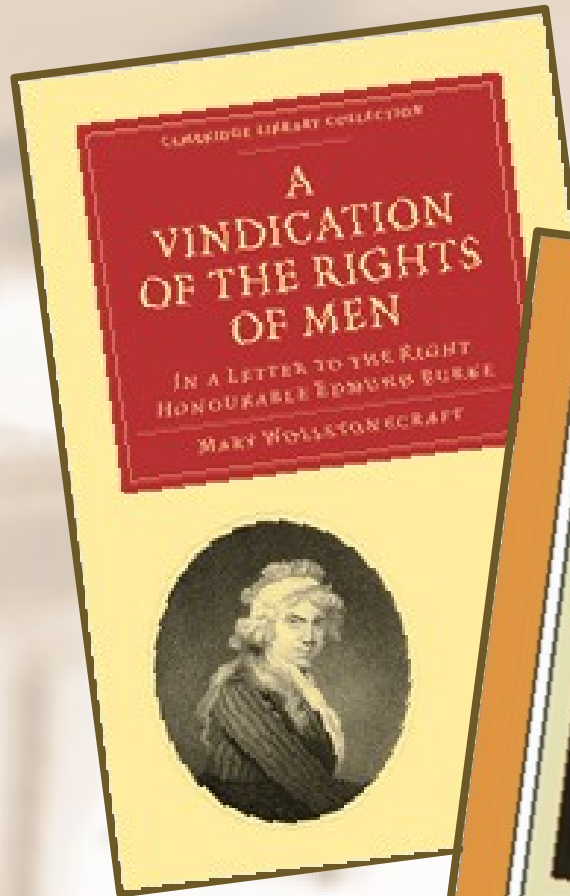




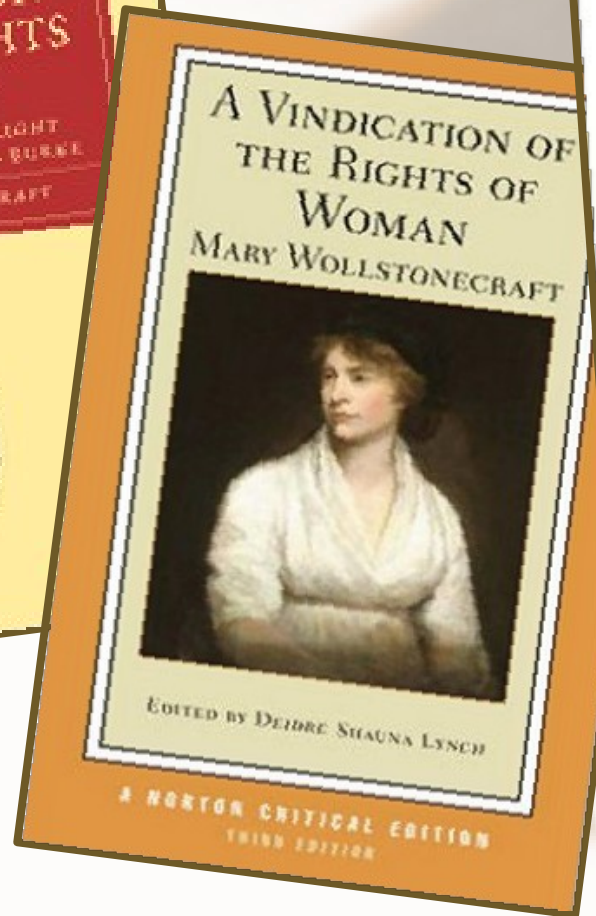
A Vindication of the Rights of Machines

David J. Gunkel
Northern Illinois University
dgunkel@niu.edu

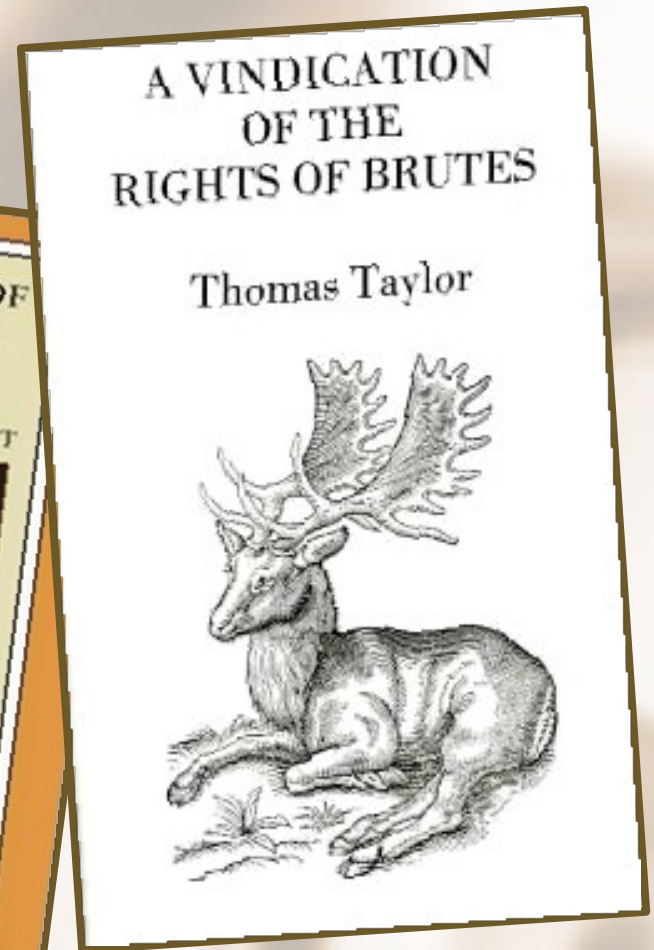
Introduction



1790

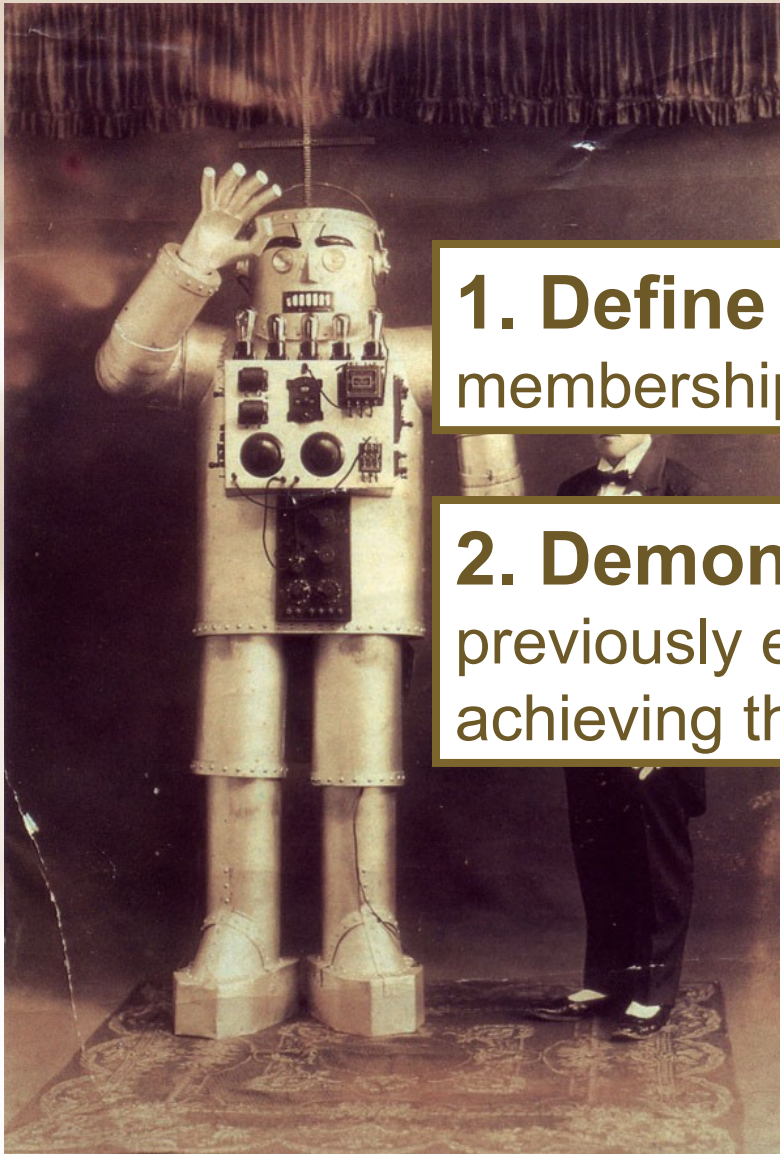


1792



1792

Introduction



1. Define Criteria - conditions for membership in “the club of *consideranda*”

2. Demonstration – Show that some previously excluded entity is capable of achieving threshold level for inclusion

Method and Approach

Instead of demonstrating that machines or at least one representative machine is able to achieve the necessary and sufficient conditions for moral standing, the following both contests this procedure and demonstrates the opposite, showing how the very criteria that have been used to decide the question of moral considerability necessarily fail in the first place.





The background of the slide is a blurred image of several interlocking metal gears. The gears are a light brown or tan color, and the image is out of focus, creating a bokeh effect. The lighting is soft, with some highlights on the gear teeth.

