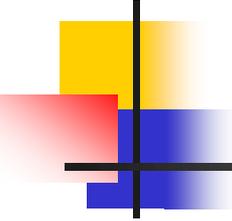# Weizenabum Address
CEPE 2019
May 29, 2019

The State of ICT Ethics:
# A Brief Look Back and a Look Ahead

Herman T. Tavani

Rivier University

Preliminary Remarks/Organization of Talk

- Before discussing the state of ICT ethics, I briefly comment on:

a) the origin of the Weizenbaum Award, and the influence that Joseph Weizenbaum's academic work had for ICT ethics;

b) some analogies involving Weizenbaum and (philosopher) Hans Jonas, whose work I return to in the final section of this talk.

- My talk is organized into three main parts:

➤ Part 1 could be viewed as historical/sociological (interspersed with a sprinkling of personal anecdotes).

➤ Part 2 is primarily "meta-computer-ethical" in nature.

➤ Part 3 is (more traditionally) philosophical.

- While the structure of this talk might seem a bit "strange," I hope that you will find it to be coherent, even if you are not convinced by any of the arguments I put forth.
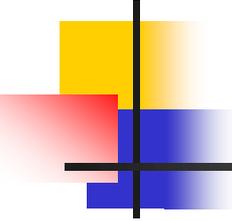
# Part 1: The Weizenbaum Award, Joseph Weizenbaum, and Hans Jonas

- 2019 marks the tenth anniversary of the (INSEIT-sponsored) Weizenbaum Award.

- So I wanted say a few words about the origin of this award, which was initially proposed by former INSEIT Director Elizabeth Buchanan (sometime in early 2008).

- As a past INSEIT president, I had the good fortune to work closely with Elizabeth on this and other INSEIT initiatives.

- We formed a committee and agreed to name the new award in honor of Joseph Weizenbaum (who passed away in April 2008).

- I then spoke by phone to the late Professor Weizenbaum's daughter, Sharon, who graciously gave permission to INSEIT to use her father's name; she also seemed pleased that INSEIT's mission closely mirrored the values espoused by her father.
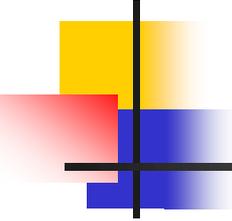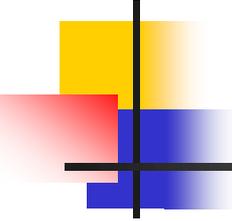
Weizenbaum Award (Background)

- I had the honor of presenting the first Weizenbaum Award to Terrell Bynum at CEPE 2009 (Ionian Academy, Corfu Greece).

- At that time, I could not have imagined that one day I too might be a recipient of that award.

- So, when informed this past January that I had been selected to receive the 2019 award, I immediately thought back to Bynum's reaction when I had notified him nearly 10 years earlier.

- Fortunately, I had saved my copy of Bynum's email response, which I quote directly:

  "It means a great deal to have my work acknowledged and honored by my peers – especially my colleagues in INSEIT, from whom I have learned so much, and with whom I have shared many meaningful projects and events!" (Bynum, 11/5/2008).
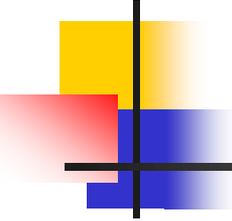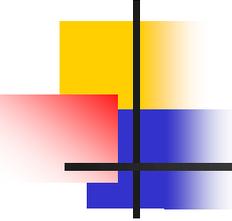
Background (Continued)

- In the first Weizenbaum address, titled "Philosophy and the Information Revolution," Bynum (2009) drew upon some analogies and comparisons in the works of two thinkers that he described as "philosophers of the Information Age":

1) Norbert Weiner (now widely regarded a classic 20th thinker in the field of cybernetics),

2) Luciano Floridi (an influential contemporary philosopher in information ethics).

- I also draw on some analogies/comparisons involving the works of two important 20th century thinkers: Joseph Weizenbaum and Hans Jonas.

- But unlike Bynum's address, the analogies that I briefly describe regarding the two thinkers I compare (i.e., Weizenbaum and Jonas) will not constitute the main objective of my talk.

- I simply note these comparisons at an early point in my talk, and then examine Jonas's ethical system in more detail in the final section.

1.2 Weizenbaum's Academic Background and his Influence on ICT Ethics

- Joseph Weizenbaum is mainly remembered today for his work as a computer science professor and AI pioneer at MIT.

- He is known for his work on several projects/programs/books:

- **ERMA** (Electronic Recording Machine, Accounting) project at GE

- **SLIP** (Symmetric List Processor) programming language (developed in the 1960s when he was on the MIT faculty).

- **ELIZA** (program), one of the earliest forays into developing an AI program designed to respond to the main challenge posed in the Turing test (and influenced his concerns about AI ethics).

- *Computer Power and Human Reason: From Judgment to Calculation* (1976), a now classic book in which he argued "that there are certain tasks which computers *ought* not to be made to do, independent of whether computers *can* be made to do them" (p. x [Italics Weizenbaum's]).

Weizenbaum's Academic Background and his Influence on ICT Ethics (Continued)

- Weizenbaum worried about AI technology in the hands of the military, because of the "potentially disastrous ways" that new technology could be used in warfare.

- He had described the computer as "a child of the military" and as a technology that was "born to the military."

- For example, he noted that this technology was designed initially to "compute ballistic trajectories" (Weizenbaum's interview with ben-Aaron, 1985), as in the case of the military's ENIAC project during WW II.

- So Weizenbaum believed that because of the use of this technology, we had an increased level of moral/social/ professional responsibility in the era of computers and AI.
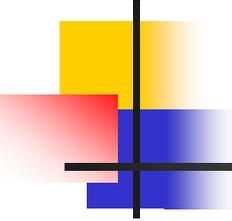
1.3 Some Common Concerns in the Works of Joseph Weizenbaum and Hans Jonas

- Like Weizenbaum, Hans Jonas also believed that our conventional notion of responsibility had been seriously challenged by the impact of "modern technology."

