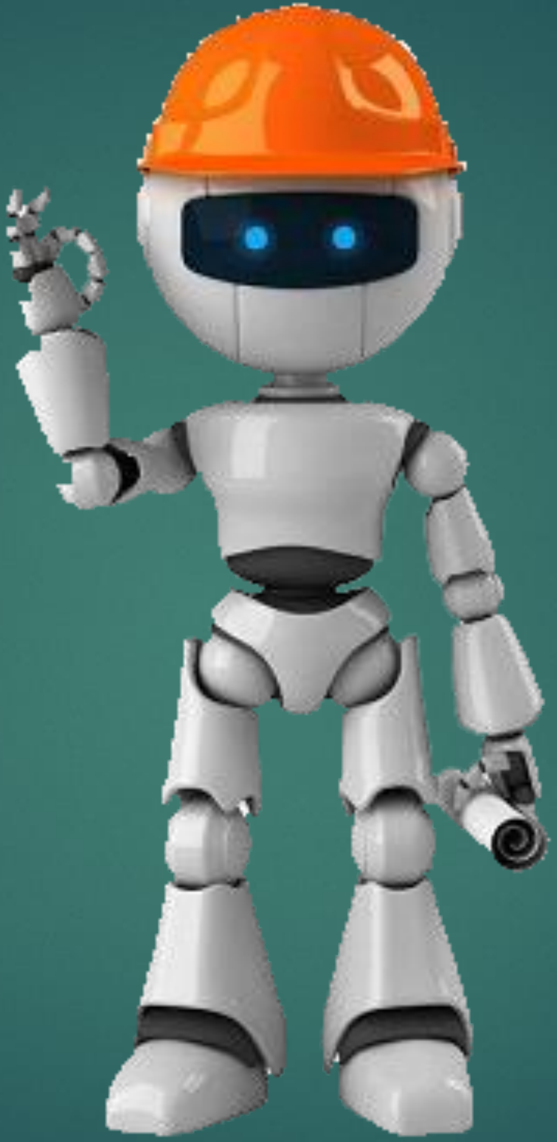
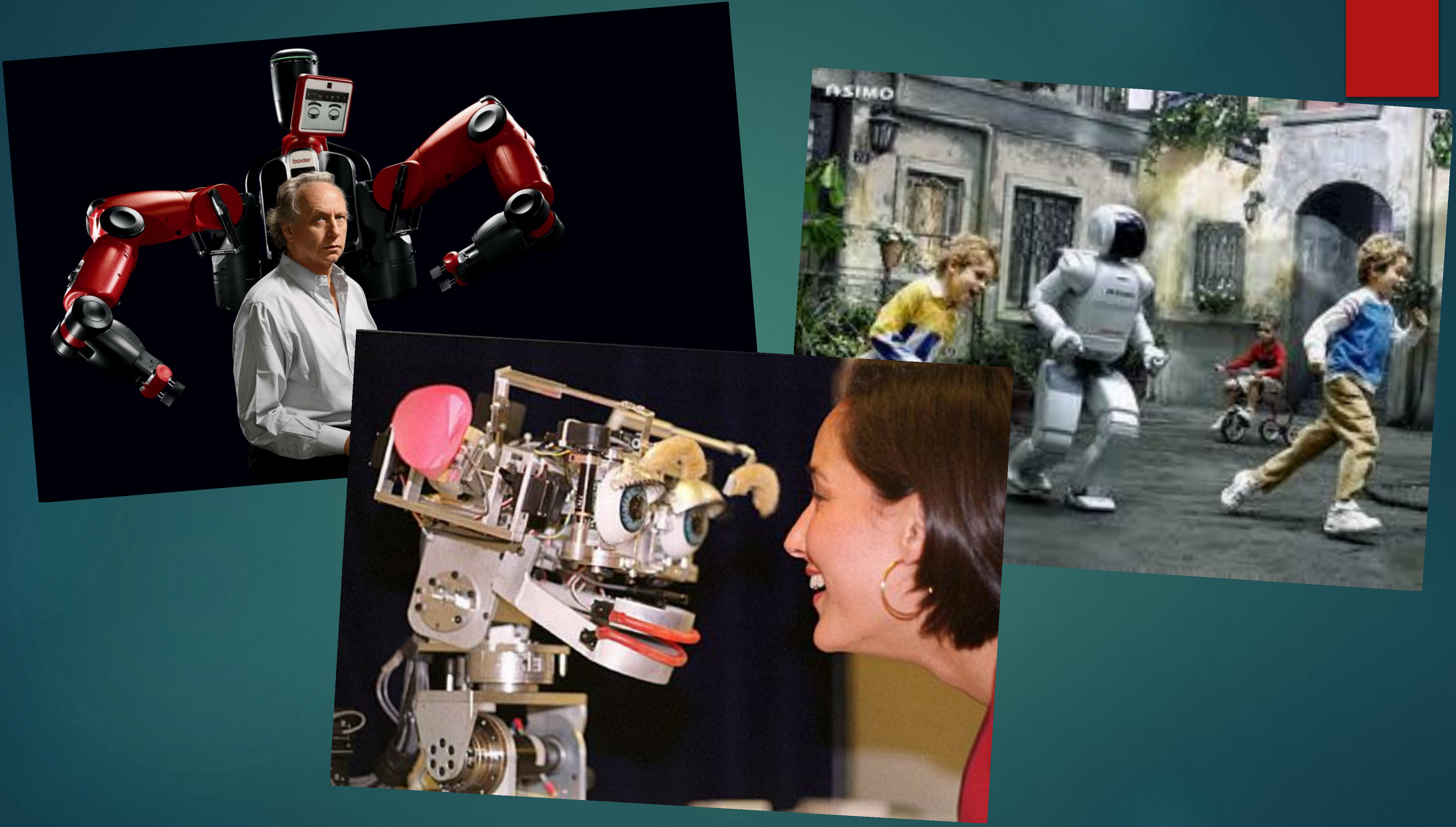
COMS 493