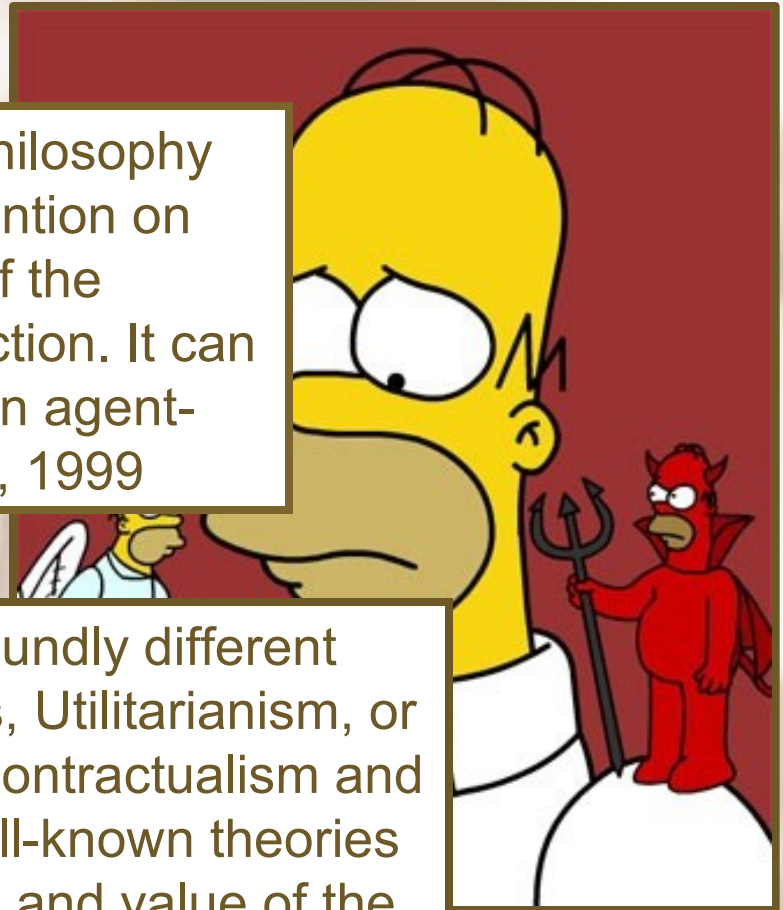
1

Moral Agency

Moral Agency

Ancient: Virtue ethics, and Greek philosophy more generally, concentrates its attention on the moral nature and development of the individual agent who performs the action. It can therefore be properly described as an agent-oriented, “subjective ethics.”– Floridi, 1999

Modern: Developed in a world profoundly different from the small, non-Christian Athens, Utilitarianism, or more generally Consequentialism, Contractualism and Deontology are the three most well-known theories that concentrate on the moral nature and value of the actions performed by the agent. – Floridi, 1999