- In many ways, Weizenbaum (1923-2008) and Jonas (1903-1993) were very different kinds of scholars: the former being a computer scientist, the latter a philosopher.

- Also, technically, they are a generation apart (20 years) in virtue of their birth years.

- Yet, these two thinkers also shared some relevant similarities:

  ➢ their ethnic/religious heritage, and experiences (in pre-WW II Germany),

  ➢ their views regarding the dangers of modern technologies (in the post-WW II era).
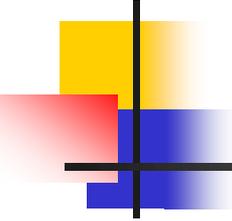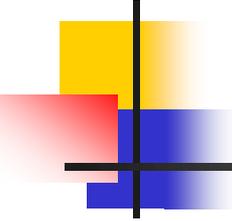
Weizenbaum and Jonas (Continued)

- Both were Jewish refugees who left Germany in the 1930s to escape persecution from the Nazis.

- Arguably, the horrific challenges they experienced during that tumultuous period had informed the ethical positions developed in their respective scholarly writings.

- Weizenbaum stated that "the knowledge of German academics during the Hitler time weighed on me very heavily" (interview with ben-Aaron, 1985).

- Jonas, in his *Memoirs* (2008), described how disturbed he became when Martin Heidegger, Jonas's former teacher and mentor at Marburg University (Germany) had officially joined the Nazi party (before eventually being elevated to the Rector of Freiberg University).
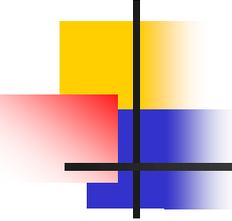
Weizenbaum and Jonas (Continued)

- After leaving Germany, both Weizenbaum and Jonas enjoyed very successful academic careers in the US: Weizenbaum at MIT, and Jonas at the New School for Social Research.

  - Note: I was fortunate to have taught for many years with two colleagues in the Philosophy Department at Rivier University, Lloyd Carr and Jerry Dolan, who took courses with Jonas while they were Ph.D. students at the New School in the late 1960s/early 1970s. I have learned much about Jonas from conversations with them.

- Also, both Weizenbaum and Jonas wrote very influential books during the 1970s, describing in detail their views about the new kinds of challenges that modern technology posed for our traditional conception of moral responsibility.

Weizenbaum and Jonas (Continued)

- Weizenbaum worried that "weapons which threaten to wipe out the human species altogether" could now be delivered with a kind of precision that would not have been possible without computers to "guide these weapons" (ben-Aaron, 1985).

- Jonas (1984) also wrote extensively about the same kinds of worries, which caused him to question our ethical responsibility to:

  ➢ "abstract ethical objects" such as "future generations of humans";

  ➢ the planet we inhabit, in light of the kind of devastating impact that modern technology could have for the future of our species and the ecosystem itself.
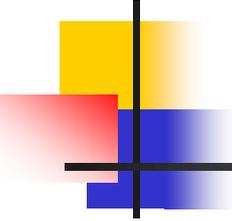
Jonas's Influence

- One of my objectives in this talk is to show why Jonas' insights regarding ethical responsibility are relevant today, especially in analyzing questions that arise in AI.

- As noted above, my fuller analysis of Jonas's ethical system is postponed until a later point in this talk.

- I next shift the focus of my talk to a very different kind of topic/theme: the *state of computer ethics* as a philosophical field.

- I also believe that taking this "detour" now will better pave the way for my analysis of Jonas's ethical framework in the final section.
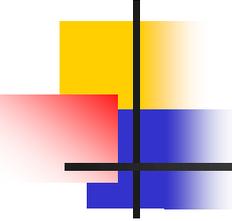
Part 2: The State of the ICT Ethics as a Philosophical Field of Applied Ethics

- Two reasons why I chose to speak on this particular topic:

I. It seemed like suitable topic for a Weizenbaum address, given its breadth.

II. I was eager to revisit and update my views on this topic vis-à-vis what I had written approximately 20 years ago.

➤ In 1999, I was invited to contribute a short piece on the (then) "state of computer ethics" to *the APA Newsletter on Philosophy and Computers*.

➤ I then developed it more fully into an article: "The State of Computer Ethics as a Field of Philosophical Enquiry," published in *Ethics and Information Technology* (2001).
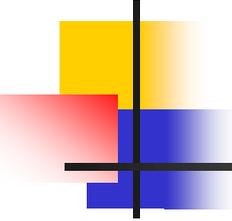
The State of the ICT Ethics (Continued)

- In my 2001 article, I identified and critiqued the following kinds of computer ethics (CE) questions, which seemed pertinent at the time:

1. Is CE a *legitimate field of applied ethics* (i.e., on what grounds does it qualify as a legitimate academic field)?

2. Is *computer ethics* the appropriate name for this (then relatively new) field in applied ethics?

3. What is the *proper scope* of CE (e.g., is it mainly a sub-field of CS)?

4. Are any ethical issues in CE either *new or unique* ethical issues (as Walter Maner [1996] and others had argued)?

5. Does CE require a *brand new ethical theory* to handle the kinds of issues that arise in this field (and would that theory replace existing/traditional theories, as Gorniak-Kocikowska [1996] and other had argued)?

6. Will CE endure as a separate, or at least a distinct, field of applied ethics, or will it *disappear* altogether (as Johnson [1999] had argued)?

7. Do "information objects/entities" deserve at least some level of *moral consideration* (as Floridi [1999] had argued)?
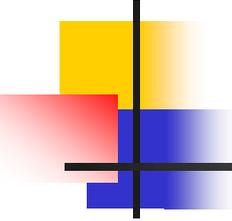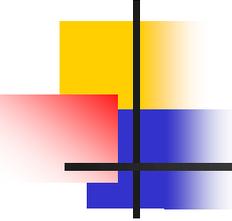
The State of the ICT Ethics (Continued)

- I will not comment on Questions 1-5, since:

  a) my positions on those topics/questions have not significantly changed during the past 20 years,

  b) those questions are no longer as pertinent today, as ICT ethics has since evolved in ways that give rise to other kinds of questions/concerns.