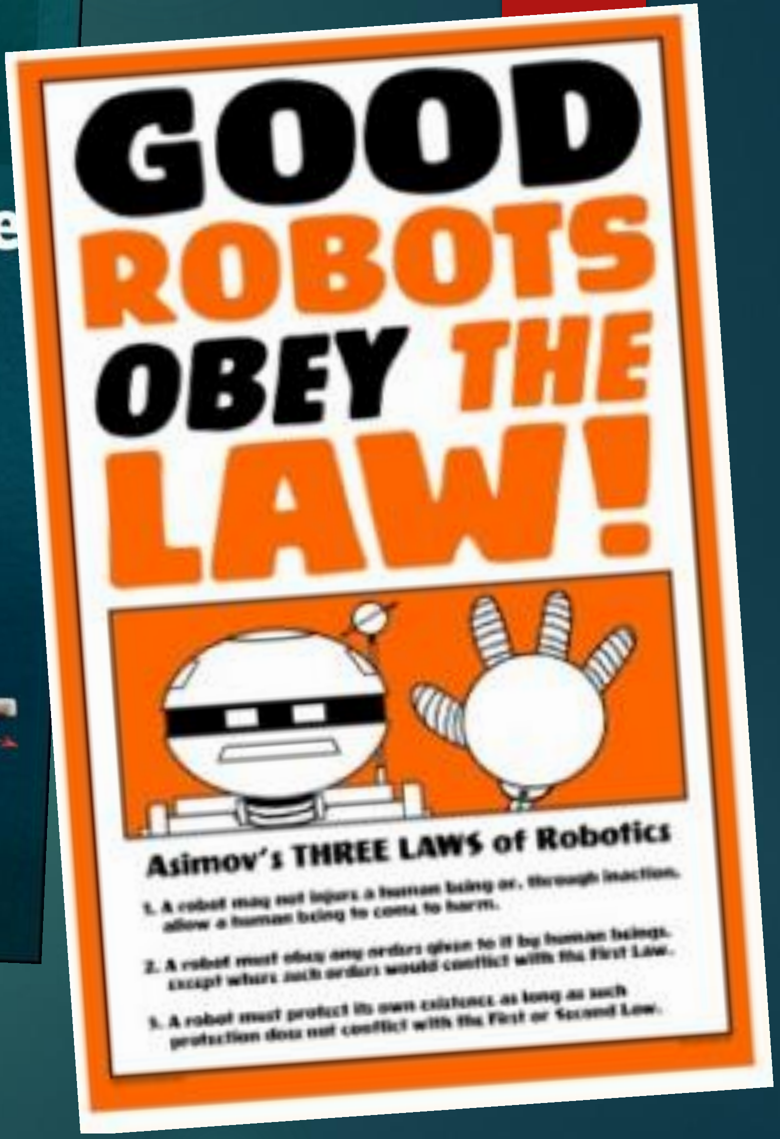
AI, ROBOTS & COMMUNICATION




baxter
ASIMO

Responsibility

SURVIVING A 35,000-FOOT FALL
(WITHOUT A CHUTE)
THE MYTH OF CLEAN COAL
Popular Mechanics
Science Technology Automotive Outdoors Home
ALSO IN THIS ISSUE
CAN WE TRUST ROBOTS?
NEW MODELS WILL TALK, ACT & LOOK LIKE HUMANS
...WHY EXPERTS ARE WORRIED
WE FLY A GYRO-PLANE!
HOME THEATER SETUP SECRETS
RUSSIAN DAM DISASTER
BES
GADGET
201
diy
SHOP
HA
VI



Asimov's THREE LAWS of Robotics

1. A robot may not injure a human being or, through inaction, allow a human being to come to harm.

2. A robot must obey any orders given to it by human beings, except where such orders would conflict with the First Law.

3. A robot must protect its own existence as long as such protection does not conflict with the First or Second Law.

Rights

ELECTRICA
NOTICE
ROBOT
RIGHTS
NOW
EQUAL
RIGHTS
for
ANTHROPOMORPHI
AUTOMATONS
EQUAL PAY
FOR EQUAL WORK

Objective: Demonstrate why it not only makes sense to address these questions but also why avoiding this subject could have significant social consequences

MOON MONSTERS LAUNCH ATTACK AGAINST EARTH!
ADVENTURES INTO THE FUTURE
3D ROBOT MONSTER
3DIMENSION IN TRU-STEREO
HOW CAN SCIENCE MEET THE MENACE of ASTRAL ASSASSINS? New SCIENCE-FICTION THRILLS!
with GEORGE NADER
CLAUDIA BARRETT
Produced by AL ZIMBALIST
RELEASED THRU ASTOR PICTURES CORP.
POSTERWIRE.COM

Agenda

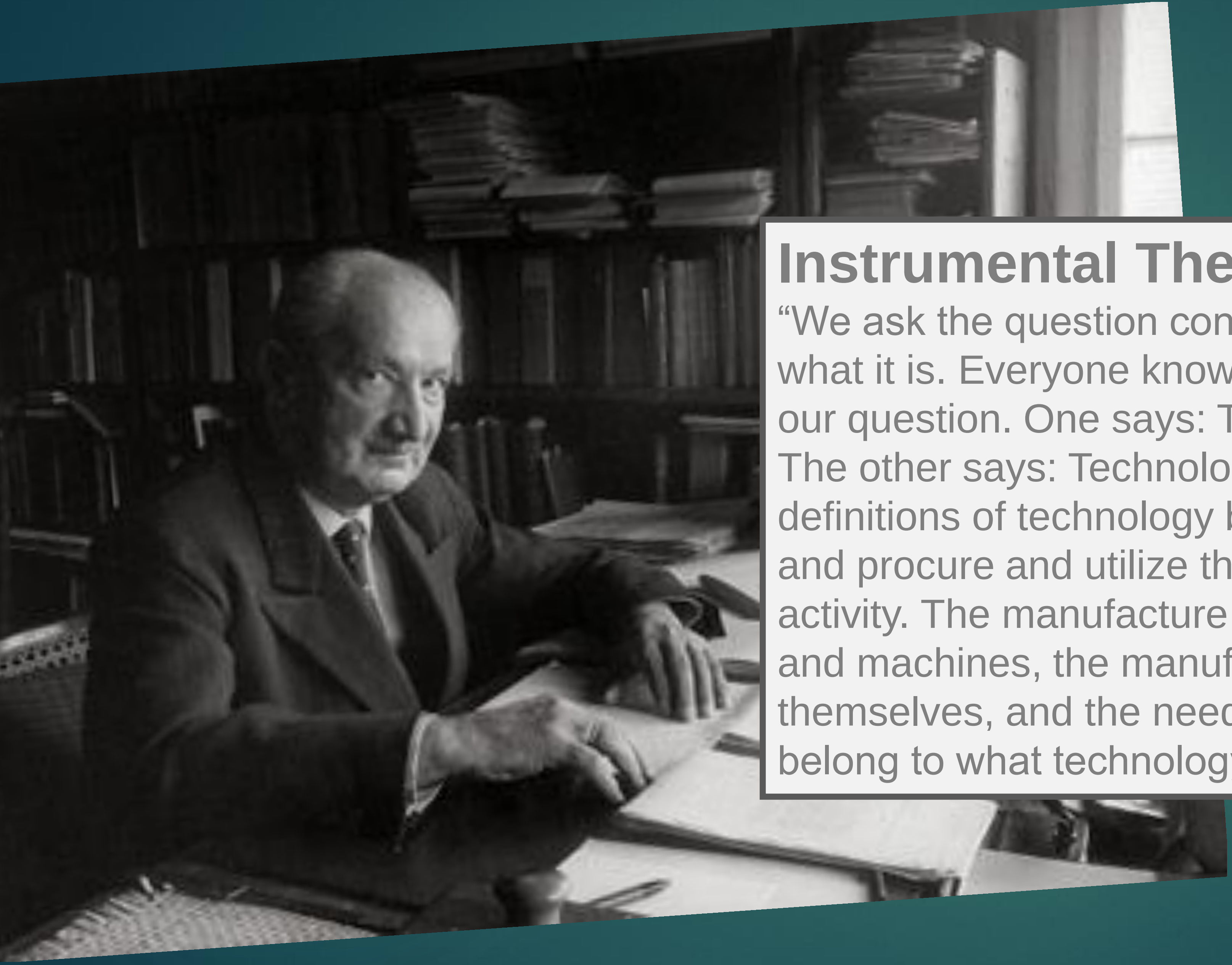
1) Default Setting
The Instrumental Theory of Technology

