



Center for
**LifeLong
Learning
& Design**

University of Colorado at Boulder

**Wisdom is not the product of schooling
but the lifelong attempt to acquire it.
- Albert Einstein**

Bill Joy: "Why the Future Does Not Need Us"

**Gerhard Fischer and Hal Eden
Spring Semester 2005**

February 2, 2005

Some of the Players: Bill Joy

- principal designer and programmer of the Berkeley version of Unix → Joy: “a personal ‘success disaster’ - so many people wanted it that I never finished my PhD”
- chief scientist and cofounder of Sun Microsystems Inc.
- primary figure in the development of
 - SPARC (**S**calable **P**rocessor **ARCH**itecture) microprocessors
 - Java programming language (with Jim Gosling): object-orientation, platform independence
 - JINI (an open architecture that enables developers to create network-centric services)
- in summary: definitely not a person who is hostile to technological developments

Some of the Players: Ray Kurzweil

- **an inventor:** Kurzweil was the principal developer of the first omni-font optical character recognition, the first print-to-speech reading machine for the blind, the first CCD flatbed scanner, the first text-to-speech synthesizer, the first music synthesizer capable of recreating the grand piano

- **Books:**
 - The Age of Intelligent Machines (1990)
 - The Age of Spiritual Machines: When Computers Exceed Human Intelligence (2000)
 - The Singularity Is Near: When Humans Transcend Biology — Kurzweil's belief about a coming technological singularity (to be released on August 18, 2005)

- In futurology, a **technological singularity** is a predicted point in the development of a civilization at which technological progress accelerates beyond the ability of present-day humans to fully comprehend or predict. *The Singularity* can more specifically refer to the advent of smarter-than-human intelligence, and the cascading technological progress assumed to follow

Some of the Players: John Searle

- a philosopher who studies consciousness
- **"strong AI"**
 - A belief of strong AI is that if a machine were to pass a Turing test, then it can be regarded as "thinking" in the same sense as human thought
 - Intentionality lies at the heart of Searle's Chinese Room argument against artificial intelligence which proposes that since minds have intentionality, but computers do not, computers cannot be minds.
 - The core of Searle's argument is the distinction between syntax and semantics.
- **Chinese room thought experiment:** a person who understands no [Chinese](#) sits in a room into which written [Chinese characters](#) are passed. In the room there is also a book containing a complex set of rules (established ahead of time) to manipulate these characters, and pass other characters out of the room. This would be done on a *rote* basis, eg. "*When you see character X, write character Y*". The idea is that a [Chinese-speaking](#) interviewer would pass questions written in Chinese into the room, and the corresponding answers would come out of the room appearing from the outside as if there were a native Chinese speaker in the room.

“Why the Future Does Not Need Us”

- **major conclusion:** robotics, nanotech, and genetic engineering are emerging so quickly that—if we weren't careful—they could endanger the human species

- **Joy's self-reflection:**
 - “I may be working to create tools which will enable the construction of the technology that may replace our species”
 - “once an intelligent robot exists, it is only a small step to a robot species - to an intelligent robot that can make evolved copies of itself.”

Some of the Concepts

▪ Luddites

- The Luddites were a group of English workers in the early 1800s who protested (often by destroying machines) against the changes produced by the Industrial Revolution that they felt threatened their jobs.
- is Bill Joy a "neo-Luddite"? → the article is published in **“Wired!”**

▪ Moore's law

- an empirical observation: our rate of technological development and advances in the semiconductor industry doubles every 18 months.
- It is attributed to Gordon E. Moore (a co-founder of Intel)

▪ Pandora's Box

- Pandora's Box is the box owned by the mythological figure Pandora, which released all the sorrows of mankind into the world when opened; hope was the only thing left in the box.

▪ “Faustian bargain”

- “Faustian bargain” is one in which a person is willing to make extreme sacrifices for power or knowledge without considering the ultimate cost
- The Faustian Bargains of Nanotechnology, of Stem Cell Research, of

Faustian Bargain: the Atomic Bomb

- Robert Oppenheimer — Hitler might obtain nuclear weapons. Energized by this concern, he brought his strong intellect, passion for physics, and charismatic leadership skills to Los Alamos and led a rapid and successful effort by an incredible collection of great minds to quickly invent the bomb.
- As the physicist Freeman Dyson later said, "The reason that it was dropped was just that nobody had the courage or the foresight to say no."
- It's important to realize how shocked the physicists were in the aftermath of the bombing of Hiroshima, on August 6, 1945. They describe a series of waves of emotion: first, a sense of fulfillment that the bomb worked, then horror at all the people that had been killed, and then a convincing feeling that on no account should another bomb be dropped.