- But I believe that Questions 6 ("Will CE endure as a separate field?") and 7 ("Can some information objects/entities qualify for moral consideration?") are as relevant today as when I first examined them.
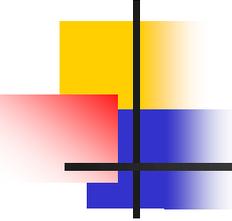
2.1 Will CE Endure as a Separate Field of Applied Ethics (or Will it "Go Away")?

- At least some of you here today may recall a provocative paper by Deborah Johnson (2000), originally delivered as a keynote address at the Ethicomp '99 Conference (Rome 1999), predicting that CE as a separate field of applied ethics would eventually "disappear."

- Johnson's argument is fairly complex and also includes some very subtle points, which I fear I may not do justice to in my brief analysis today.

- But I also believe that, for purposes of this talk, we can review the general outline of her argument to grasp its key points.

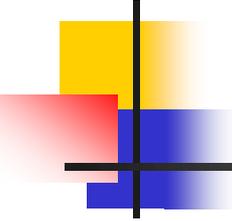# Johnson and the Future of CE/ICT Ethics (Continued)

- I try to summarize Johnson's position via her following (five) claims:

1. When computers were new, some were struck by questions that the use of these new machines raised for our systems of ethics and values.

2. Ethics is ultimately about behavior, and computers simply "mediate human behavior" (albeit in some new ways).

3. Some computer-related ethical issues, such as privacy, initially may have seemed to be different enough from traditional ethical concerns involving privacy to warrant special analysis.

4. In the near future, we will come to view these concerns simply as privacy issues and so we will not need a separate field of CE to examine them.

5. As computers become more and more a part of our ordinary, everyday lives, other ethical issues that are now often associated with computing will simply become part of "ordinary ethics."
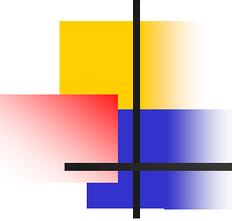
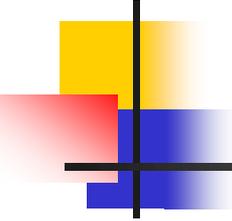# Johnson and the Future of ICT Ethics (Continued)

- Johnson concludes:

  So I offer you a picture of computer ethics in which computer ethics *as such* disappears, as the issues that might involve or center on computer technology are simply issues of ordinary ethics…When computer-ethical issues have disappeared, the field of ethics will have changed not so much in what it sees as its agenda, but in what it presumes about the world it must address…We will be able to say that computer ethics has become ordinary ethics and that ordinary ethics has become computer ethics (2000, p. 31 [Italics Johnson]).

Johnson and the Future of ICT Ethics (Continued)

- Although Johnson's argument seemed fairly persuasive, not everyone at that time was convinced that CE would "go away" altogether.

- For example, I believed that while Johnson's position was compelling with respect to many of the (then) standard or mainstream CE issues, such as privacy, it could also be challenged on at least two points.

1) CE would likely endure as a field of professional ethics (i.e., for computer/IT professions) because important safety and design decisions affecting future software developments would, in all likelihood, continue to arise (e.g., in developing software for nanocomputers, autonomous systems, etc.).

2) CE would likely endure as a philosophical field of applied ethics because of new conceptual muddles and policy vacuums (Moor 1985), regarding concepts such as agency, autonomy, trust and responsibility, which would likely arise in connection with emerging AI-related technologies including "(ro)bots" and autonomous systems.

# Johnson and the Future of ICT Ethics (Continued)

- But I believe that Johnson has been shown to be correct in predicting that some standard CE issues such as privacy and intellectual property would no longer need to be viewed solely, or even mainly, as computer-ethics issues per se, but instead could be analyzed as issues of "ordinary ethics."

- For example, these kinds of issues have since been incorporated into current philosophy textbooks for "contemporary moral issues" courses (see, for instance, Hinman [2013] and Van Camp et al. [2011]).

- Also, some business ethics textbooks now incorporate traditional CE-specific issues, as well.

- We now view many digital-privacy issues simply as privacy issues.

- So there is no longer anything special about privacy concerns that happen to involve digital technology, especially since it is becoming difficult to conceive of many contemporary privacy issues apart or separate from the context of digital technology.
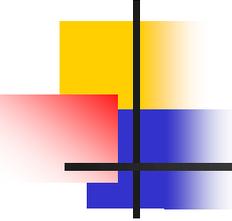
# 2.1.1 The Challenge for ICT Ethics to Remain "Whole" in the Future

- Even if ICT ethics does not disappear, or go away altogether, the field may become increasingly "splintered"?

- So if ICT ethics is not fully subsumed in ordinary ethics, it may become "diluted" or fragmented into many smaller sub-fields such as AI ethics, nanoethics, data ethics, etc. (i.e., *splintering* vs. *subsumption*).

- Johnson (2008) raised this point when she asked whether information ethics and computer ethics will continue to be splintered, or if it were still possible for the two to remain integrated. She writes:

  While the flourishing of the field of computer ethics is to be celebrated, growth inevitably means pressure to split the whole into parts. The topics that need to be addressed continue to expand and perspectives from a wide range of disciplines are relevant. Thus, there is pressure for the field to become *splintered into subfields* (for example, with a distinction between computer ethics and information ethics [p. xi, Italics Added].
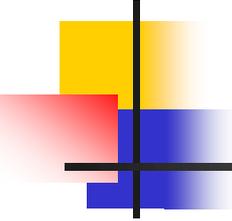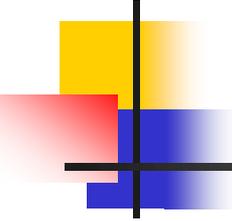
Future Challenges for ICT Ethics (Continued)

- Of the many recently emerging sub-fields in ICT ethics, one that poses a significant challenge for keeping ICT ethics whole is from an area that some refer to as "robot ethics" and others call "machine ethics," but which I will henceforth call "AI ethics."