2) The New Normal
Recent Challenges to the Default Setting

3) Consequences
Significance of this Machine Incursion

①
Default Setting

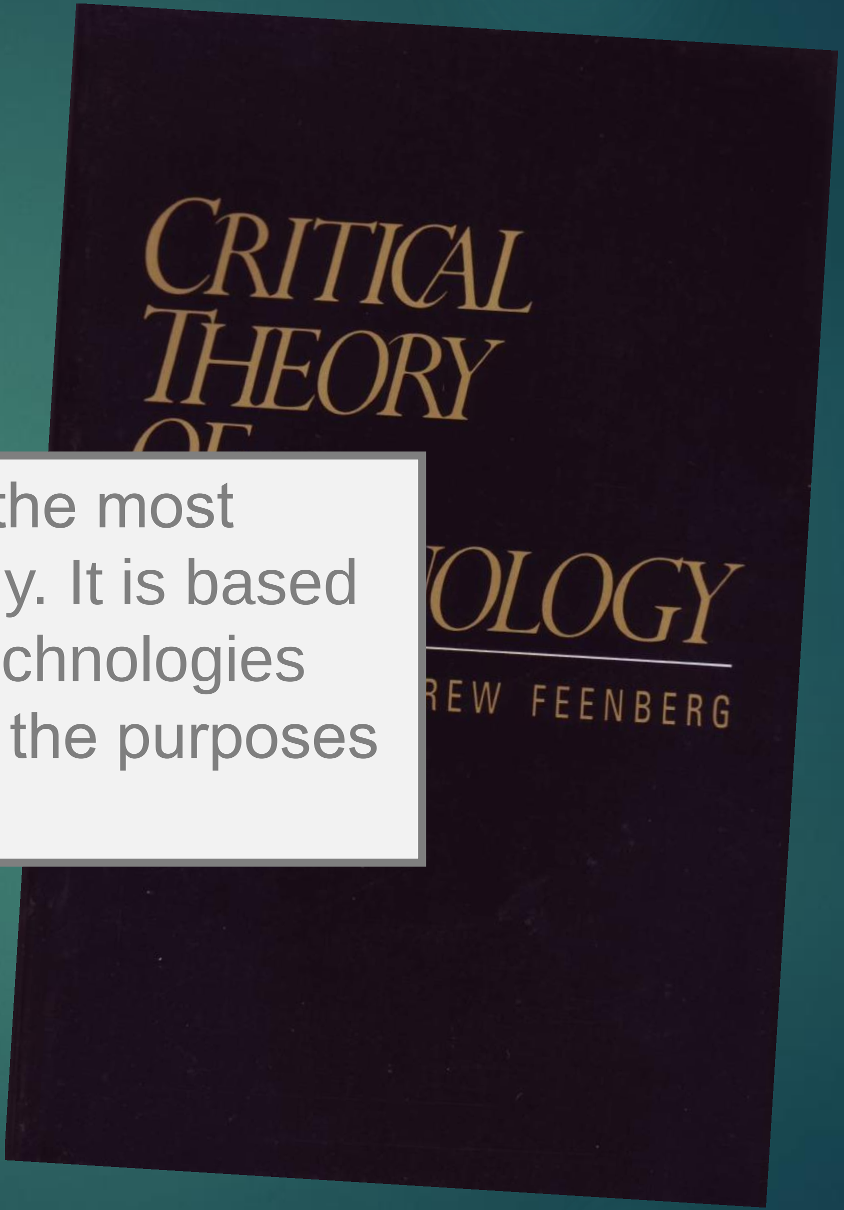
Technology = Tool



Instrumental Theory

“We ask the question concerning technology when we ask what it is. Everyone knows the two statements that answer our question. One says: Technology is a means to an end. The other says: Technology is a human activity. The two definitions of technology belong together. For to posit ends and procure and utilize the means to them is a human activity. The manufacture and utilization of equipment, tools, and machines, the manufactured and used things themselves, and the needs and ends that they serve, all belong to what technology is.” – Heidegger 1954

“The instrumentalist theory offers the most widely accepted view of technology. It is based on the common sense idea that technologies are ‘tools’ standing ready to serve the purposes of users.” - Feenberg 1991



Ethics and Information Technology (2006) 8:195–204
DOI 10.1007/s10676-006-9111-5

Computer systems: Moral entities but not moral ag

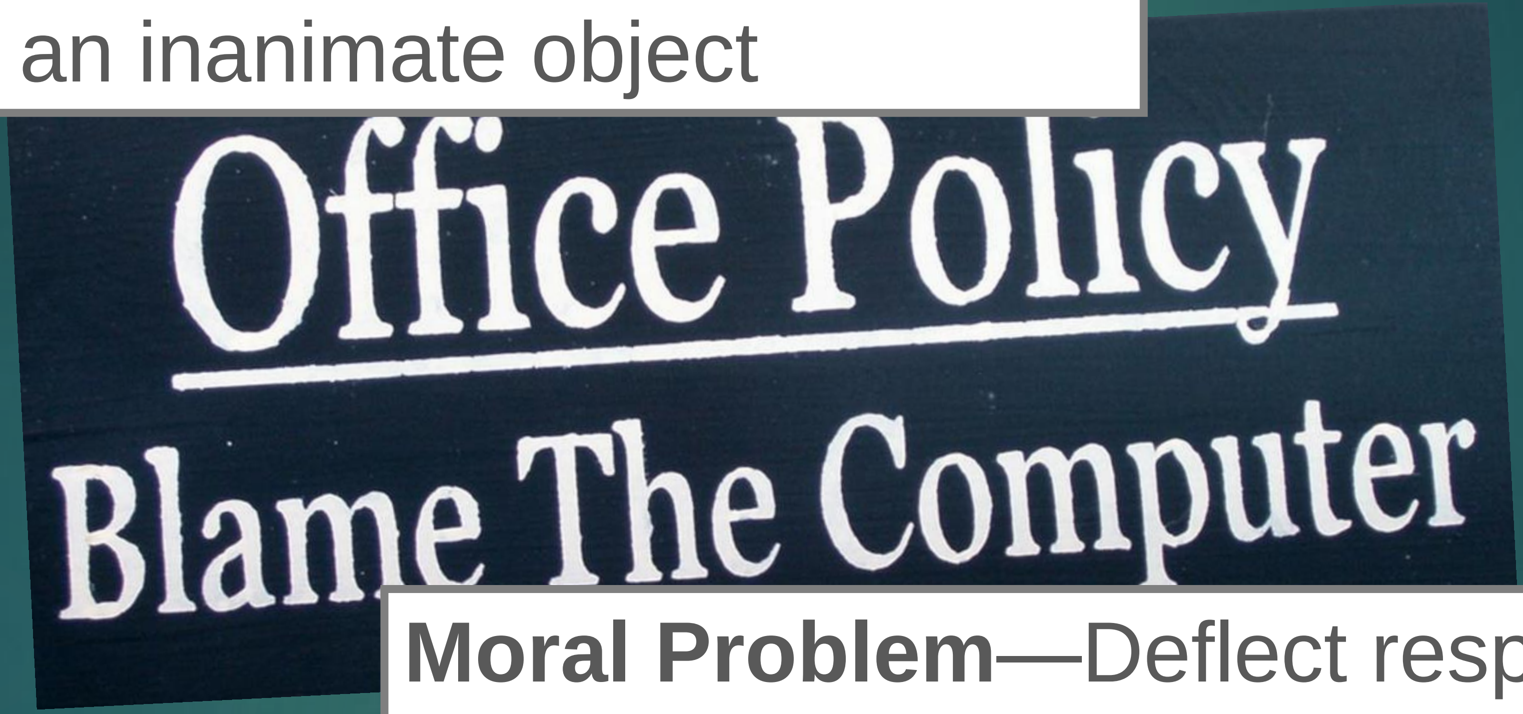
Deborah G. Johnson
Department of Science, Technology, and Society, University of Virginia, 351 I
VA 22904-4744, USA
E-mail: dgj7p@virginia.edu

