distributive justice and (hidden technological) power

## 4. Applications

- Technological reconstruction of our shared environments and us: Internet of Things; Big Data enabled automated systems of, about, and around us; nudging ...

# ***Can Humans Not-think?***

- 1. The human side of the Turing line**
2. Turing Test: A brief overview and literature review
- 3. Intelligence: Thinking and not-thinking**
  - A. (ir)rationality test
  - B. common sense test
4. Technologically Extended Minds
5. Autonomy and Choice
  - A. difficulties in measuring degrees of freedom and identifying exercise of choice
  - B. prediction / predictability of agents' behavior: A proxy test
6. The Environment Game
- 7. Generalizing the inquiry: What are we testing on the human side of the Turing line?**

# Turing Line

**Humans**

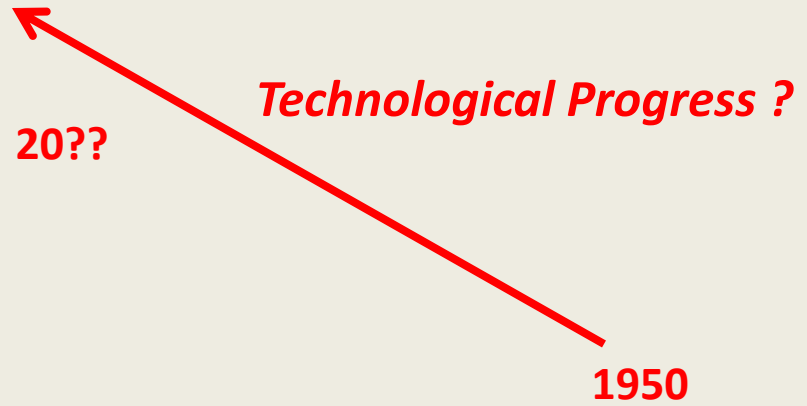
**Machines**



# Turing Line

Humans

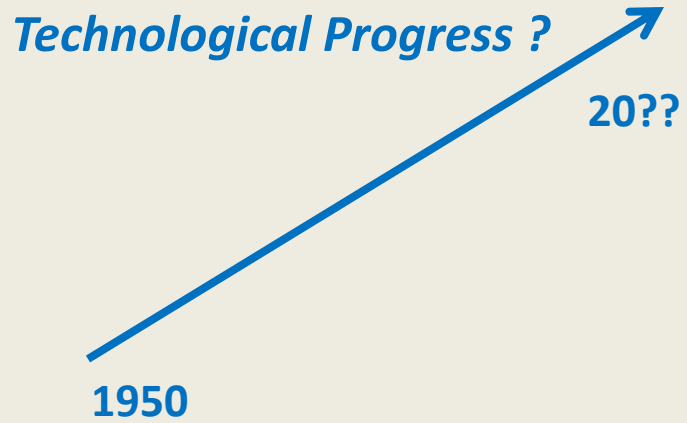
Machines



# Turing Line

Humans

Machines



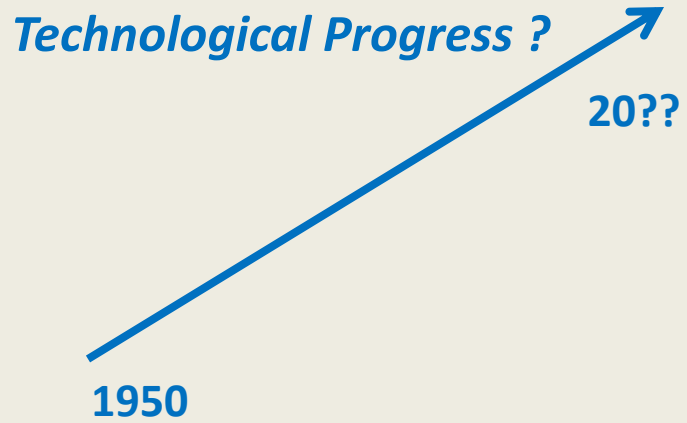


- Under what conditions and/or when are humans indistinguishable from machines?
- Can humans be programmed or constructed to be indistinguishable from machines?
- Can environments dehumanize?
- How and/or when are human beings constructed (via technology, social context, and the environment within which we live and through which our preferences and beliefs are formed) to be indistinguishable from machines?