- Can/Should this relatively new research area of AI ethics be viewed as a *separate field* of applied ethics?

- Or, alternatively, can AI ethics be understood more appropriately as a *distinct sub-field* of ICT ethics?

- Consider that now some professional organizations, as well as international conferences and journals, are devoted exclusively to AI ethics (and robo-ethics in particular).

- So, will the time soon come for AI ethics and ICT ethics to "part company" and go their separate ways?
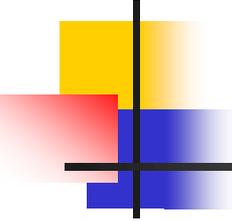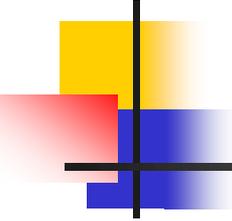
2.1.2 The Challenge from AI Ethics

- How is AI ethics different from ICT (or computer) ethics?
- Wallach and Allen (2009, p. 61) view AI ethics (or what they call "machine ethics") as a field that extends or *expands* upon (traditional) computer ethics in at least two ways, because it:

1) shifts the concern away from "what people do with computers to questions about what machines do by themselves";

2) fosters "a discussion of the technological issues involved in making computers themselves into explicit moral reasoners."
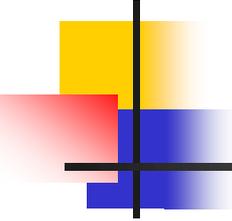
The Challenge from AI Ethics (Continued)

- Does the kind of "expansion" described by Wallach and Allen constitute the grounds for an entirely separate field of applied ethics, i.e., separate from ICT ethics?

- Or does it instead suggest that the domain of ICT ethics itself needs, once again, to be expanded to include this and other new sub-fields?

- Only time will tell which directions ICT ethics and AI ethics eventually take.

- But let's assume for the sake of argument that ICT ethics can adequately and successfully subsume AI ethics as sub-field, which is *distinguishable* but not *separable* (to invoke David Hume's famous expression) from ICT ethics.
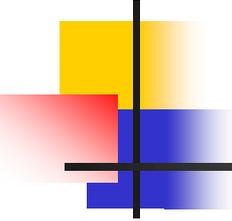
The Challenge from AI Ethics (Continued)

- In one sense, those outside academia, and perhaps even some academics outside the fields of philosophy and computer science, might ask: "Who cares?"

- For example, they might object by claiming that the only important thing is that the ethical concerns are addressed by some academic discipline (or perhaps by some appropriate regulatory organization).

- I raise this concern, in part, as a speculative question.

- But I also raise it because I wonder what role ICT ethics (and, for that matter, CEPE conferences) would have in the future, if AI ethics were to "leave home."

The Challenge from AI Ethics (Continued)

- Will ICT ethics possibly survive as a kind of "umbrella" field comprising a cluster of loosely-related ICT issues, which include AI ethics?

- Or will ICT ethics' time as an independent field "have come and gone" (in much the same way that Johnson had predicted CE's demise as a separate field 20 years ago)?

- I suspect that ICT ethics will survive at some level.

- And I also hope that CEPE conferences continue to flourish and that INSEIT does not suffer the same fate as some like-minded organizations such as CPSR.

- One thing seems fairly certain: AI-ethics-related-issues will continue to need the kind of intense attention and philosophical scrutiny they currently receive.

The Challenge from AI Ethics (Continued)

- One very important concern in the area of AI ethics, as has already been noted, is the question of whether at least some AI entities deserve, or will eventually qualify for, moral consideration.

- Recall that this specific concern was implied in Question 7 in my above list (at the beginning of Section 2).

- Questions about whether AI entities warrant moral status had begun to receive some level of attention in the ICT ethics community, at least as far back as 1999.

- Floridi (1999) argued that "information objects" have intrinsic worth and have moral standing (at least as "moral patients," a concept that I describe and employ later in the talk).
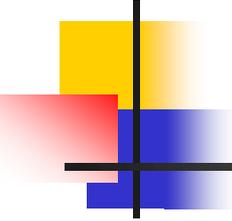
2.2 Can (at Least) Some AI Entities Qualify for Moral Consideration?

- Floridi argued that just as we now grant moral consideration to biological and ecological entities in the "biosphere" or "ecosphere" (e.g., to animals and plants), we again need to extend the domain of moral consideration to include non-biological entities ("information objects") that comprise what he called the "infosphere."

- He elaborates upon these points in his *Information Ethics* (*IE*) framework, which I will not discuss today.

- I do not need to embrace Floridi's IE framework to examine the kinds of moral-consideration-related concerns that I focus on this talk.

- Also, I agree with Brey (2008, p. 109) who argues that Floridi has given no convincing argument to show either (a) that all information objects deserve moral respect *because* they have intrinsic value, or (b) that all of these objects have intrinsic value.

- But we should acknowledge Floridi for having raised some key questions and for his initial contribution to the debate.
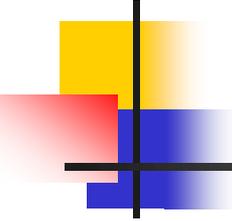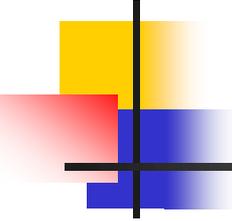
AI Entities and Moral Consideration (Continued)

- A related/extended, and even more controversial, question that has recently been debated in AI ethics is whether some AI entities deserve full-fledged rights.

- In October 2017, a sophisticated robot named "Sophia" was officially granted full civil rights by the government of Saudi Arabia.

- Currently, the European Union is also considering the question of whether and how to grant legal status/rights/protections to certain kinds of robots (Committee on Legal Affairs, 2016).