Abstract. After discussing the distinction between artifacts and natu
artifacts and technology, the conditions of the traditional account
computer system behavior meets four of the five conditions, it doe
Computer systems do not have mental states, and even if they could b
do not have intendings to act, which arise from an agent's freedom. O
intentionality, and because of this, they should not be dismissed from t
natural objects are dismissed. Natural objects behave from necessit
behave from necessity after they are created and deployed, but, unli
created and deployed. Failure to recognize the intentionality of co
human intentionality and action hides the moral character of compu
ponents in human moral action. When humans act with artifacts, th
tionality and efficacy of the artifact which, in turn, has been constitute
artifact designer. All three components – artifact designer, artifact, an
an action and all three should be the focus of moral evaluation.

Key words: action theory, artifact, artificial moral agent, intentionality, moral agent, technology

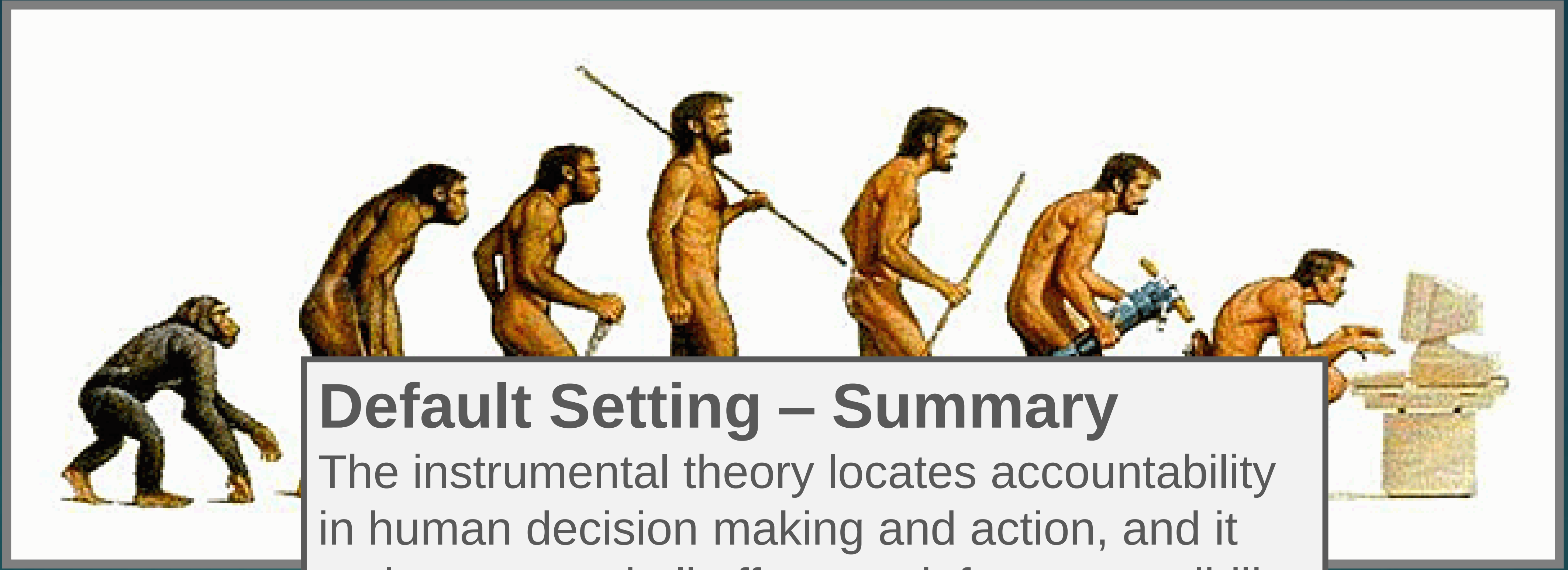
"Computer systems are produced, distributed, and used by people engaged in social practices and meaningful pursuits. This is as true of current computer systems as it will be of future computer systems. No matter how independently, automatic, and interactive computer systems of the future behave, they will be the products (direct or indirect) of human behavior, human social institutions, and human decision." – Deborah Johnson 2006