Moral Agency

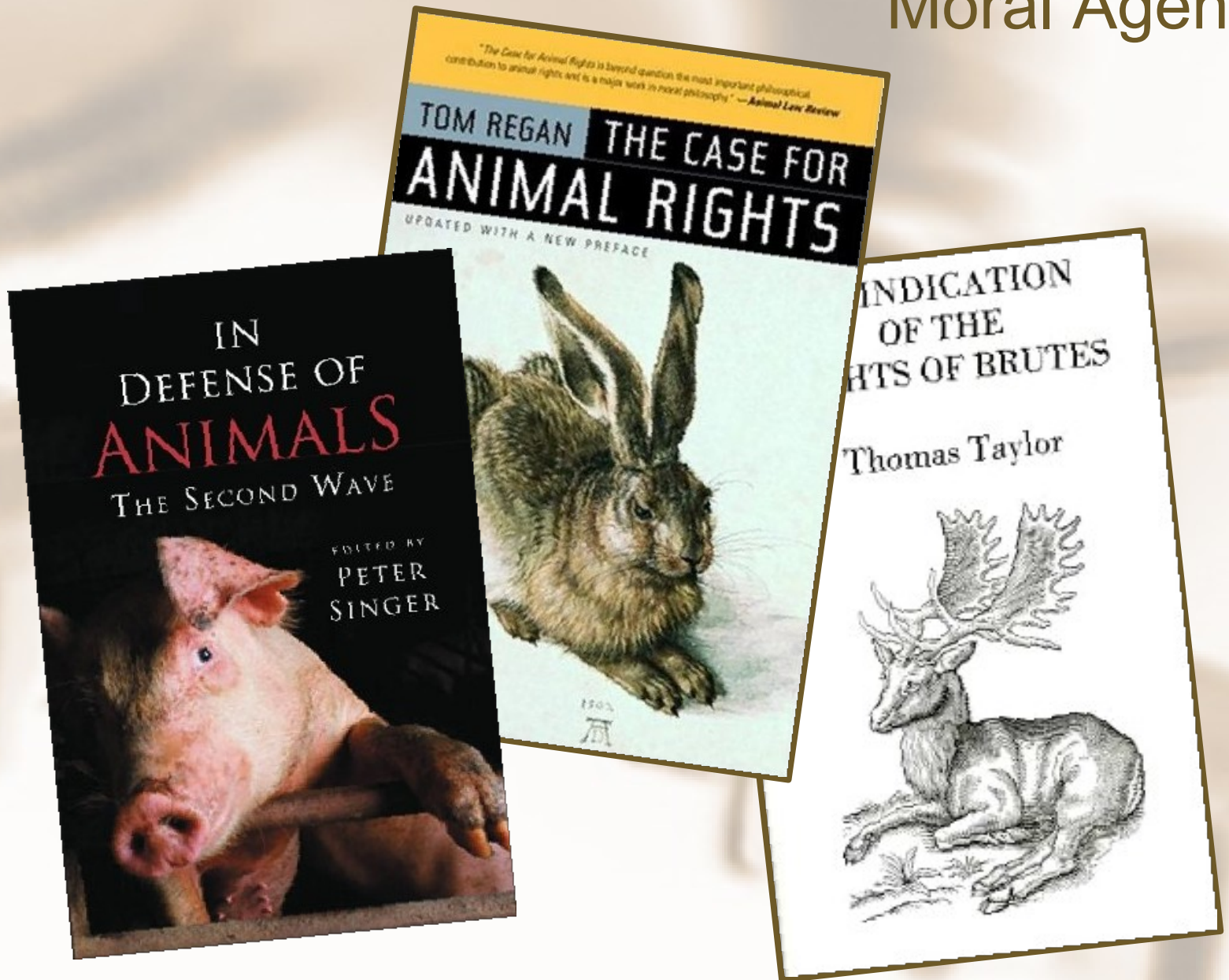
who



what



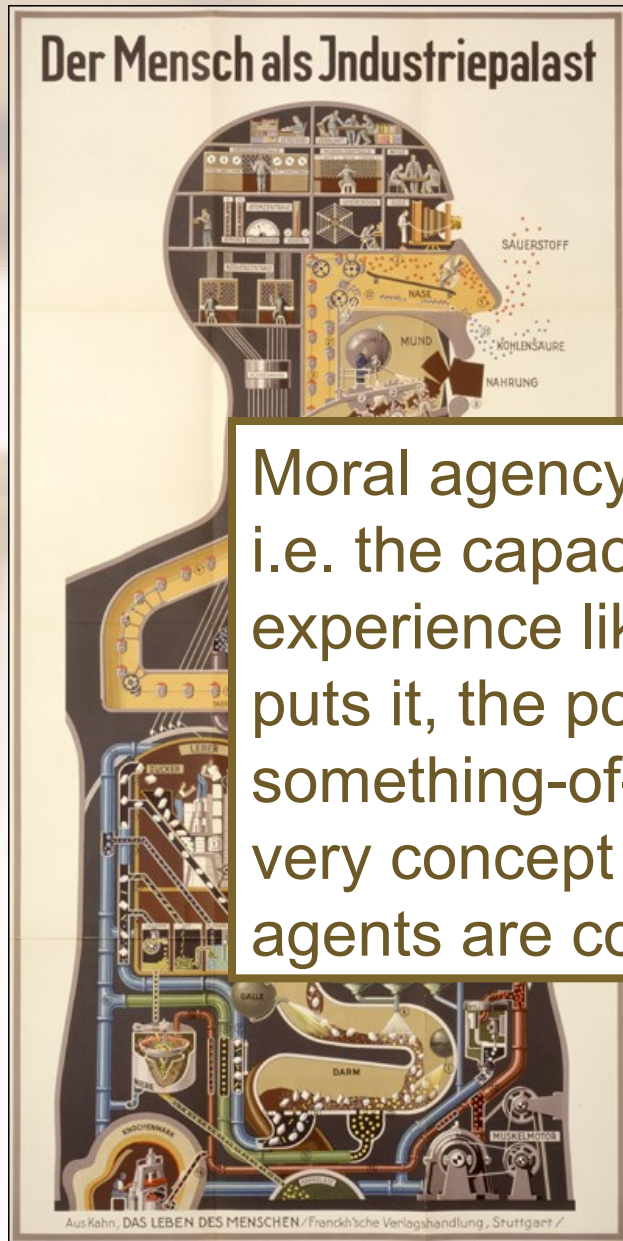
Moral Agency



Moral Agency

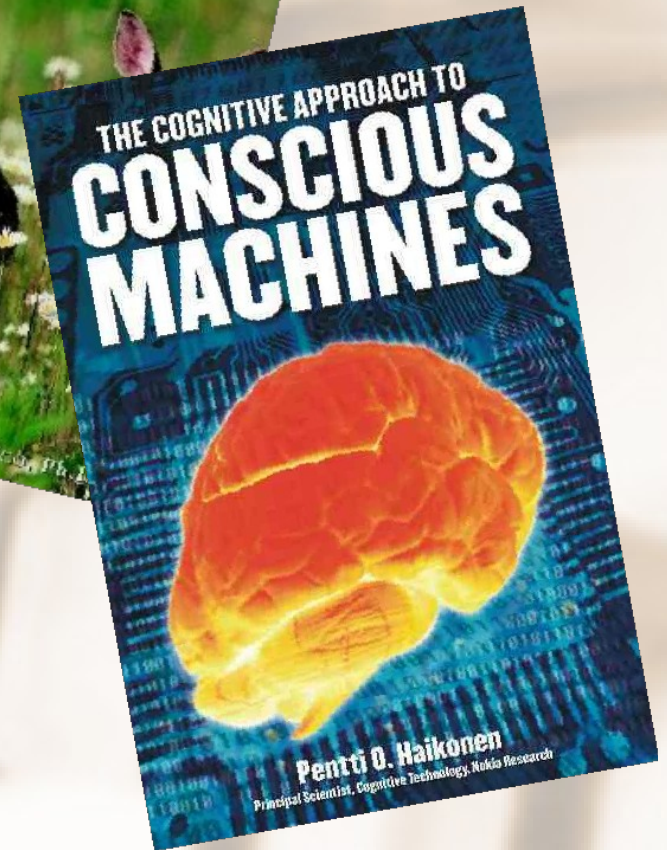
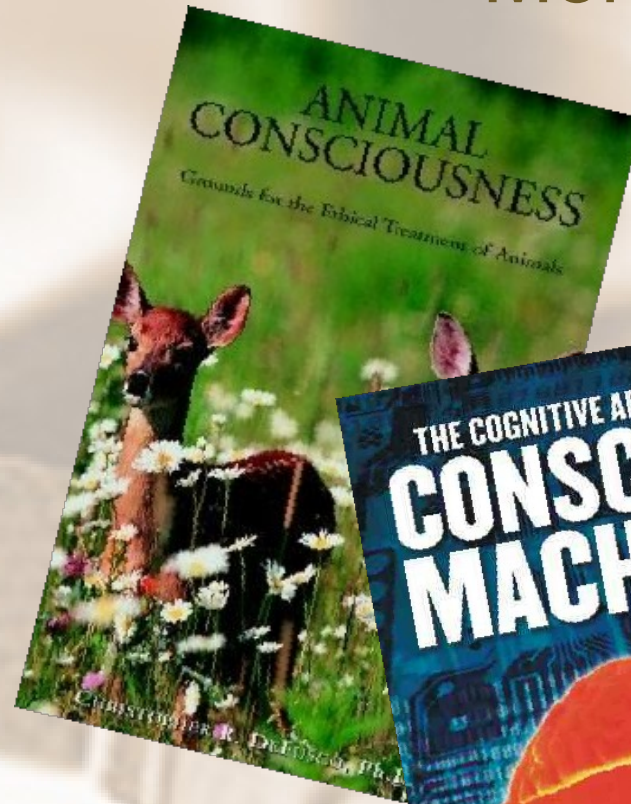
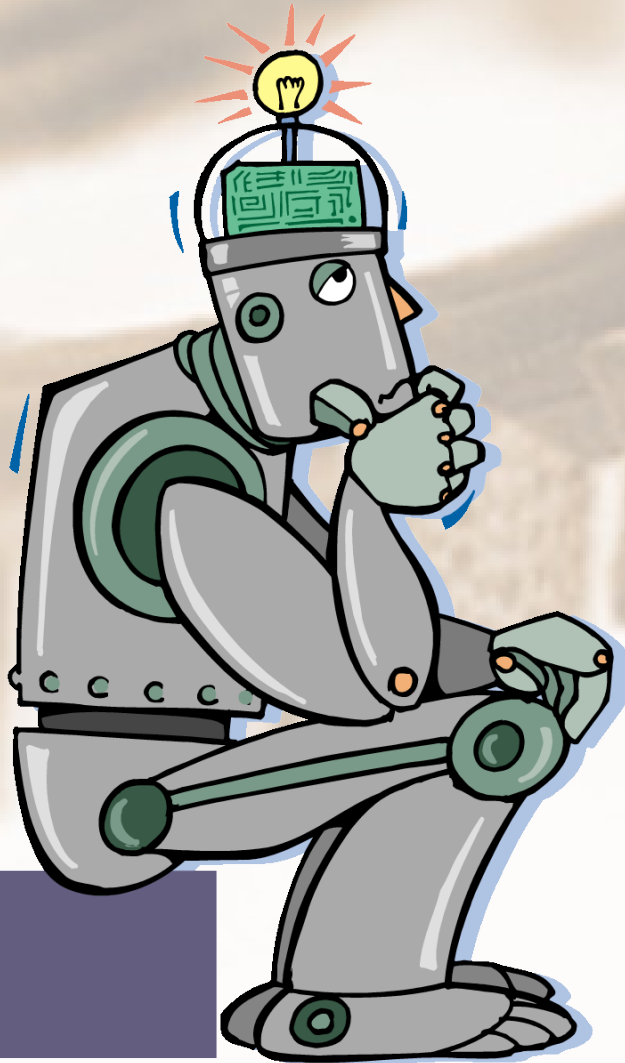


Moral Agency



Moral agency presupposes consciousness, i.e. the capacity for inner subjective experience like that of pain or, as Nagel puts it, the possession of an internal something-of-which-it-is-to-be and that the very concept of agency presupposes that agents are conscious. – Himma, 2009

Moral Agency



Moral Agency

The term means many different things to many different people, and no universally agreed core meaning exists. – Velmans, 2012

The image shows the front cover of the book 'Understanding Consciousness' by Max Velmans. The top half of the cover features a colorful, abstract pattern of overlapping triangles and lines in shades of green, blue, and purple. Below this pattern, the title 'Understanding Consciousness' is printed in a large, white, serif font. Underneath the title, 'Second Edition' is written in a smaller, teal-colored font. At the bottom of the cover, the author's name 'Max Velmans' is printed in a white, serif font. The background of the bottom half of the cover is a solid teal color.

Understanding Consciousness

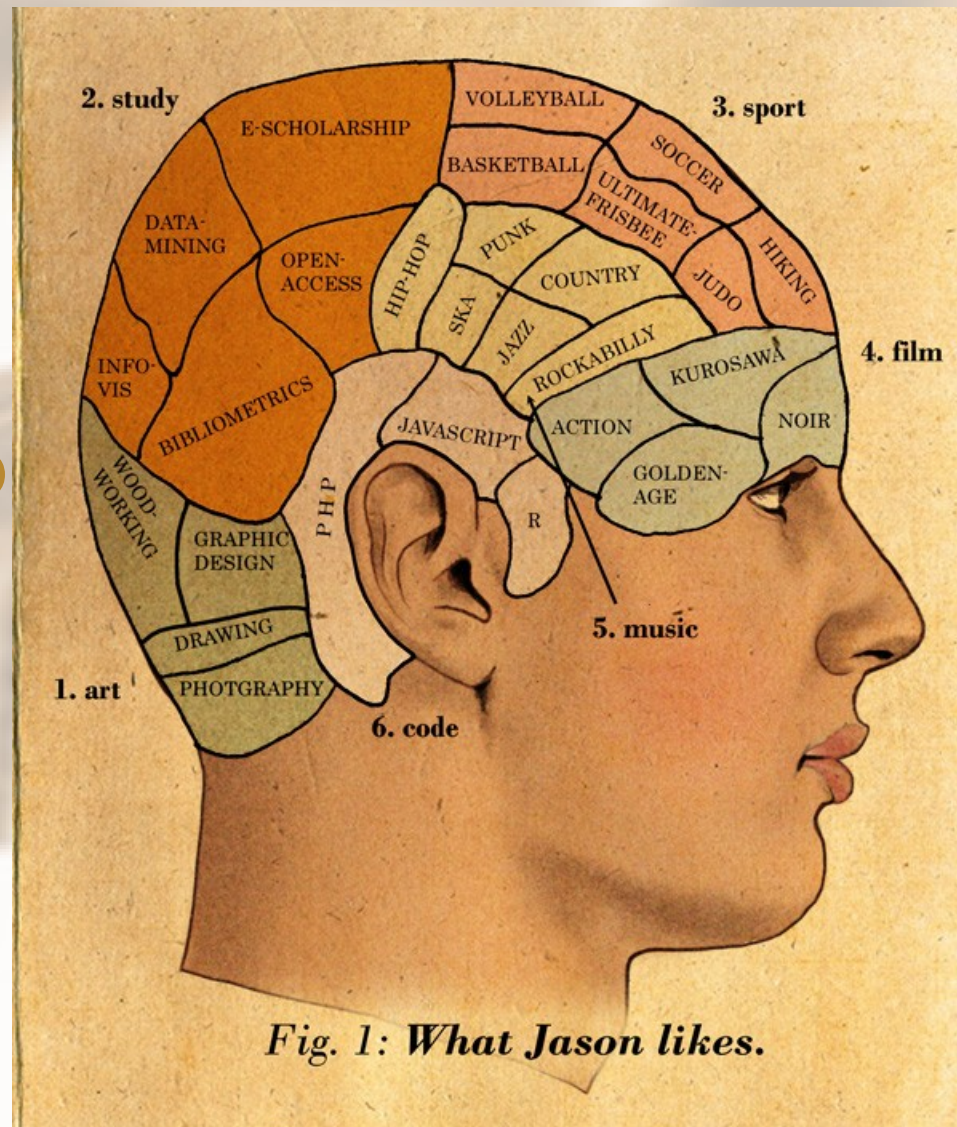
Second Edition

Max Velmans

Moral Agency

b) P

tion



Moral Agency

MATTER and CONSCIOUSNESS

REVISED EDITION

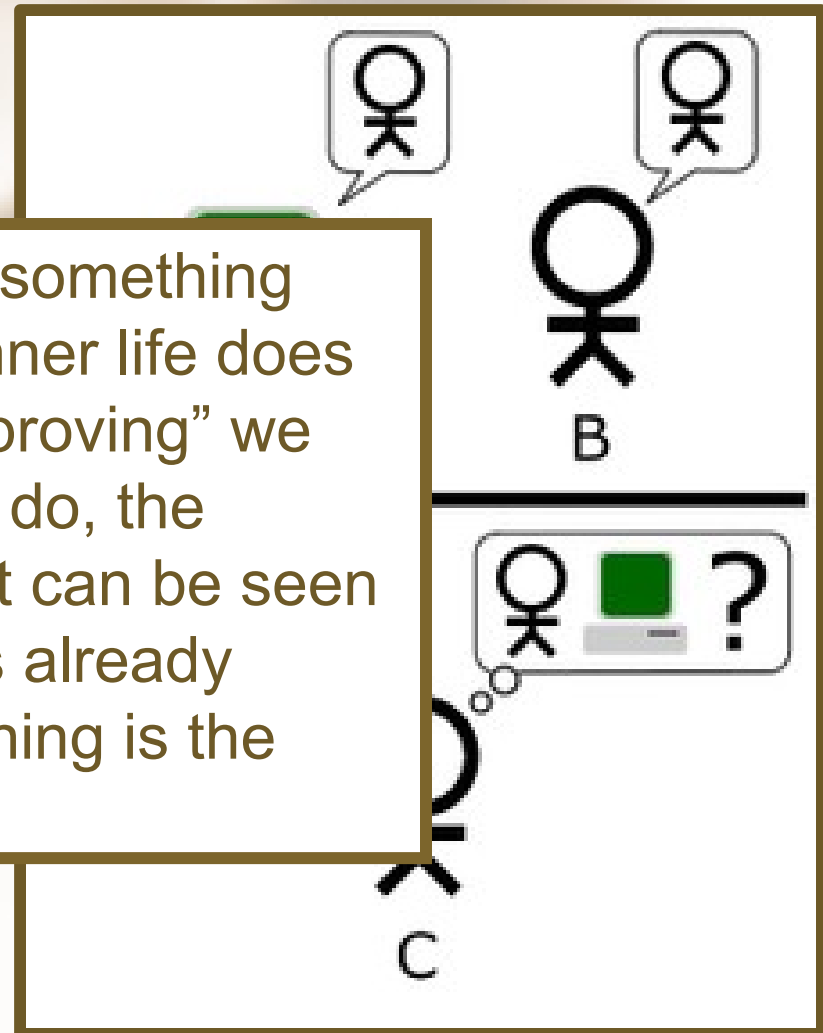
Paul M. Churchland



How does one determine whether something other than oneself—an alien creature, a sophisticated robot, a socially active computer, or even another human—is really a thinking, feeling, conscious being; rather than, for example, an unconscious automaton whose behavior arises from something other than genuine mental states? – Churchland, 1999