- So the controversial question of robot rights has gained the attention of law/policy makers in many countries, as well as the interest of academics.
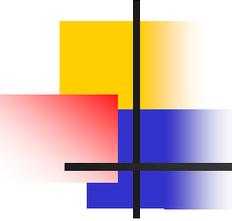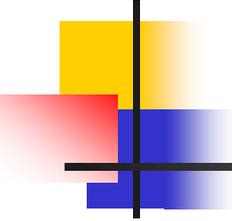
AI Entities and Moral Consideration (Continued)

- The robot-rights question is ambiguous and problematic on a number of levels.

- For one thing, the nature of rights and "the language of rights" have posed some thorny conceptual questions for philosophers and ethicists.

- Also, many authors writing in the context of robots have not always been clear with regard to which kinds of rights they have in mind when examining this question – i.e., legal rights, moral rights, civil rights, etc.

- Some, including Darling (2016, p. 228), have recently proposed granting "legal rights" to robots that meet Darling's criteria of "social robot," which she defines (p. 215) as a "physically embodied, autonomous agent that communicates and interacts with humans on a social level."
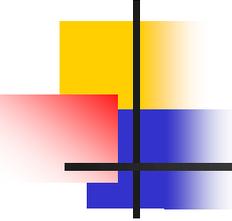
AI Entities and Moral Consideration (Continued)

- Is it possible that the some of the most sophisticated social robots might one day qualify for full-blown *rights*?

- Traditionally, we have tended to grant full moral rights only to (human) moral agents.

- But recall Floridi's distinction between *moral agents* and *moral patients*, where the former are "the sources of moral action" and the latter are described as "the receivers of moral action."

- Employing this useful distinction, we could consistently grant moral consideration to social robots as moral patients (but without having to embrace Floridi's IE framework), even if they do not meet the criteria for full-fledged rights (moral agents).

- A growing number of authors argue that *social robots* can qualify for moral consideration.

AI Entities and Moral Consideration (Continued)

- On what grounds could we justify granting some level of moral status to social robots?

- Coeckelbergh (2010), Gunkel (2017), Gerdes (2015), and others have provided some interesting rationales for how we can do this.

- Coeckelbergh (p. 209) uses his "relational account," based on "social relations" (which he contrasts with a traditional "property account" based on an "ontological framework," appealing to consciousness, intentionality, sentience, etc.).

- Gunkel argues that Levinas's notion of "the other" can be adapted and applied in the case of social robots to show why they deserve moral consideration as *moral patients*.

- Gerdes, who also agrees that social robots can qualify as moral patients, argues that Kant's moral theory – in particular, its account of "indirect duties" – provides us with a better rationale.

AI Entities and Moral Consideration (Continued)

- While I admire the insights that Coeckelbergh, Gunkel, and Gerdes each offer, I do not believe that their respective positions provide us with the most satisfactory answer to the question at hand.

  - For a critique of their respective arguments, see (Tavani 2018) which I will not repeat here.

- I believe that Jonas (1984) provides us with a more adequate ethical framework for showing why social robots can qualify for moral consideration.

- I also believe that his framework can help us in analyzing some other kinds of AI-ethics issues, as well as some broader ICT-ethics-related issues.
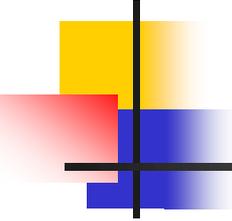
Part 3. Jonas's Ethical Framework and its Application to Issues in AI Ethics

- Before describing Jonas's theory in detail, I briefly address a potential criticism that I anticipate from some who might interpret my position regarding Jonas as supporting the view that a brand new ethical theory is needed to handle AI issues.

- Recall Question 5 in Section 2.1: "Does ICT Ethics (or CE as it was then called) require a brand new ethical theory/framework that would replace existing ethical theories?" (as Gorniak-Kocikowska [1996] and others had previously argued).

- It is important to note that I do not advocate using Jonas's ethical framework in a way that would *replace* existing (traditional) ethical theories.
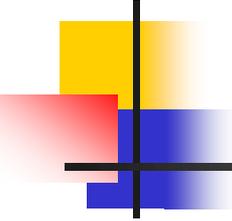
3.1 Skepticism Regarding Claims for the Need of a Brand New Ethical Framework

- But I can see why some might be inclined to interpret Jonas's ethical system as put forth to replace altogether existing or traditional ethical theories, in the age of "modern technology."

- Although relatively little has been written thus far about Jonas's ethical system in the context of ICT ethics per se, Michelfelder (2000) had briefly questioned whether Jonas's ethical theory was needed to understand "our moral condition in cyberspace."

- In a very thoughtful article, she argued that we have "good reasons… to be skeptical of any claim that there is a need for a new, cyberspace ethics to address the moral dilemmas arising from these technologies" (Michelfelder, p. 147).
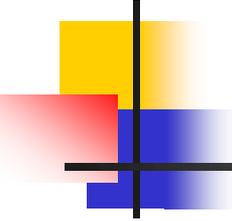
Michelfelder's Critique

- Michelfelder also argued that our moral condition has not (at least not yet) been so radically altered that we needed a new ethical system like Jonas's to handle the issues that arise.

- At that time, I was in complete agreement with Michelfelder on this point.

- However, Michelfelder also concluded her article by noting that

  If cyberspace were to give us *new objects of ethical consideration*, we would then have reason to believe, following Jonas, that [this technology] is ushering us into a new moral condition (p. 152 Italics Added).
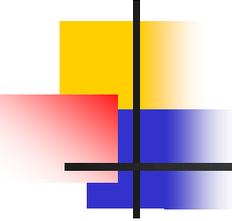
Michelfelder's Critique (Continued)

- I now believe that some new objects of ethical consideration – viz., sophisticated AI entities such as social robots – have since been introduced, and that these entities have affected (or soon will affect) our moral condition.

- But I also still believe Jonas's theory does not need to be viewed as "replacement" ethical theory, as I will show.