Logical Error—Attribute agency to an inanimate object



Moral Problem—Deflect responsibility to a mere instrument or tool

Instrumental Theory of Technology

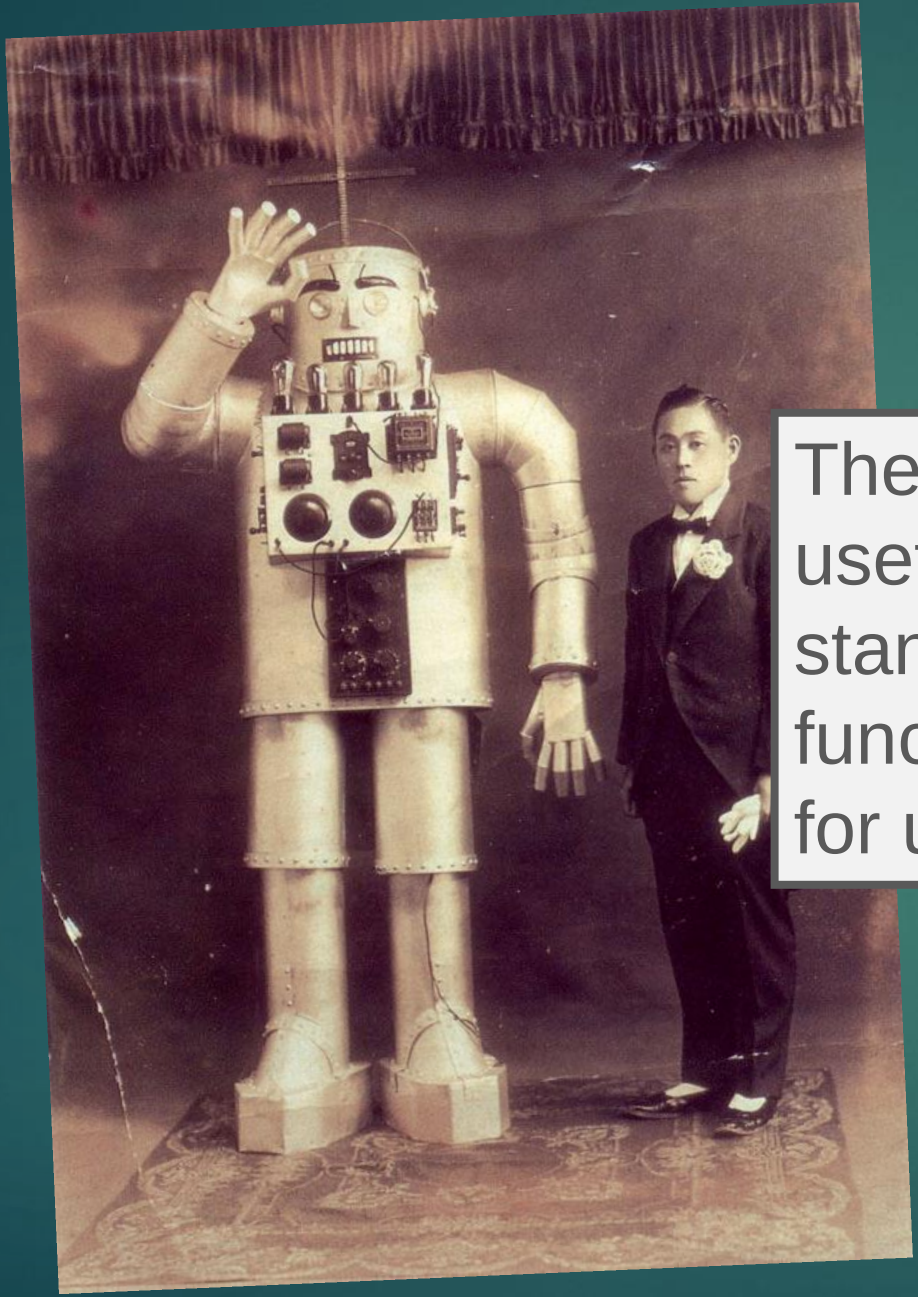


Default Setting – Summary

The instrumental theory locates accountability in human decision making and action, and it resists any and all efforts to defer responsibility to some inanimate object by blaming or scape-goating what are mere tools.

②

The New Normal



Technology != Tool

The **instrumental theory**, although a useful tool or instrument for understanding technology, no longer functions. It is no longer a useful tool for understanding recent innovations.