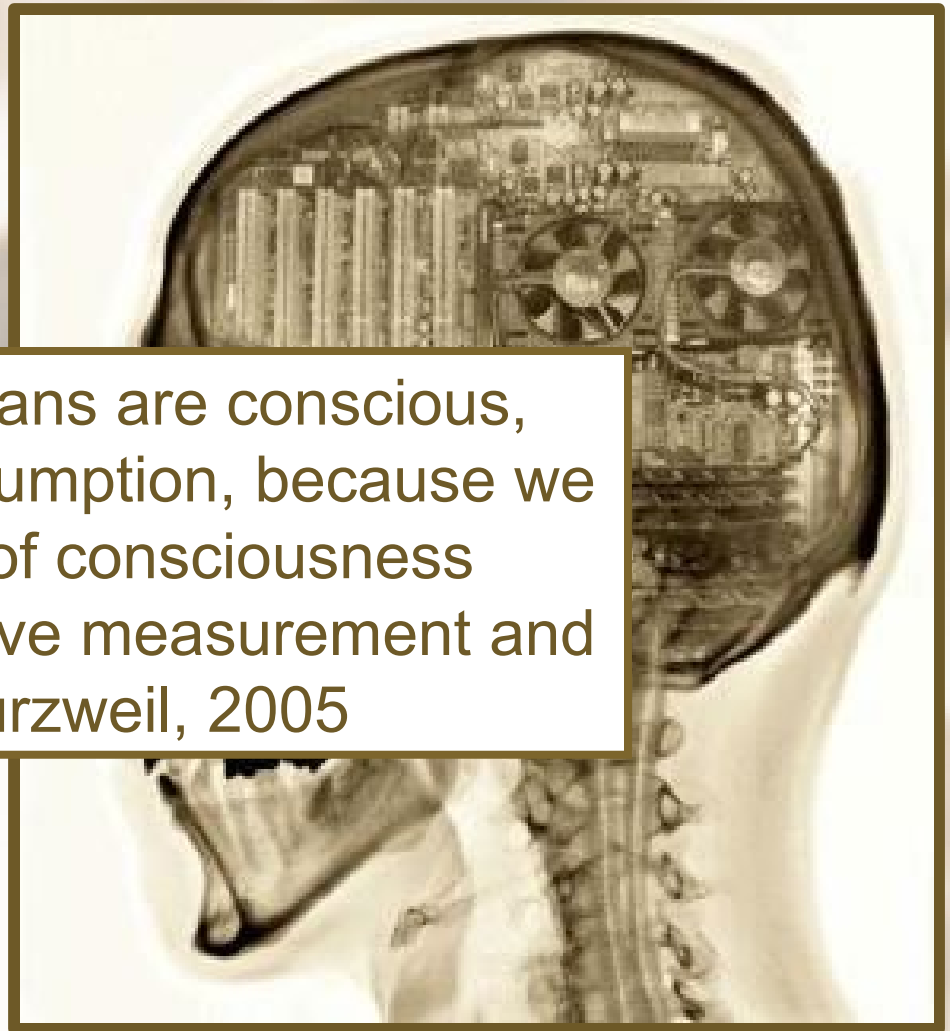
Moral Agency

There is no proving that something that seems to have an inner life does in fact have one—if by “proving” we understand, as we often do, the evincing of evidence that can be seen to establish by principles already agreed upon that something is the case. – Dennett, 1998

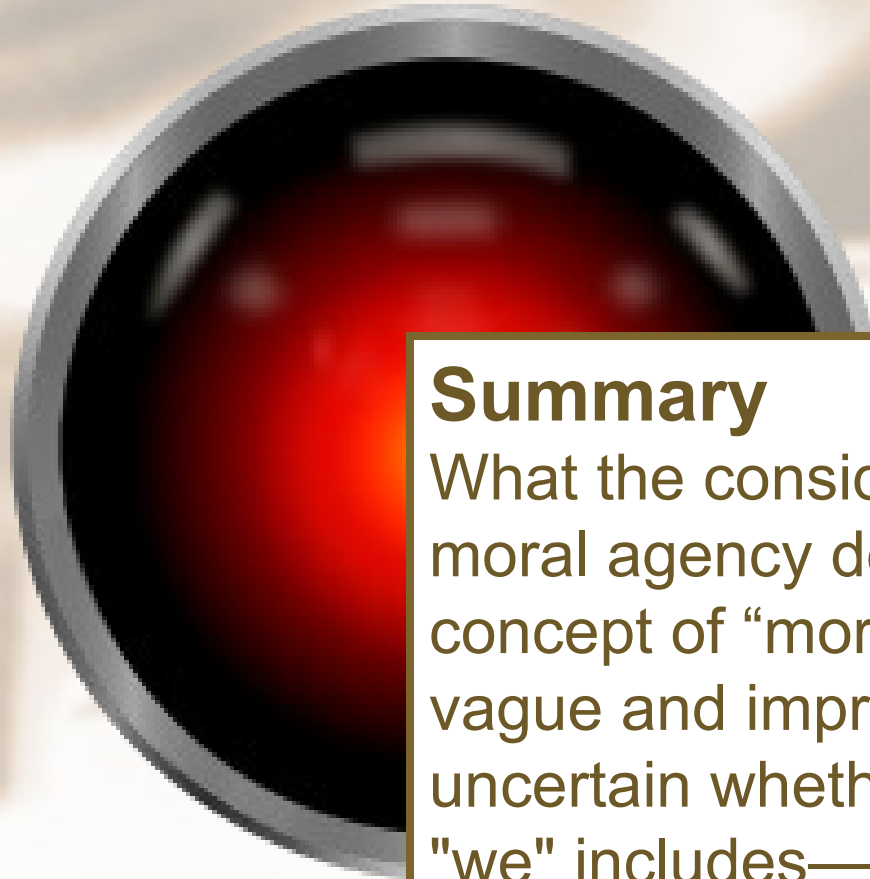


Moral Agency

We assume other humans are conscious, but even that is an assumption, because we cannot resolve issues of consciousness entirely through objective measurement and analysis (science) – Kurzweil, 2005



Moral Agency



Summary

What the consideration of machine moral agency demonstrates is that the concept of “moral agency” is already vague and imprecise such that it is uncertain whether we—whoever this “we” includes—are in fact moral agents.

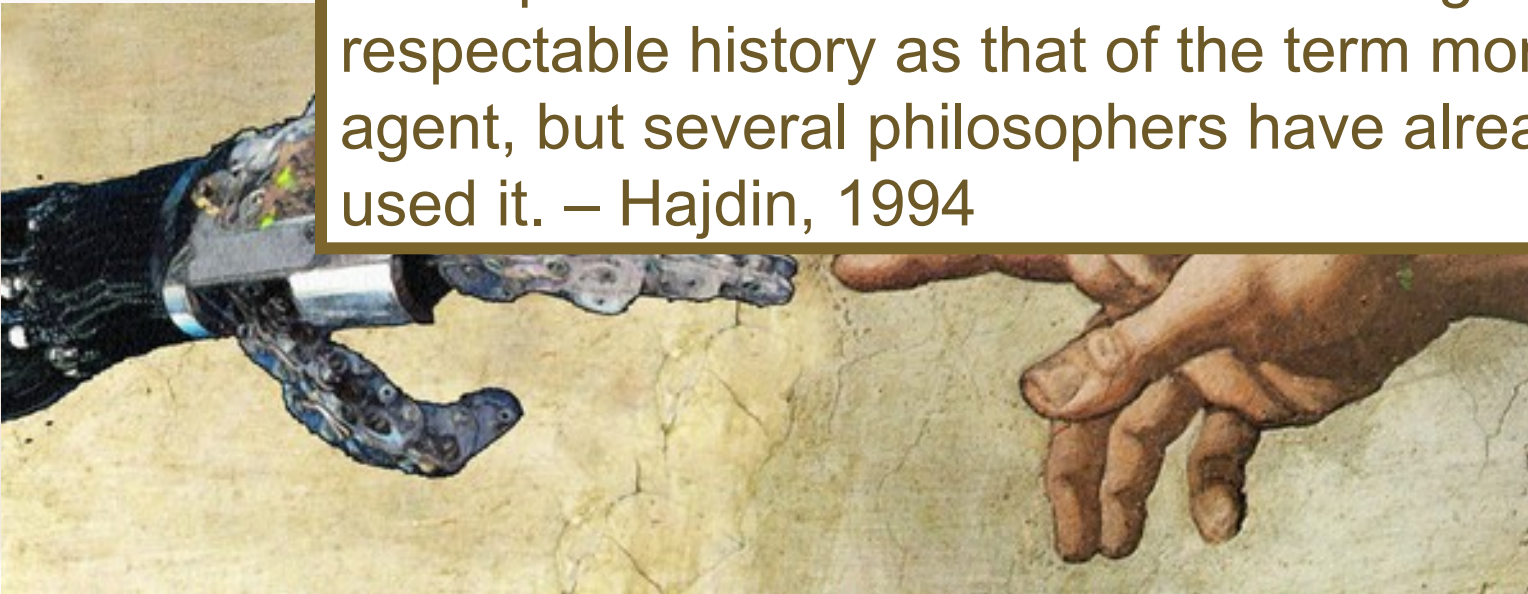
The background of the slide features a blurred image of several interlocking metal gears. The gears are a light tan or beige color, and the image is out of focus, creating a sense of depth and mechanical complexity. The lighting is soft, highlighting the textures of the gear teeth.