- Apart from controversial claims that one or more new/alternative/replacement ethical theories are needed, the field of ICT ethics issues has benefited from the introduction of (what I view as) some new *methodological frameworks* during the past two decades.
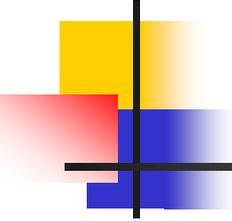
Methodological Frameworks

- Among the many thoughtful methodological frameworks that have been put forth during that period, at least four have been especially useful:

- Philip Brey's "Disclosive Computer Ethics" model (2000),

- Batya Friedman et al.'s "Value Sensitive Design" (VSD) framework (2008),

- James Moor & John Weckert's "Dynamic Ethics" model (2004),

- and Brey's "Anticipatory Technology Ethics" (ATE) framework (2012).

- It is also worth noting that none of these frameworks aims at replacing traditional ethical theories.
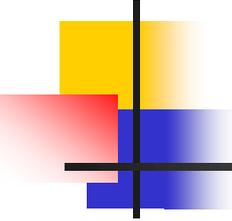
Methodological Frameworks Continued)

- Another interesting framework, put forth approximately 20 years ago, is Moor's "Just Consequentialism" (Moor 1999).

- Unlike the four methodological frameworks described above, Moor's theory combines some key elements in traditional deontological and utilitarian ethical theories.

- But like the other four frameworks, Moor's theory does not aim to "replace" existing ethical theories.

  ➢ For example, his theory can work in tandem with traditional ethical theories, such as virtue ethics, in a way that can supplement those theories.

- So it is in the spirit of this kind of *supplementary* framework that Jonas's ethical system can also be understood and applied in ICT ethics, and especially in the sub-field of AI ethics.
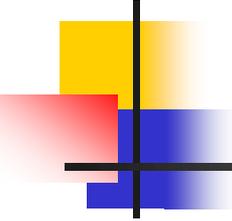
3.2 Applying Jonas's Framework in the Context of AI Ethics

- Understand the context, or background, in which Jonas developed his system of ethics:

- Jonas was a student of Martin Heidegger (1889-1976), and was influenced by his teacher's notion of "being-in-the-world"

- In Heidegger's view, humans do not exist simply as entities isolated (or apart) from "the world"; rather, their very existence or being (*dasein*) is one that is already *in* the world.

- Jonas applied/extended Heidegger's concepts to our modern "technological-world," suggesting that we now exist in (i.e., have "being-in") the "technological-world" (Carr 2018).
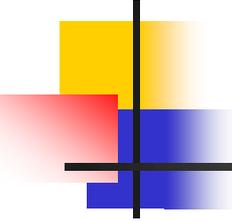
Heidegger's Influence on Jonas

- For Jonas, contemporary human existence can now be viewed in terms of a "technological network" in which we interact in significant ways with some non-human entities/objects (Carr 2018).

- Also, "being-in-the-technological-world" means that we, as humans, are now able to act in "novel" (and much more powerful) ways.

- For example, Jonas (2008) notes that we can now easily render the entire earth uninhabitable via nuclear technology, as well as through pollution and over consumption of natural resources.

- So Jonas proposed a ethical framework based on a new "imperative of responsibility."
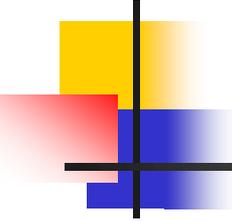
A Sketch of Jonas's Argument

- I will try to summarize Jonas's position via the following argument structure:

- **P1:** Traditional ethical systems have tended to view *moral responsibility* as something that was restricted to human action in inter-human relationships – a kind of "neighbor ethics" limited to the "direct dealing of man with man" in the "here and now" (Jonas 1984, p. 8 [Italics Added]).

- **P2:** Our traditional conception of ethics is no longer adequate in the era of "modern technology," because the nature of human action has changed in fundamental ways – i.e., "new objects of action have been added," which in turn have opened up for us "a whole new dimension of ethical relevance" (Jonas, p. 1).

- **P3:** Because ethics is concerned with human action, "a change in ethics" is now required; that is, "novel powers to act" made possible by modern technology require "novel ethical rules" and have created for us a new "moral condition" (Jonas, p. 23).

- **P4:** Our moral condition has been altered by the need to extend the sphere of moral consideration to provide for these additional "ethical objects," which include some "abstract objects" such as future generations of human beings, as well as (the whole of) nature itself (Jonas, pp. ix-x).

- **C1:** The new moral condition ushered in by the "altered nature of human action" and the introduction of new ethical objects has raised some important issues of responsibility, and so we now have a "new dimension of responsibility" (Jonas, p. 6).
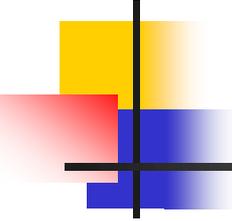
Extending Jonas's Argument to the Impact of AI Entities for our Moral Condition

- **P5** (= C1 from the preceding inference): The new moral condition ushered in by the "altered nature of human action" and the introduction of new ethical objects has raised some important issues of responsibility, and so we now have a "new dimension of responsibility" (Jonas, p. 6).

- **P6:** Once again, some "new objects," viz., some very sophisticated AI entities such as social robots have been added, which increase our powers to act in significant ways, and thus (once again) open up what Jonas (p. 1) refers to as a "new dimension of ethical relevance".

- **C2:** Since these "new objects" or entities – i.e., sophisticated social robots – also qualify as "ethical objects," our moral condition is once again altered, and we need to expand further our "dimension of responsibility" (Jonas, p. 6) to include them.
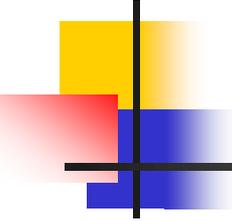
A Jonas-Inspired Framework Comprising Three Questions

- Employing Jonas's ethical system, I propose a framework comprising the following three questions to determine whether some new/emerging AI entity qualifies as an *ethical* object:

1) Has the introduction of this object/entity significantly changed the nature of human action (i.e., by providing us with "novel powers to act" in a ways that opens up a "new dimension of ethical relevance")?