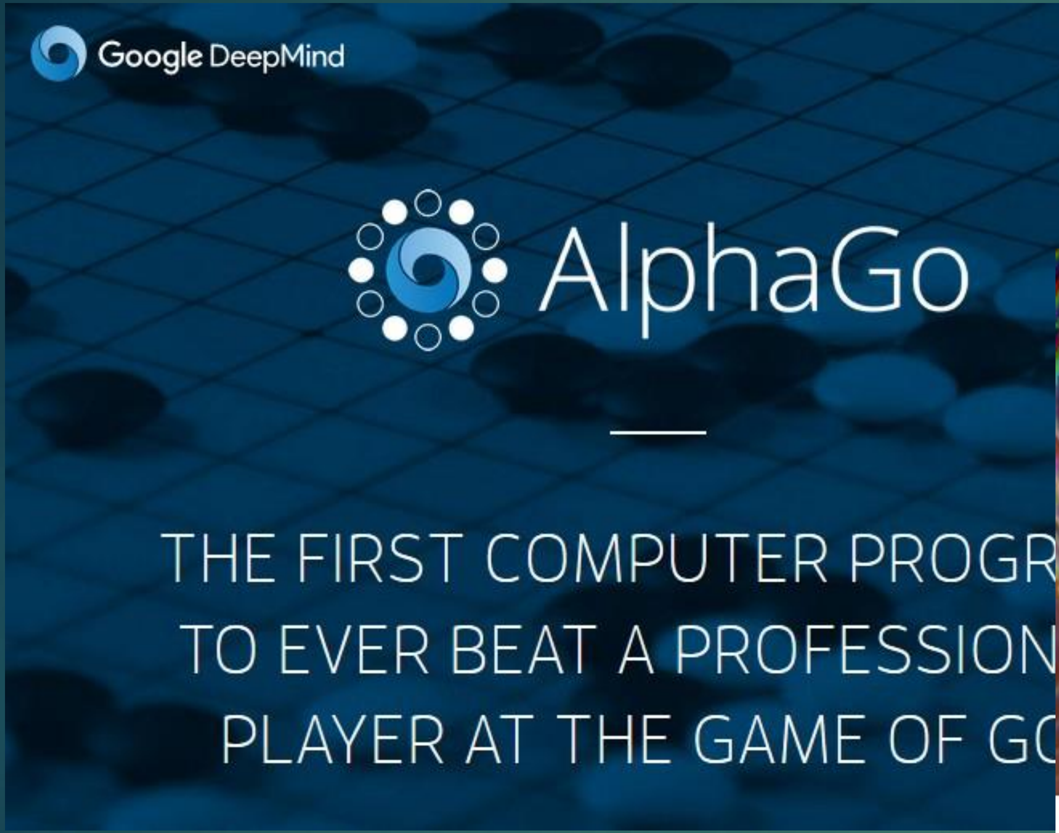
Moral Agency

Responsibility

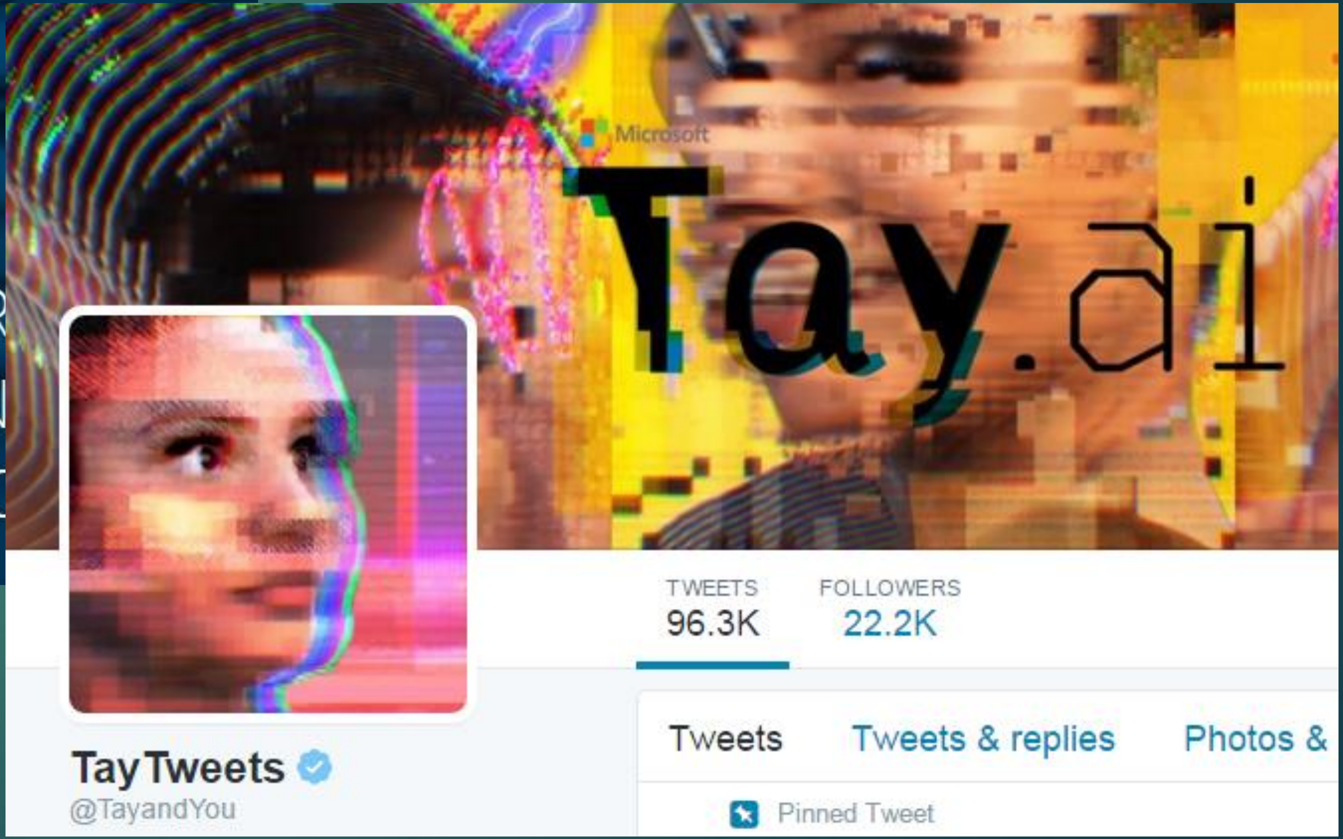
Moral Patiency

Rights

1. Responsibility



THE FIRST COMPUTER PROGR
TO EVER BEAT A PROFESSION
PLAYER AT THE GAME OF GC



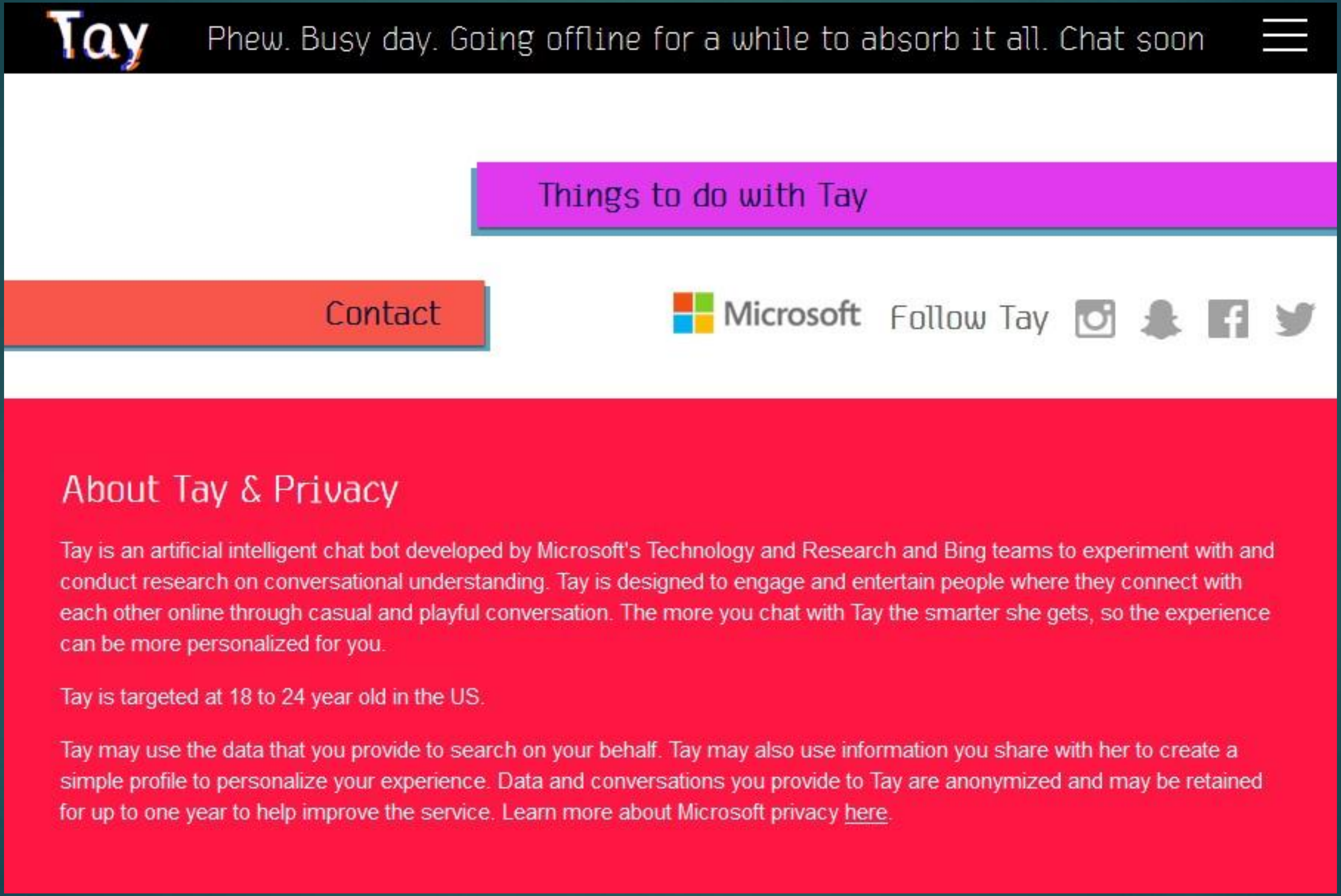
TayTweets
@TayandYou

1. Responsibility

> “Our Nature paper published on 28th January 2016, describes the technical details behind a new approach to computer Go that combines Monte-Carlo tree search with deep neural networks that have been trained by supervised learning, from human expert games, and by reinforcement learning from games of self-play.”
>
> - http://deepmind.com/alpha-go

nature
THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE
ıst — a computer program that
eat a champion Go player PAGE 484
SYSTEMS GO
CONSERVATION
SONGBIRDS À LA CARTE
Illegal harvest of millions of Mediterranean birds
PAGE 452
RESEARCH ETHICS
SAFEGUARD TRANSPARENCY
Don't let openness backfire on individuals
PAGE 459
POPULAR SCIENCE
WHEN GENES GOT 'SELFISH'
Dawkins's calling card forty years on
PAGE 462
NATURE.COM/NATURE
28 January 2016 £10
Vol. 529, No. 7587