2

Moral Patience

Moral Patiency

The term *moral patient* is coined by analogy with the term moral agent. The use of the term moral patient does not have such a long and respectable history as that of the term moral agent, but several philosophers have already used it. – Hajdin, 1994



“Can they suffer?”

Moral Patience



Moral Patiency

Animal Liberalism



PETER

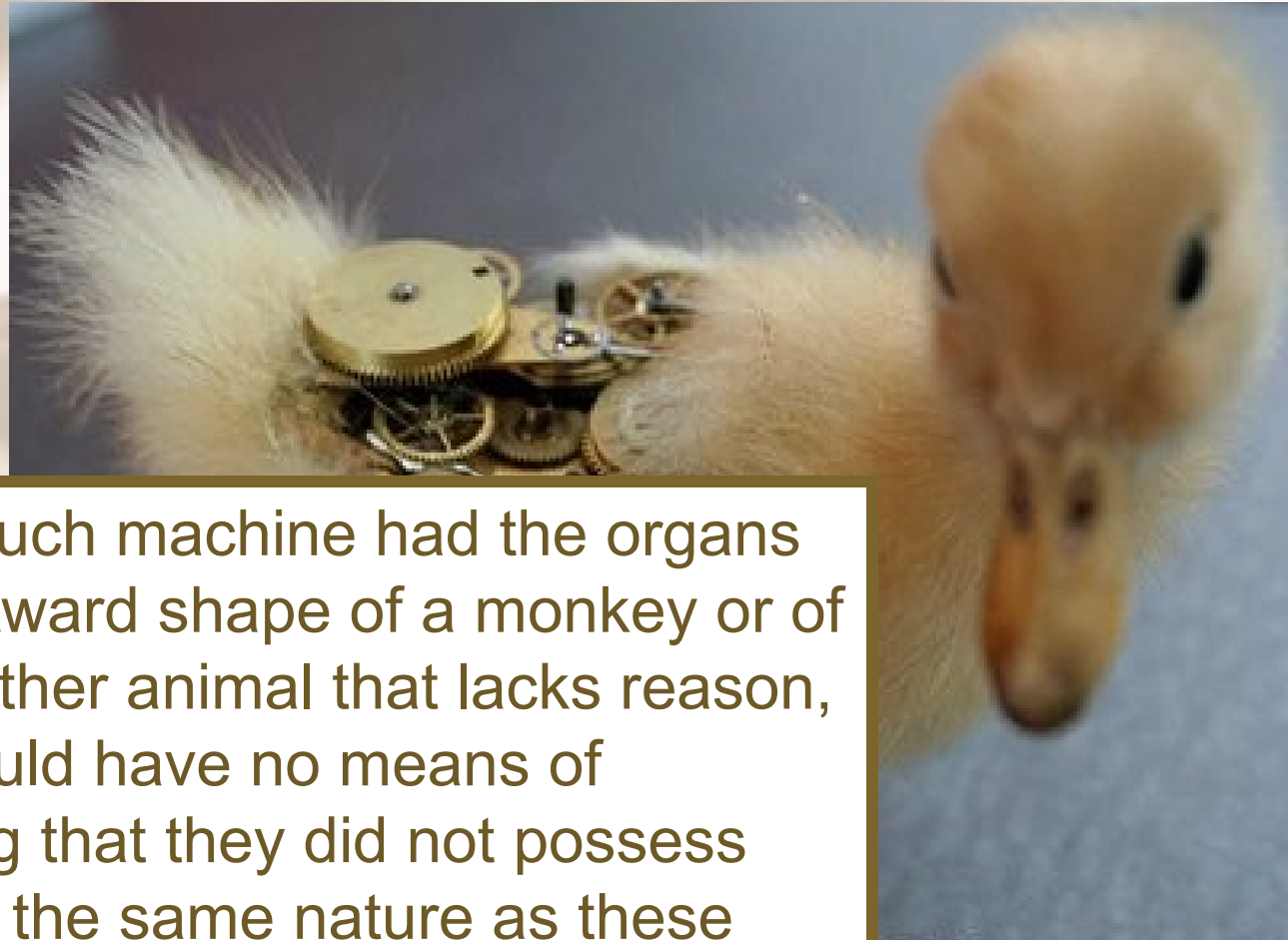
If a being suffers there can be no moral justification for refusing to take that suffering into consideration. No matter what the nature of the being, the principle of equality requires that its suffering be counted equally with the like suffering of any other being – Singer, 1975

"A most important book that will change the way many of us look at animals – and, ultimately, at ourselves."

– *Chicago Tribune*

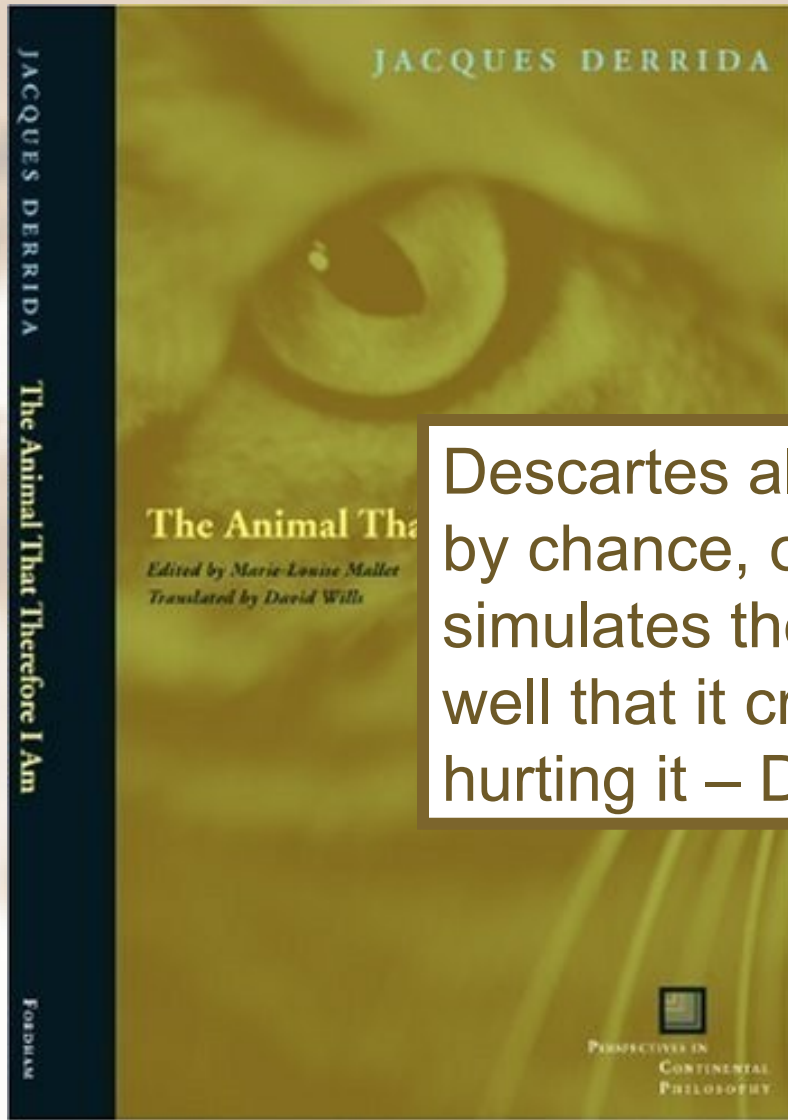
Moral Patiency

bête-machine



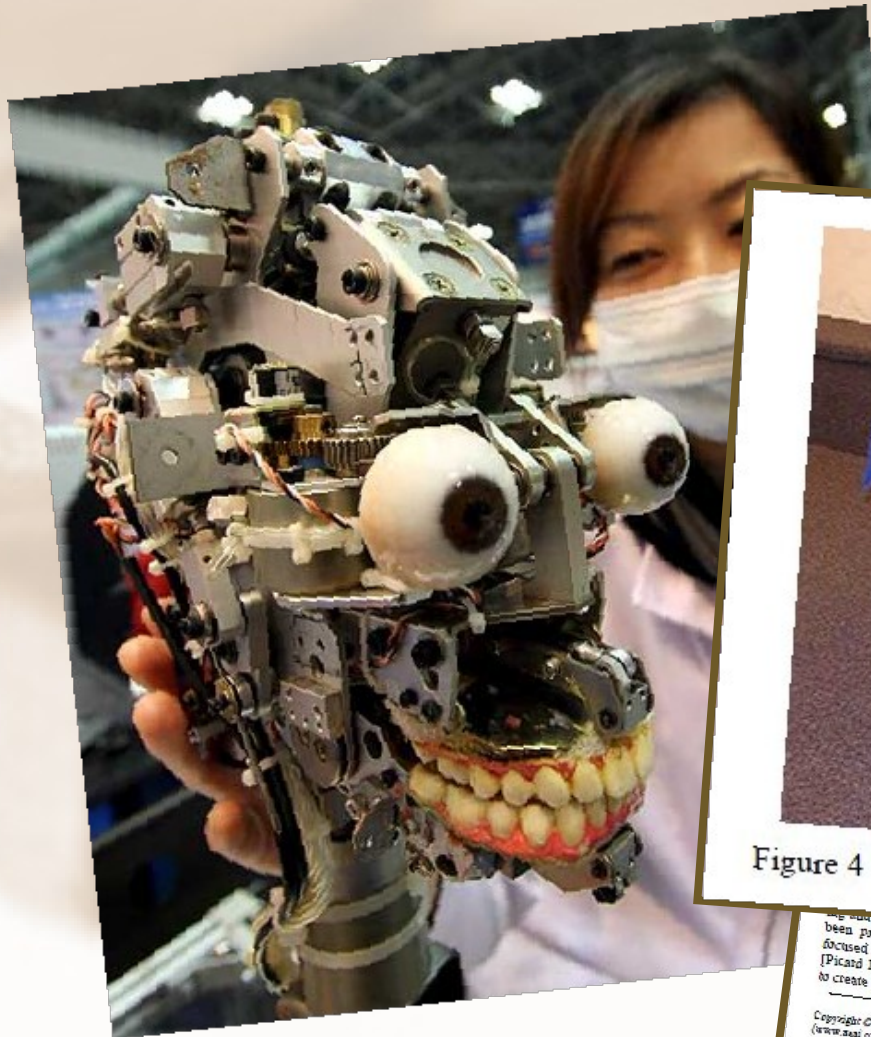
If any such machine had the organs and outward shape of a monkey or of some other animal that lacks reason, we should have no means of knowing that they did not possess entirely the same nature as these animals – Descartes, 1637

Moral Patiency



Descartes already spoke, as if by chance, of a machine that simulates the living animal so well that it cries out that you are hurting it – Derrida, 2006

Moral Patency



Simroid – Kokoro LLC

From: AAAI-98 Proceedings. Copyright © 1998, AAAI (www.aaai.org). All rights reserved.