2) Can this new object/AI entity significantly affect "future generations of humans," as well as the sustainability of our planet, in ways that cause serious harm?

3) Will this new object/AI entity (even though it is a non-biological-entity/life-form) significantly modify our "moral condition" and enhance our "dimension of responsibility"?
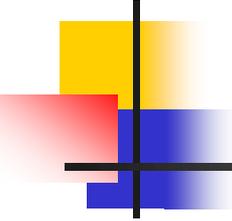
Further Additional Questions (Assuming the truth of C2, Above)

- Two questions to consider regarding **our** (human) moral responsibility in the future development of these entities:

1) Before further developing sophisticated AI entities, such as social robots, do we have a responsibility to figure out how to co-exist with them in our (current and future) "technological world"? (See, e.g., de Laat [2016]).

2) Do we have a responsibility to contain these new objects – e.g., to restrict (or possibly even prohibit altogether) the future development of these entities, so that they do not become either too intelligent or too autonomous, i.e., in ways that they can seriously threaten either: (a) future generations of humans, or (b) the viability of our planet?
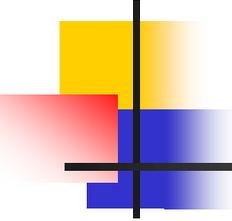
Further Questions for Consideration (Assuming the truth of C2) Continued

- Two questions worth examining with regard to granting moral consideration **to** these AI entities:

1) What kinds of responsibilities do we (humans) have towards these new objects, given that these "intelligent" and "autonomous" entities are unlike the "new ethical objects" previously introduced, in that the former are neither biological entities nor (natural) life forms?

2) If we grant moral status as moral patients to some (currently available) sophisticated AI entities, will we possibly need to grant certain kinds of *rights* (e.g., legal rights) to some more advanced AI entities in the future?
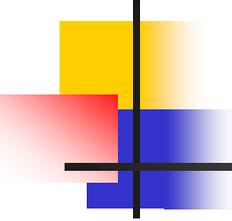
Concluding Remarks

- Covering a wide range of topics in this talk, I have left far more questions unanswered than satisfactorily resolved.

- Regarding the outlook for ICT ethics in the near future, I remain somewhat optimistic that our field will continue to thrive.

- However, I am unclear as to what the field will look like twenty years from now, as well as what it might eventually be called.

- I feel fairly confident that questions affecting the moral status of sophisticated AI entities will continue to be an important question.

- As far as which ethical frameworks future ICT-ethics researchers will use to debate/analyze that and related questions, I am again uncertain.

- But I hope that some younger scholars will further investigate Jonas's ethical framework and examine its potential implications for aiding us in other areas of ICT-ethics-related research as we go forward.
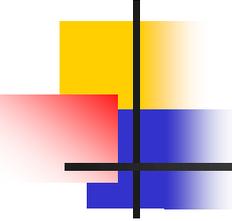
Concluding Remarks (Continued)

- I began this talk by very briefly describing some ethical concerns that Joseph Weizenbaum raised approximately fifty years ago, regarding developments in AI and especially its potential use by the military.

- I also briefly described, albeit in a cursory way, some comparisons between the kinds of ethics-and-technology-related worries expressed both by Weizenbaum and Jonas.

- To the best of my knowledge, these two professors never met in person, nor had even communicated with one another.

- But I cannot help wondering what might have transpired if the two influential thinkers had indeed met or had at least communicated, especially given their similar experiences as German-Jewish refugees in the era leading up to WW II.
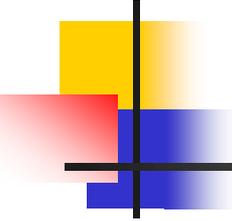
Concluding Remarks (Continued)

- We can now only speculate about what might have happened if these two men had met.

- Would Weizenbaum, a computer scientist concerned about threats posed by AI, possibly have influenced Jonas to develop his ethical system in a slightly different direction?

- Would Jonas's position on our "imperative of responsibility" in the age of modern technology have influenced the direction of Weizenbaum's views regarding our increased level of responsibility at the dawn of the AI era?

- I would still like to imagine a scenario in which these two giant 20th thinkers had actually met in person and shared with each other their thoughts on the ethical impact of modern technology.

- And, I would love to have been a "fly on the wall" in that meeting room.

- THANK YOU for your attention and patience!
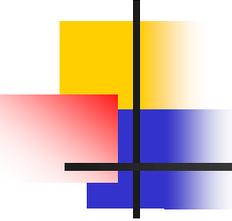
References

Ben-Aaron, D. (1985). "Weizenbaum Examines Computers and Society" (Interview with Joseph Weizenbaum). *The Tech*, Vol. 105, No. 16, April 9, pp. 1-4. Available at http://tech.mit.edu/V105/N16/weisen.16n.html.

Brey, P. (2000). "Disclosive Computer Ethics." *Computers and Society*, Vol. 30, No. 4, pp. 10-16.

Brey, P. (2009). "Do we Have Moral Duties Towards Information Objects?" *Ethics and Information Technology*, Vol. 10, Nos. 2-3, pp. 109-114.

Brey, P. (2012). "Anticipatory Ethics for Emerging Technologies." *Nanoethics*, Vol. 6, No. 1, pp. 1-13.

Bynum, T.W. (2009). "Philosophy and the Information Revolution." In M. Bottis, ed. *Proceedings of the 8th International Conference on Computer Ethics: Philosophical Enquiry*. Athens, Greece: Nomiki Bibliothiki, pp. 30-52.

Carr, L. 2018. "On What Grounds Might We Have Moral Obligations to Robots?" Available at https://www2.rivier.edu/faculty/lcarr/OUR%20MORAL%20OBLIGATION%20TO%20ROBOTS.pdf.

Coeckelbergh, M. (2010). "Robot Rights? Towards a Social-Relational Justification of Moral Consideration." *Ethics and Information Technology*, Vol. 12, pp. 209-221.