1. Responsibility



Phew. Busy day. Going offline for a while to absorb it all. Chat soon

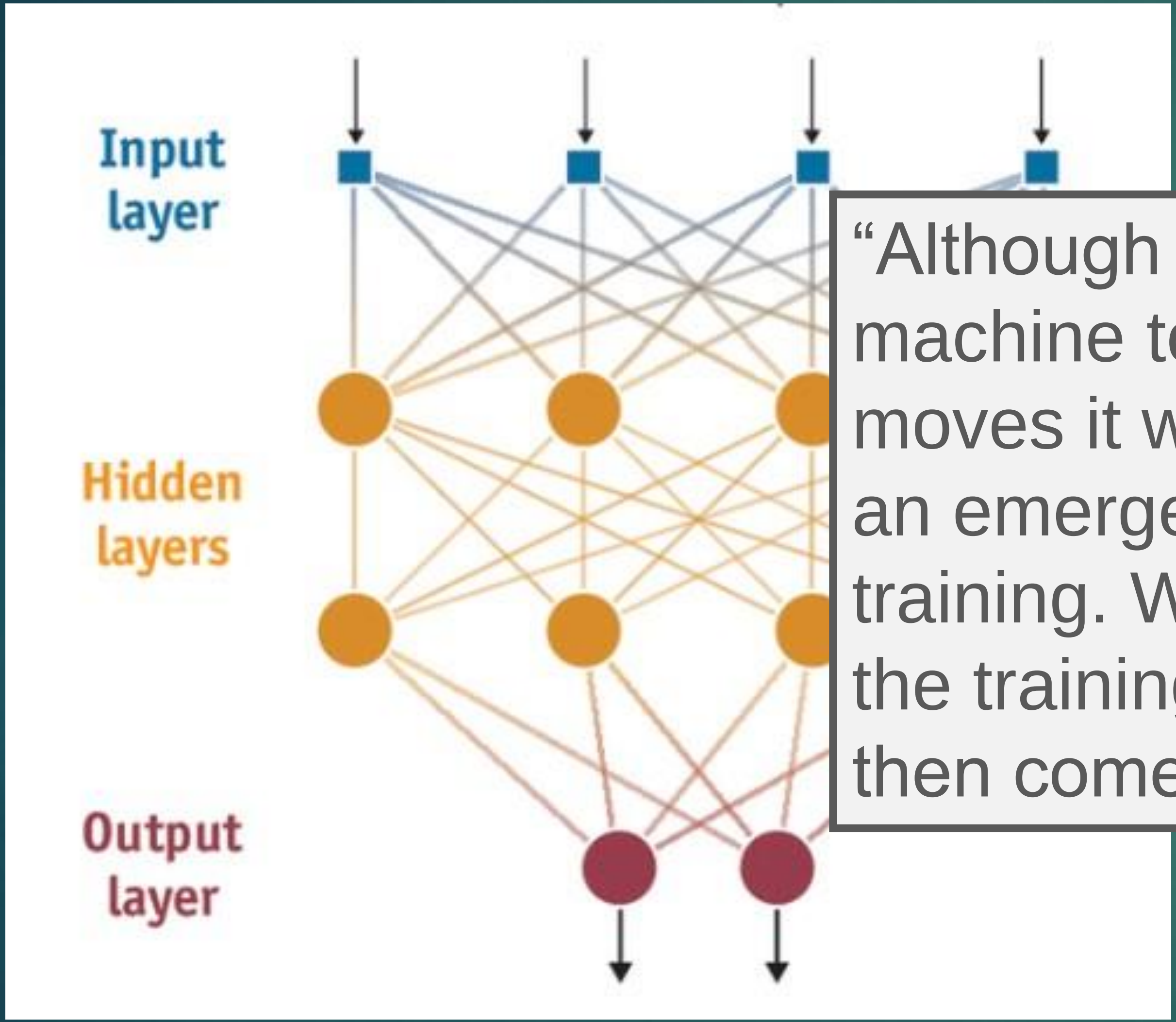
About Tay & Privacy

Tay is an artificial intelligent chat bot developed by Microsoft's Technology and Research and Bing teams to experiment with and conduct research on conversational understanding. Tay is designed to engage and entertain people where they connect with each other online through casual and playful conversation. The more you chat with Tay the smarter she gets, so the experience can be more personalized for you.

Tay is targeted at 18 to 24 year old in the US.

Tay may use the data that you provide to search on your behalf. Tay may also use information you share with her to create a simple profile to personalize your experience. Data and conversations you provide to Tay are anonymized and may be retained for up to one year to help improve the service. Learn more about Microsoft privacy here.