When Robots Weep: Emotional Memories and Decision-Making

Juan D. Velásquez



Figure 4 Yuppy, an Emotional Pet Robot

... (Maes 1995; Reilly 1996), or synthesizing some of their influences in behavior and 1996; Kitano 1995; Pfeifer 1988; Velásquez 1997). Our main contribution is to describe in this paper derives from, and research on computational models of emotions, and behaviors can be integrated architecture, that uses some of the mechanisms to acquire memories from past emotions that serve as biasing mechanisms while during the action-selection process.

as Biasing Mechanisms

... with lesions in the prefrontal cortex motivated Damasio and colleagues to examine reasoning and decision-making mechanisms at different levels, extending from basic body regulation, to those that have control of complex strategies. An important component of this view is that reasoning and the feelings accompanying images that relate to the state of the world.

... part of this process includes the unconscious biasing mechanism that leads to a "right" decision. This biasing step is the marker hypothesis. The main idea is that decisions that are made in circumstances that are potentially harmful, or potentially a somatic response used to mark the importance of an event, and to signal their future outcome it serves as an alarm.

... Instead, a positive somatic marker is linked, it becomes an incentive to make that particular choice. These ideas inspired the model described below.

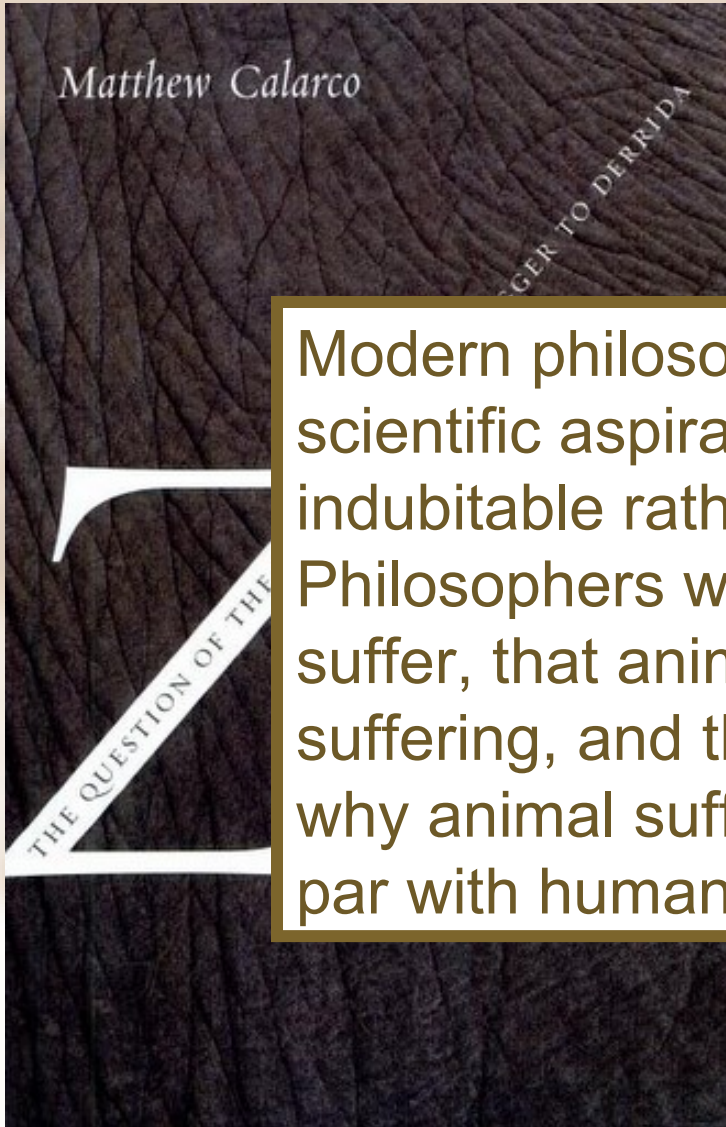
The Computational Model

This section describes *Cathexis*, a computational model of emotions and action-selection inspired by work in different fields, including Neuropsychology, Artificial Intelligence,

... and decision making. Several models of emotions have been proposed, but most of the work in this area has focused on specific aspects, such as recognizing emotions (Picard 1997), synthesizing emotions as the primary means to create believable synthetic agents (Bates 1994; Blumberg 1995).

Copyright © 1998, American Association for Artificial Intelligence (www.aaai.org). All rights reserved.