Faustian Bargains: Strategic Defense Initiative, Robotics and Genetic Engineering

- **Strategic Defense Initiative** = proposed by the Reagan administration, was an attempt to design such a shield against the threat of a nuclear attack from the Soviet Union
- Similar difficulties apply to the construction of shields against **robotics** and **genetic engineering**
- **claim:** “these technologies are too powerful to be shielded against in the time frame of interest; even if it were possible to implement defensive shields, the side effects of their development would be at least as dangerous as the technologies we are trying to protect against”

Dialogs, Discussions and Collective Wisdom: Challenges for Learning and Collaboration

- I believe that we all wish our course could be determined by our **collective values, ethics, and morals**. If we had gained more collective wisdom over the past few thousand years, then a dialogue to this end would be more practical, and the incredible powers we are about to unleash would not be nearly so troubling.
- One would think we might be driven to such a dialogue by our instinct for self-preservation. Individuals clearly have this desire, yet as **a species** our behavior seems to be not in our favor. In dealing with the nuclear threat, we often spoke dishonestly to ourselves and to each other, thereby greatly increasing the risks. Whether this was politically motivated, or because we chose not to think ahead, or because when faced with such grave threats we acted irrationally out of fear, I do not know, but it does not bode well.
- The new **Pandora's boxes** of genetics, nanotechnology, and robotics are almost open, yet we seem hardly to have noticed. Ideas can't be put back in a box; unlike uranium or plutonium, they don't need to be mined and refined, and they can be freely copied. Once they are out, they are out.

Bill Joy's Recommendations

- We must do more **thinking up front** if we are not to be similarly surprised and shocked by the consequences of our inventions.
- My continuing professional work is on improving the reliability of software. Software is a tool, and as a toolbuilder I must struggle with the uses to which the tools I make are put. I have always believed that **making software more reliable, given its many uses, will make the world a safer and better place**; if I were to come to believe the opposite, then I would be morally obligated to stop this work. I can now imagine such a day may come.
- My immediate hope is to **participate in a much larger discussion** of the issues raised here, with **people from many different backgrounds**, in settings **not predisposed to fear or favor technology** for its own sake.

How Good are our Predictions?

- **claim** by John Seely Brown and Paul Duguid in P. Denning “The Invisible Future”, p 126:
 - where real things occur (e.g.: WWW), it is rarely predicted, and
 - where it is predicted, it often fails to occur, e.g.: strong AI, Natural Language processing, paperless office, distance does not matter,

AI Questions in 1963 — and Before

- What is a computer? Is it just a “number factory”?
- Is it possible for computing machines to think?
- But doesn't a computer do exactly what it is told to do and no more?
- Is it true that a computer will be a chess champion because the computer is so fast that it can examine all possible moves and their consequences?
- What is a heuristic program?
- What are some unsolved problems of Artificial Intelligence?
- What are the limits of artificial intelligence?

Claims

- Making computers intelligent helps us to understand intelligence.
- Intelligent computers are more useful computers.
- Intelligent computers have their own will.
- Intelligent computers will make mistakes.
- Intelligent computers will / should replace human beings (e.g.: expert systems, ABS, flight computers)
- Intelligent computers will / should empower and augment human capabilities (IA = Intelligence Augmentation, Einstein: “My pencil is cleverer than I”!)

Different Views about AI

Person_A: My students have programmed computers to solve mathematical problems, play games, even to understand some English.

Person_B: I know those programs and I don't feel that they really think. They only do what they are programmed to do. People are individuals with free will. They have creative intellects.

Person_A: Newborn babies do only what they are programmed to do by their genes and environments. And they never do anything very creative until programmed by parents, school and the culture at large.

Person_B: That's not being programmed! At least the children learn it themselves, by walking around, seeing, touching, feeling, loving. That's not being programmed. That's *experiencing*.

Different Views about AI – Continued

Person_A: Well, if you like. But we are beginning to make robots which can see, touch manipulate, make mistakes and learn while doing all that. The gap is closing.

Person_B: Irrelevant. Even if computers were to show the same *outer* behavior as people, you can be sure that what is going on inside is totally different. As different as birds and airplanes. Birds and people feel. Airplanes and computers don't.

Person_A: Birds and Airplanes *are* different. But the same science of aerodynamics helps us to understand both of them. It's only after people learned to make and fly airplanes that they knew what to look for in trying to understand birds. Perhaps we must first build thinking machines before we know the right questions to ask about thinking and feeling.

Version 1 of a Dialogue with Eliza

P: Men are all alike.

C: In what way?

P = Person

C = Computer Program

P: They are always bugging us about something or other.

C: Can you think of a specific example?

P: Well, my boyfriend made me come here.

C: Your boyfriend made you come here?

P: He says, I am depressed much of the time.

C: I am sorry to hear that you are depressed.

P: It is true, I am unhappy.

C: Do you think coming here will help you not to be unhappy?

P: I need some help, that much seems to be certain.

C: What would it mean to you if you got some help?

P: Perhaps I could learn to get along with my mother.

C: Tell me more about your family!

GUS: A Travel Planning Assistant

Computer: When do you want to leave?

User: I must be in LA at 10:00am

Computer: Continental Flight #342 arrives at 9:30am?

User: What is the next flight?

Version 2 of a Dialogue with Eliza

P: My mumble bzzz grumble grumble.

C: That's very interesting.

P: Ugh, you izzle flerp me.

C: What makes you think that I izzle flerp you?

P: Gorch.

C: Does that have anything to do with the fact that your mumble bzzz grumble grumble?

Example: Principles of Human-Centered Aircraft Automation

- **Premise:** The pilot bears the ultimate responsibility for the safety of any flight operation.

- **Axiom:** The pilot must be in command

- **Corollaries:**
 - to command effectively, the pilot must be involved
 - to be involved, the pilot must be informed
 - the pilot must be able to monitor the automated systems
 - automated systems must therefore be predictable
 - the automated systems must also be able to monitor the pilot
 - each element in the system must have knowledge of the other's intent

Dimensions of “Human-Centered Automation”