# Turing Line

Humans

Machines



# Turing Line

Humans

Machines

*Technological Progress ?*

20??

1950

**Call Center Employees**

*The Computerized Voice that Wasn't, NY Times*



# *Set-up*

- a) Play Turing test with same rules; no difference at all; examine situations where humans have been mistaken to be machines.
- b) Play Turing test as usual, except humans are told that they should try to deceive the observer; the observer doesn't know about this additional instruction.
- c) Same, but the machine programmers also are informed that humans will play strategically.
- d) Same as b, but the observer knows.
- e) Same as c, but the observer knows.
  
- f) Same as any of the above, except the machine agents are *simple machines*, or at least, are not deceptive or otherwise strategically playing the game.
- g) Same, except the observer is a computer.

# ***Intelligence-related characteristics that can be tested***

- Reason
- Rationality / Irrationality
- Willpower
- Emotion
- Phenomenological experience, such as the capacity to feel and understand the feeling of hot/cold, hunger, etc., or to see and understand the color red
- Language / capacity to construct new language or social meaning
- Common sense
- Planning for others / for the happiness of others
- Language with which to plan for future / others
- System 1 & system 2: thinking fast and slow
- ...

# (ir)rationality

- Set-up: (a) + (f) + (g): conventional Turing test w/ simple machine agents and machine observer
- Observer machine (w/ machine learning and battery of RC tests) → *perfect rationality detector*
- Thought experiment: *Nudging environment*
  - Human routinely pass test
  - What inferences?
  - Extensions of thought experiment ...

- Suppose the nudging government did not limit itself to workplace environments. Suppose the government systematically constructed nudging environments in as many places and social contexts as possible.
  - *If this seems implausible and too abstract, consider modern surveillance systems.*
- Suppose the government is not involved and the nudging environments are voluntarily constructed by private entities, such as firms or collections of people employing shared technologies. Does it matter *who* is doing the nudging?
  - *Once more, if this seems implausible and too abstract, consider modern surveillance systems.*

# common sense

- “[We] can understand common sense itself as the base of knowledge about common-sense reality that allows each of us to survive and thrive during our everyday lives. Common *beliefs* about the common-sense world are the most prominent components of this knowledge base. ... common sense also includes the widespread *abilities* that allow us to act successfully in the common-sense world.”
  - (Erion at 33).
- It entails core knowledge and skills that are shared and “used by all of us (even skeptical philosophers) during our everyday lives.” Language is critical to common sense both as knowledge and as skill. That is, competence in using language is a “subset of common sense”
  - (Erion at 36).



# Common sense test

- High bar for machines, but what about for humans?
- *What would it mean if a human were indistinguishable from a machine based on the human's performance in a common sense test?*
  - Common sense depends upon a shared core knowledge base, language, and social interactions sufficient to generate common understandings and beliefs.
  - Thought experiment: Alice lost

# Common sense and technology

- common sense may dictate resort to technology
  - But technologies are not neutral or equivalent
  - Some may substitute for common sense or disable access to key inputs (core knowledge base, language, and social interactions)
  - Reliance on different communities

1. Humans face common problems in everyday life (“everyday life problems”).
2. Humans develop and rely on common sense solutions to everyday life problems.
  - Developing common sense solutions necessarily depends on a shared core knowledge base, language, and social interactions sufficient to generate common understandings and beliefs.
  - Developing common sense solutions [*necessarily? often? usually?*] depends on experimentation and social innovation.
3. Humans develop technology to solve problems.
  - Developing technology to solve a problem depends on knowledge, experimentation, and innovation, but not necessarily on a shared core knowledge base, language, and social interactions sufficient to generate common understandings and beliefs.
4. Some technology solves everyday life problems.
5. If technology solves an everyday life problem (more efficiently than existing common sense solutions) then humans will not (are less likely to) develop common sense solutions to that problem.
6. If technology solves *all* everyday life problems, then humans will lack common sense (or a subset of common sense that concerns problem solving).
7. Humans without common sense are indistinguishable from machines, at least in one (important) respect.

# Conclusion

- What are we testing on the human side of the Turing line? What can we reasonably infer when test is passed?

*We Frogs need to be able to detect changes in water temperature before we can contemplate whether it is getting too hot, much less who is controlling the heat.*