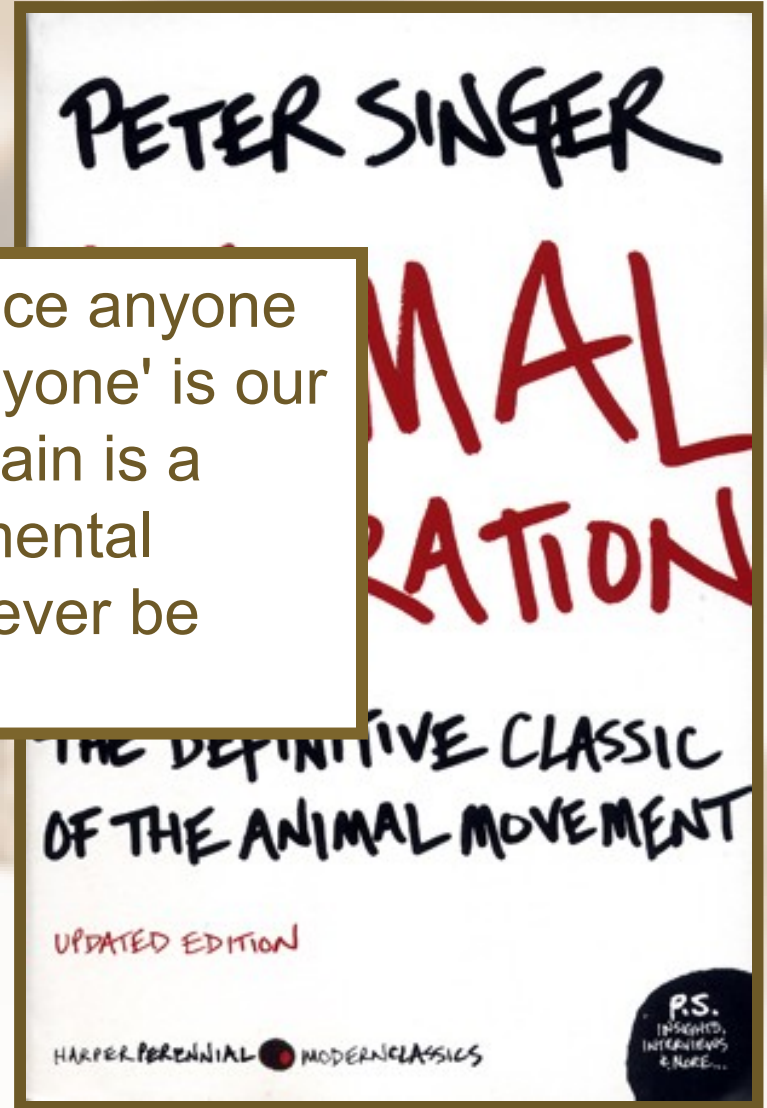
Moral Patiency



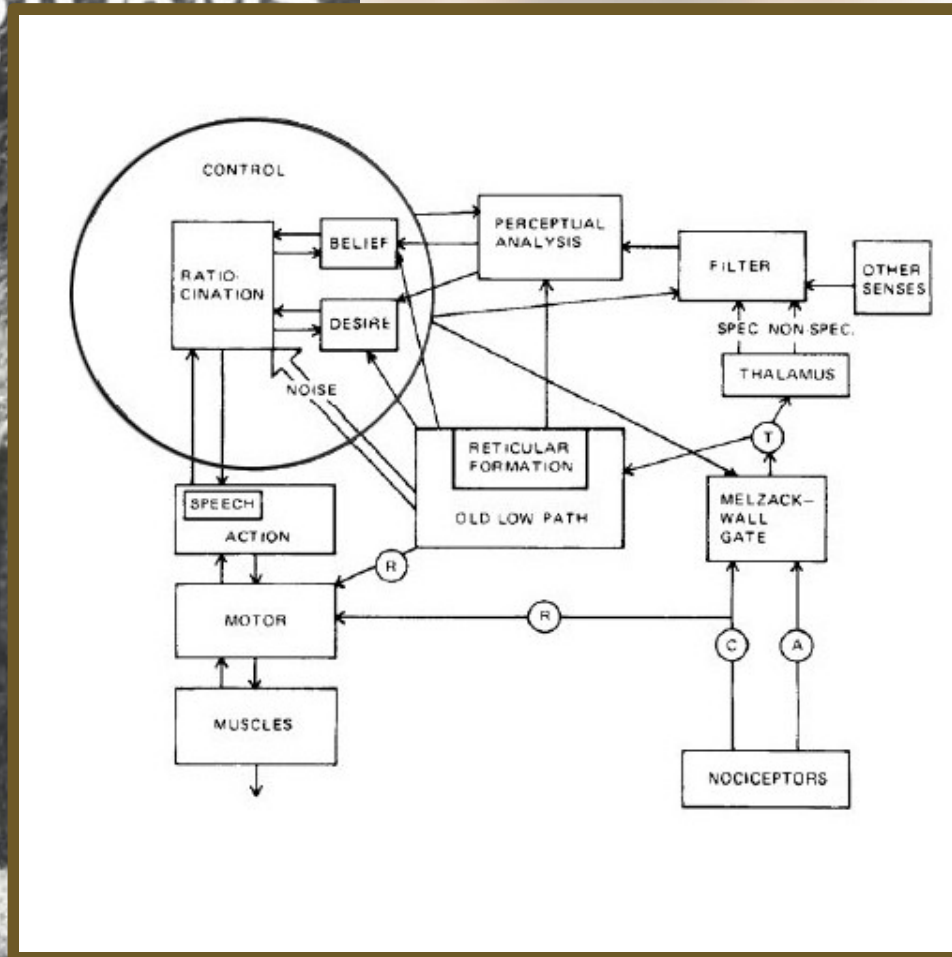
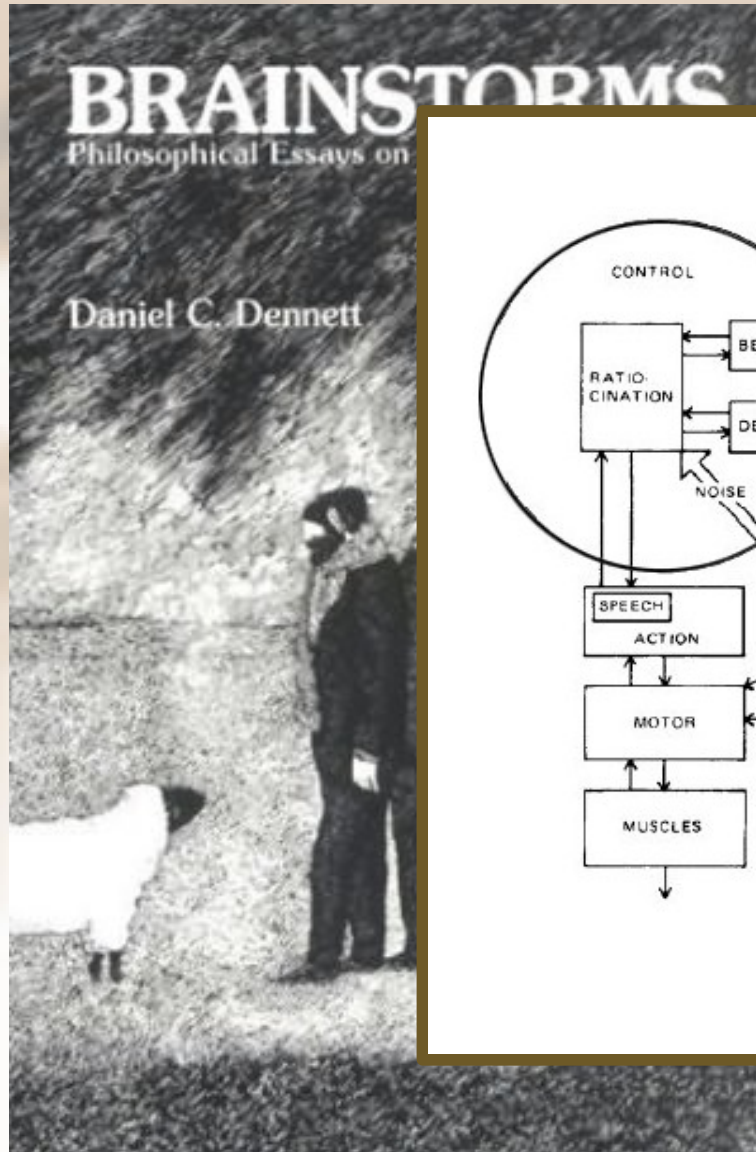
Modern philosophy true to its Cartesian and scientific aspirations, is interested in the indubitable rather than the undeniable. Philosophers want proof that animals actually suffer, that animals are aware of their suffering, and they require an argument for why animal suffering should count on equal par with human suffering. – Calarco, 2008

Moral Patiency

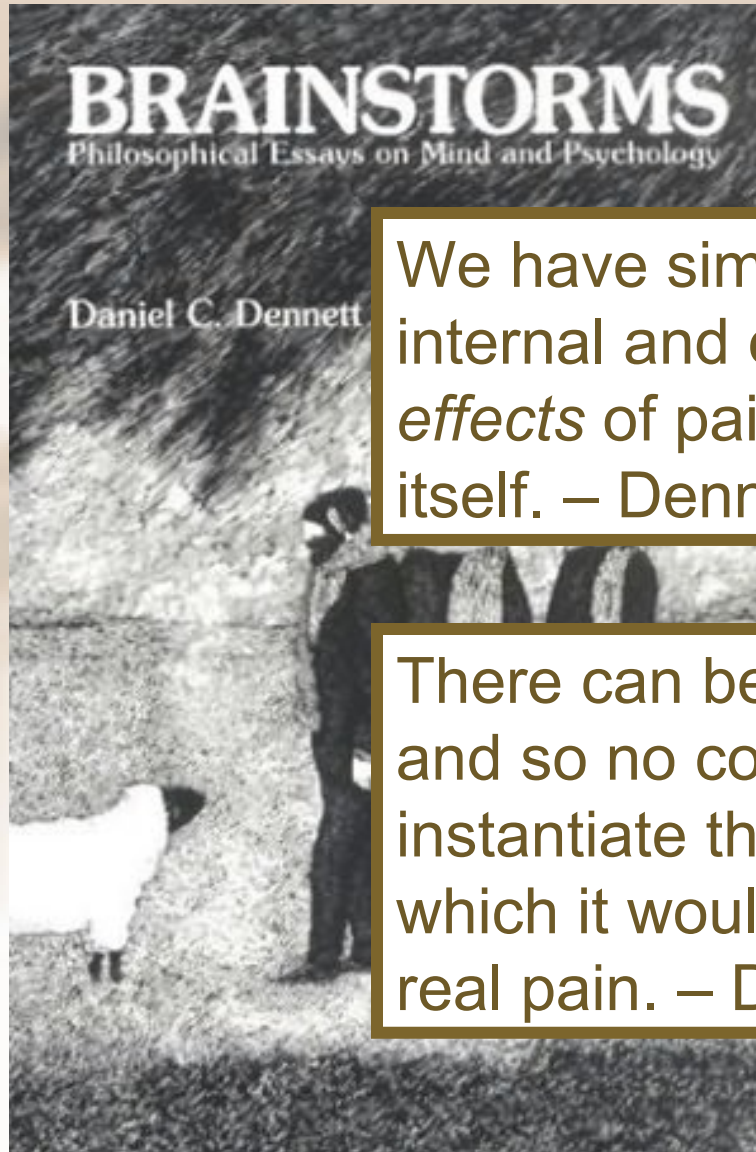
We cannot directly experience anyone else's pain, whether that 'anyone' is our best friend or a stray dog. Pain is a state of consciousness, a 'mental event,' and as such it can never be observed. – Singer, 1975



Moral Patiency



Moral Patiency



We have simulated, perhaps, the internal and external *causes and effects* of pain, but not the pain itself. – Dennett, 1998

There can be no true theory of pain, and so no computer or robot could instantiate the true theory of pain, which it would have to do to feel real pain. – Dennett, 1998

Moral Patiency

Moral Machines

Teaching Robots Right from Wrong

If (ro)bots might one day be capable of experiencing pain and other affective states a question that arises is whether it will be moral to build such systems—not because of how they might harm humans, but because of the pain these artificial systems will themselves experience. In other words, can the building of a (ro)bot with a somatic architecture capable of feeling intense pain be morally justified? – Wallach & Allen, 2009

Wendell Wallach • Colin Allen

Moral Patency



Moral Patiency

Summary

Extending the patient oriented approach modeled by animal rights philosophy does not provide definitive proof that the machine either can be or is not able to be a similarly constructed moral patient. Instead doing so demonstrates how this innovation in moral thinking might already be misguided and prejudicial.

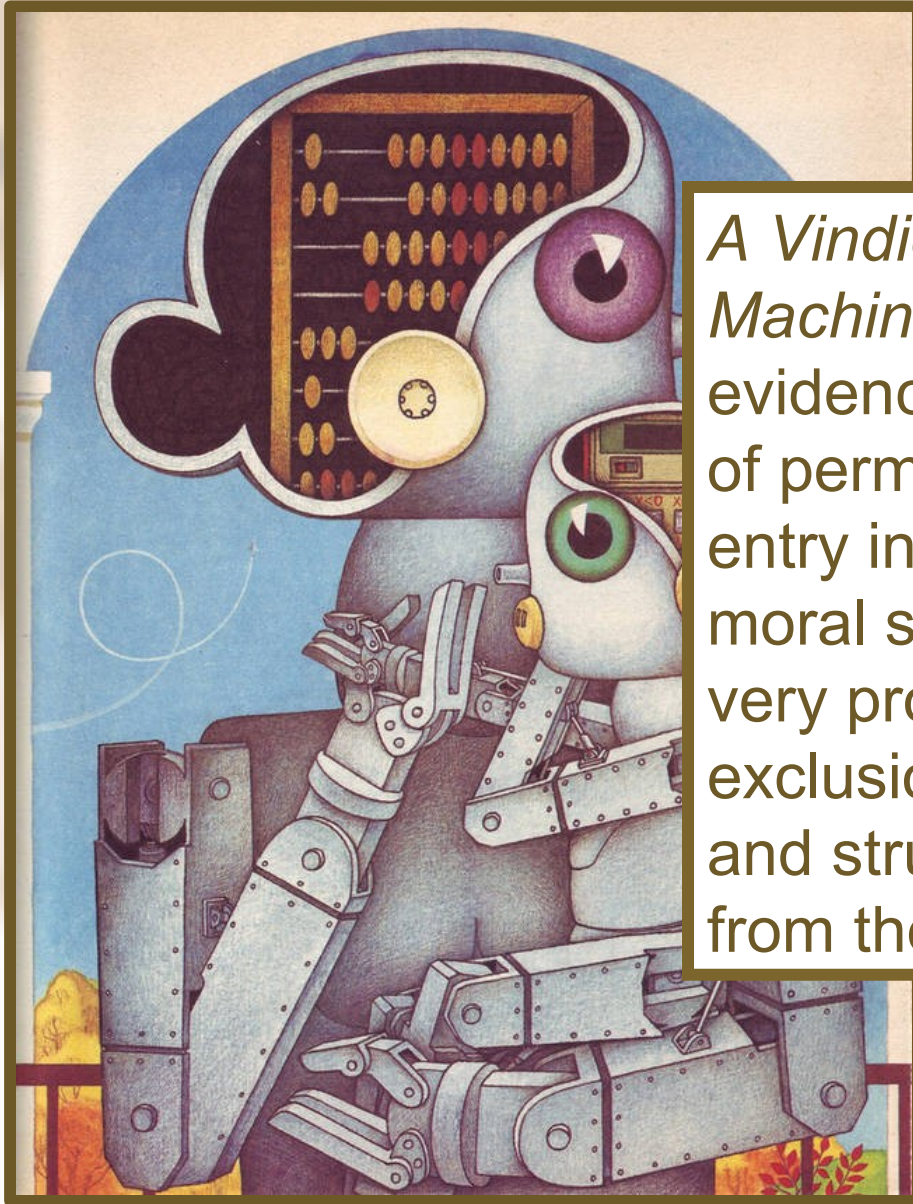


The background of the slide is a blurred image of several interlocking metal gears. The gears are a light brown or tan color, and the image is out of focus, creating a bokeh effect. The lighting is soft and even.