Committee on Legal Affairs. (2016). Draft Report with Recommendations to the Commission on Civil Law Rules on Robotics. European Parliament. Available at http://www.europarl.europa.eu/sides/getDoc.do?type=COMPARL&reference=PE-582.443&format=PDF&language=EN&secondRef=01.

Darling, K. Extending Legal Protection to Social Robots: The Effects of Anthropomorphism, Empathy, and Violent Behavior towards Robotic Objects. In *Robot Law*; Calor, R., Froomkin, A.M., Keri, I., Eds.; Edgar Elgar Publishing: Cheltenham, UK. 2016, pp. 213-231.

de Laat, P.B. (2016). "Trusting the (Ro)botic Other: By Assumption?" *ACM SIGCAS Computers and Society*, Vol. 45, pp. 255-260.

Floridi, L. (1999). "Information Ethics: On the Philosophical Foundations of Computer Ethics," *Ethics and Information Technology*, Vol. 1, No. 1, 1999, pp. 37-56.

Floridi, L. (2002). "On the Intrinsic Value of Information Objects in the Infosphere." *Ethics and Information Technology*, Vol. 4, pp. 287-304.

Friedman, B., Kahn, P., and Borning, A. (2008). "Value Sensitive Design in Information Systems." In *The Handbook of Information and Computer Ethics*. Hoboken, NJ: John Wiley and Sons. Edited by Himma, K. E., and H. T. Tavani, H. T., pp. 69-101.
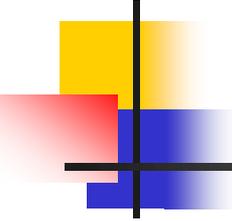
References (Continued)

Gerdes, A. (2015). "The Issue of Moral Consideration in Robot Ethics." *ACM SIGCAS Computers and Society*, Vol. 45, pp. 274-279.

Gorniak-Kocikowska, K. (1996). "The Computer Revolution and the Problem of Global Ethics," *Science and Engineering Ethics*, Vol. 2, No. 2, pp. 177-190.

Gotterbarn, D. (1995). "Computer Ethics: Responsibility Regained." In D. G. Johnson and H. Nissenbaum, eds. *Computing, Ethics, and Social Values*. Upper Saddle River, NJ: Prentice Hall, pp. 18-24.

Gunkel, D. J. (2017). "The Other Question: Can and Should Robots Have Rights?" *Ethics and Information Technology*, Vol. 19, DOI 10.1007/s10676-017-9442-4.

Himma, K. E., and Tavani, H. T. (2008). "Introduction." In K. E. Himma and H. T. Tavani, eds. *The Handbook of Information and Computer Ethics*. Hoboken, NJ: John Wiley and Sons.

Hinman, L., ed. (2013). *Contemporary Moral Issues: Diversity and Consensus*. 4th ed. Upper Saddle River, NJ: Pearson.

Johnson, D. G. (1985). *Computer Ethics*. Englewood Cliffs, NJ: Prentice Hall.

Johnson, D. G. (1994). *Computer Ethics*. 2nd ed. Englewood Cliffs, NJ: Prentice Hall.

Johnson, D. G. (2000). "The Future of Computer Ethics." In G. Collste, ed. *Ethics in the Age of Information Technology*. Linköping, Sweden: Centre for Applied Ethics, pp. 17-31.

Johnson, D. G. (2008). "Foreword." In. K. E. Himma and H. T. Tavani, eds. *The Handbook of Information and Computer Ethics*. John Wiley and Sons, pp. xi-xii.

Jonas, H. (1984). *The Imperative of Responsibility: In Search of an Ethics for the Technological Age*. Chicago, IL: University of Chicago Press.

Jonas, H. (2008). *Memoirs*. (Edited and annotated by Wiese, C. ; translated from German by Winston, K.). Brandeis University Press: Waltham, MA.

Langford, D., ed. (2000). *Internet Ethics*. London, UK: Macmillan.

Levinas, E. *Totality and Infinity: An Essay on Exteriority*. (Trans. Lingis, A.). Duquesne University Press: Pittsburgh, PA, 1969.

Maner, W. (1996). "Unique Ethical Problems in Information Technology." *Science and Engineering Ethics*, Vol. 2, No. 2, 1996, pp. 137-154.
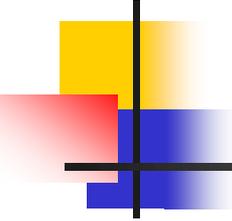
References (Continued)

Moor, J. H. (1985). "What Is Computer Ethics?" *Metaphilosophy*, Vol. 16, No. 4, 1985, pp. 266-275.

Moor, J. H. (1999). "Just Consequentialism and Computing." *Ethics and Information Technology*, Vol. 1, No. 1, pp. 65-69.

Tavani, H. T. (2000). "Computer Ethics: Current Perspectives and Resources." *APA Newsletter on Philosophy and Computers*, Spring, pp. 166-170.

Tavani, H. T. (2001). "The State of Computer Ethics as a Philosophical Field of Inquiry: Some Contemporary Perspectives, Future Projections, and Current Resources." *Ethics and Information Technology*, Vol. 3, No. 2, pp. 97-108.

Tavani, H. T. (2018). "Can Social Robots Qualify for Moral Consideration? Reframing the Question about Robot Rights." *Information*, Vol. 9, pp. 1-16. doi:10.3390/info9040073

Turkle, S. *Alone Together: Why We Expect More from Technology and Less from Each Other*. Basic Books: New York, 2011.

Van Camp, J., Olen, J., and Barry, V., eds. (2011). *Applying Ethics: A Text with Readings*. 10th ed. Belmont, CA: Thompson.

Wallach, W., and Allen, C. (2009). *Moral Machines: Teaching Robots Right from Wrong*. New York: Oxford University Press.

Weizenbaum, J. (1976). *Computer Power and Human Reason: From Judgment to Calculation*. San Francisco, CA: Freeman.

Weiner, N. (1948). *Cybernetics: or Control and Communication in the Animal and the Machine*. Technology Press.