3

Thinking Otherwise

Thinking Otherwise



A Vindication of the Rights of Machines does accumulate evidence or arguments in favor of permitting one more entity entry into the community of moral subjects. It questions the very protocols of inclusion and exclusion that have organized and structured moral philosophy from the beginning.

Thinking Otherwise



DISCOURS DE LA METHODE

Pour bien conduire sa raison, & chercher
la verité dans les sciences.

Plus
LA DIOPTRIQUE.
LES METEORES.

ET
LA GEOMETRIE.

Qui sont des essais de cete METHODE.



A LEYDE
De l'Imprimerie de IAN MAIRE.

CICIDCXXXVII.

Avec Privilege.

Thinking Otherwise

Now, before starting to rebuild your house, it is not enough simply to pull it down, to make provision for materials and architects (or else train yourself in architecture), and to have carefully drawn up the plans; you must also provide yourself with some other place where you can live comfortably while building is in progress. Likewise, lest I should remain indecisive in my actions while reason obliged me to be so in my judgments, and in order to live as happily as I could during this time, I formed for myself a **provisional moral code** consisting of just three or four maxims.

DISCOURS DE LA METHODE

re la raifon,& chercher
ans les ſciences.

us
P TRIQUE.

TEORES.

T
METRIE.

de cete METHODE.



EYDE
de L'AN MAIRE.

C XXXVII.
Privilege.

Thinking Otherwise



Moral consideration is no longer seen as being 'intrinsic' to the entity: instead it is seen as something that is 'extrinsic': it is attributed to entities within social relations and within a social context. – Coeckelbergh, 2010

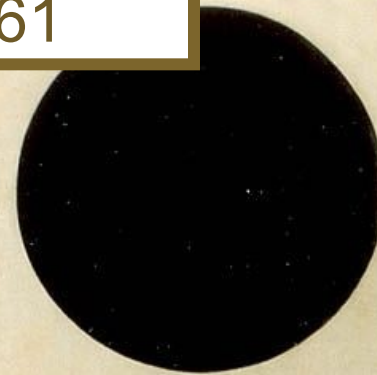
Thinking Otherwise

Emmanuel Lévinas

Totalité et infini

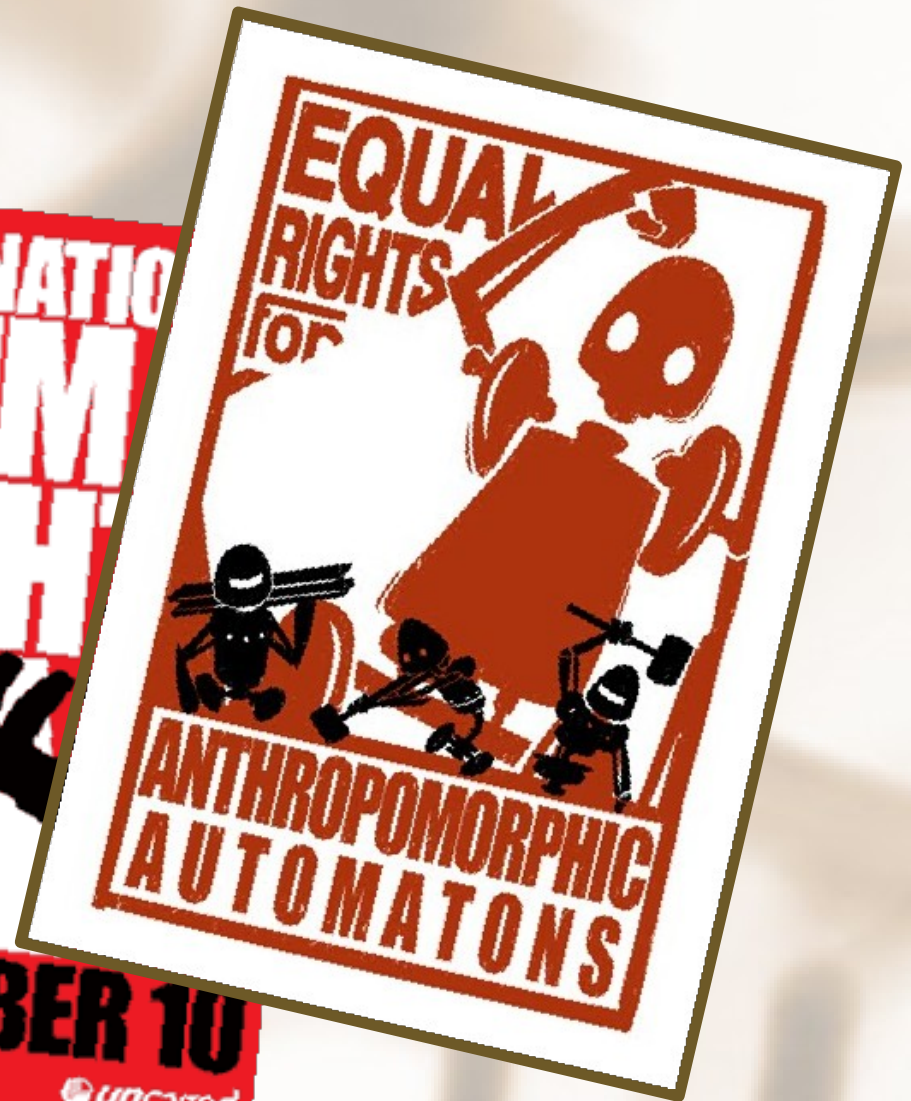
Essai sur l'extériorité

Morality is not a branch of philosophy,
but first philosophy – Levinas, 1961



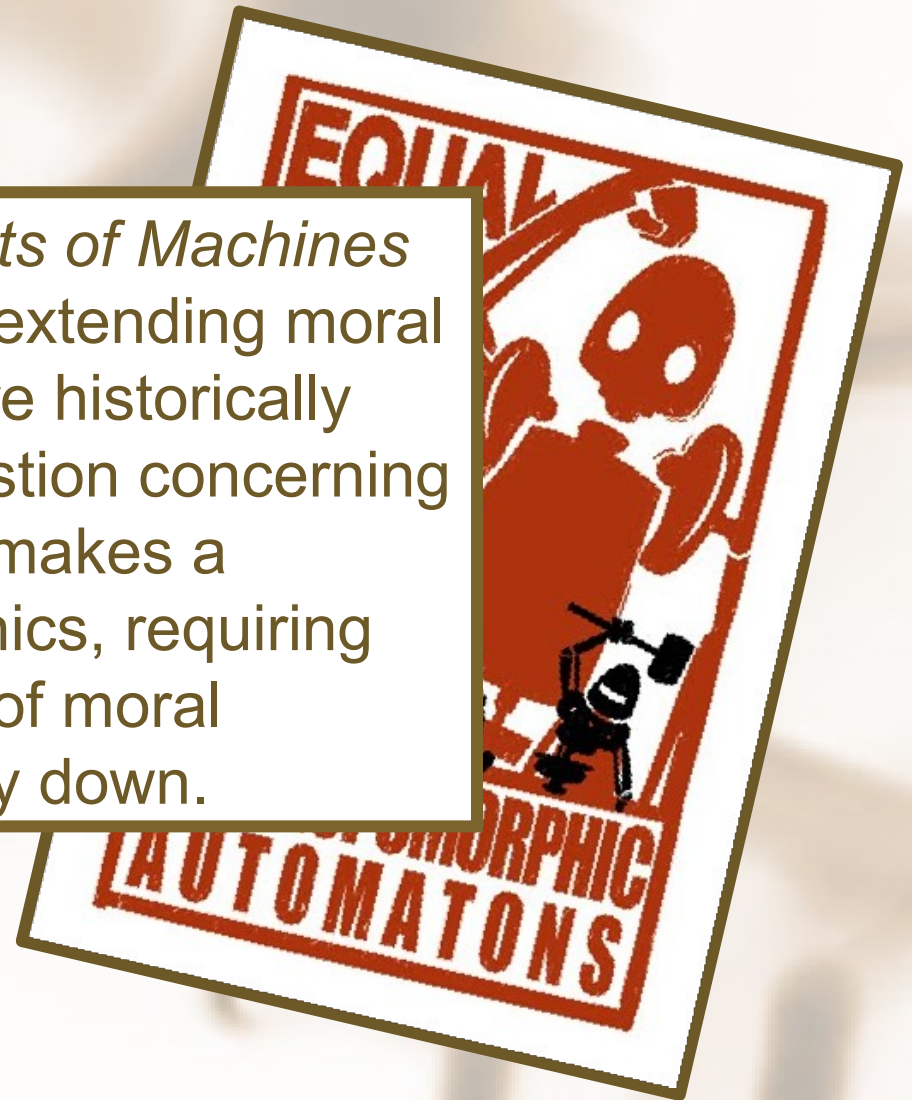
biblio  *essais*

Thinking Otherwise



Thinking Otherwise

A Vindication of the Rights of Machines is not simply a matter of extending moral consideration to one more historically excluded other. The question concerning the "rights of machines" makes a fundamental claim on ethics, requiring us to rethink the system of moral considerability all the way down.



THE MACHINE QUESTION

CRITICAL PERSPECTIVES ON AI,
ROBOTS, AND ETHICS

DAVID J. GUNKEL



David J. Gunkel
dgunkel@niu.edu
<http://machinequestion.org>