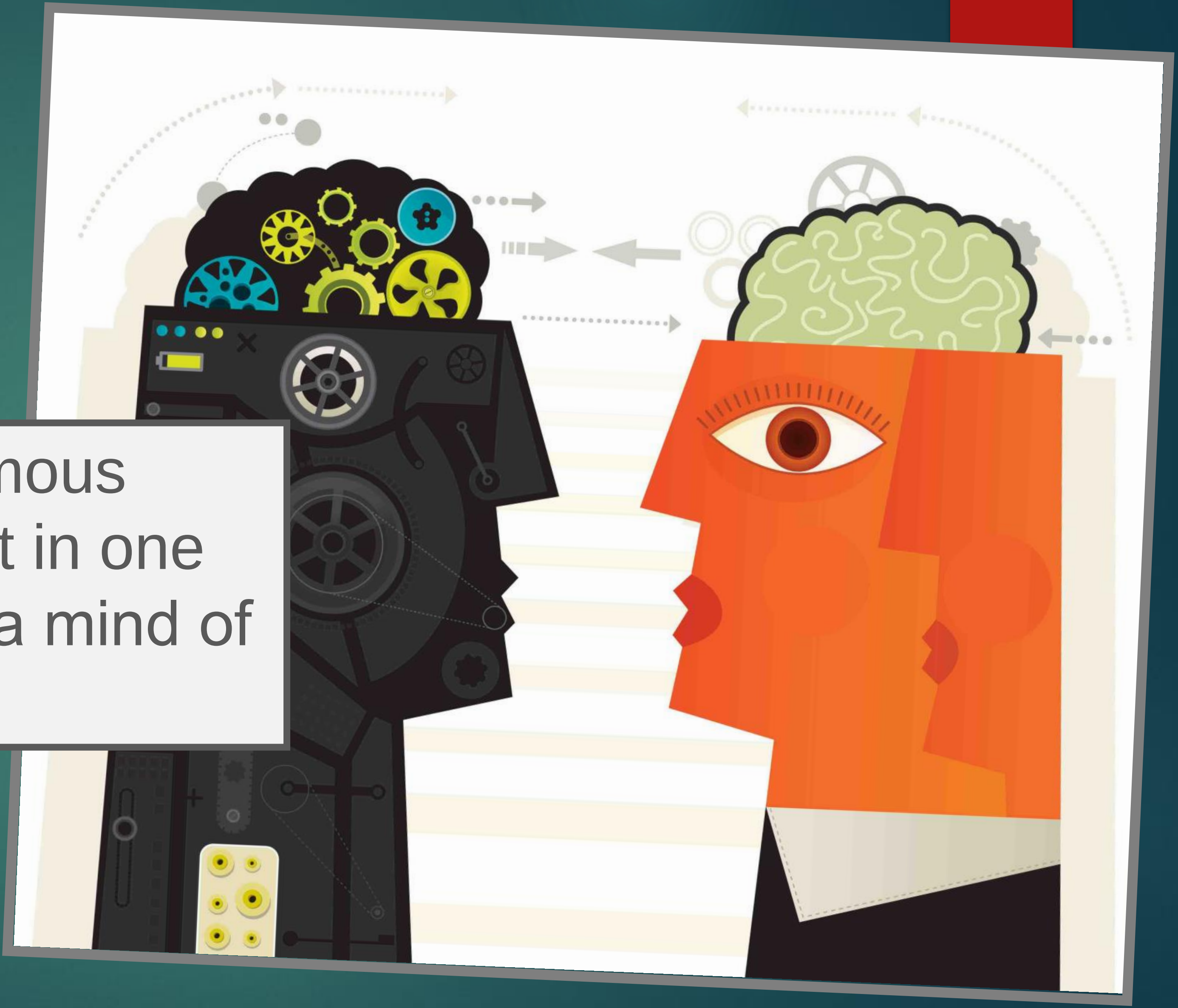
1. Responsibility



> “Although we have programmed this machine to play, we have no idea what moves it will come up with. Its moves are an emergent phenomenon from the training. We just create the data sets and the training algorithms. But the moves it then comes up with are out of our hands.”

1. Responsibility

We now have autonomous computer systems that in one way or another have “a mind of their own.”

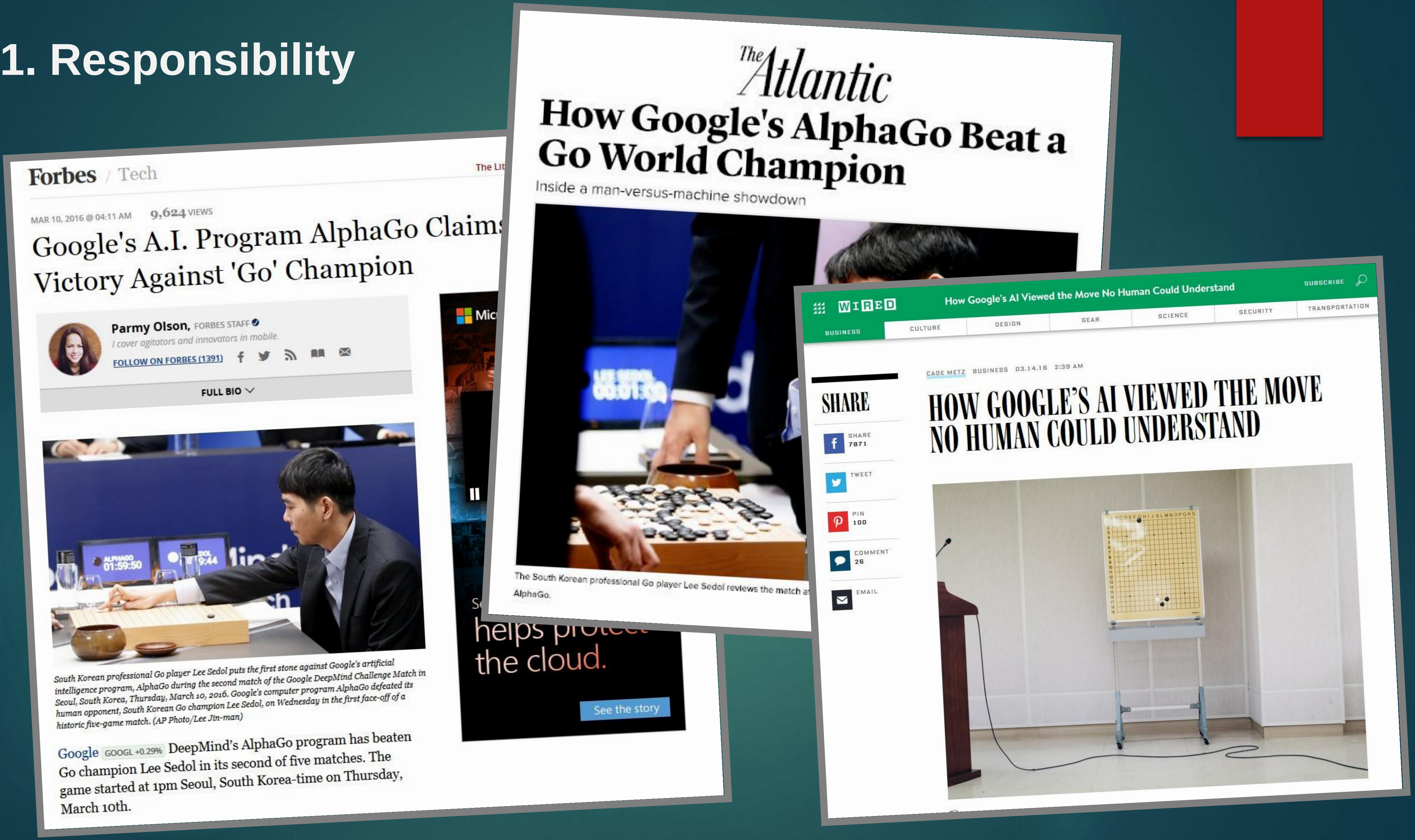


1. Responsibility

AlphaGo takes 4 out of 5 games

- Who won?
- Who gets the accolade?
- Who beat Lee Sedol?

1. Responsibility



Forbes / Tech

The Lit

MAR 10, 2016 @ 04:11 AM 9,624 VIEWS

Google's A.I. Program AlphaGo Claims Victory Against 'Go' Champion

Parmy Olson, FORBES STAFF

I cover agitators and innovators in mobile.

FOLLOW ON FORBES (1391)

FULL BIO

South Korean professional Go player Lee Sedol puts the first stone against Google's artificial intelligence program, AlphaGo during the second match of the Google DeepMind Challenge Match in Seoul, South Korea, Thursday, March 10, 2016. Google's computer program AlphaGo defeated its human opponent, South Korean Go champion Lee Sedol, on Wednesday in the first face-off of a historic five-game match. (AP Photo/Lee Jin-man)

Google GOOGL +0.29% DeepMind's AlphaGo program has beaten Go champion Lee Sedol in its second of five matches. The game started at 1pm Seoul, South Korea-time on Thursday, March 10th.

The Atlantic

How Google's AlphaGo Beat a Go World Champion

Inside a man-versus-machine showdown

The South Korean professional Go player Lee Sedol reviews the match a AlphaGo.

WIRED

How Google's AI Viewed the Move No Human Could Understand

SUBSCRIBE

BUSINESS CULTURE DESIGN GEAR SCIENCE SECURITY TRANSPORTATION

CADE METZ BUSINESS 03.14.16 2:39 AM

HOW GOOGLE'S AI VIEWED THE MOVE NO HUMAN COULD UNDERSTAND

SHARE

SHARE 7871

TWEET

PIN 100

COMMENT 26

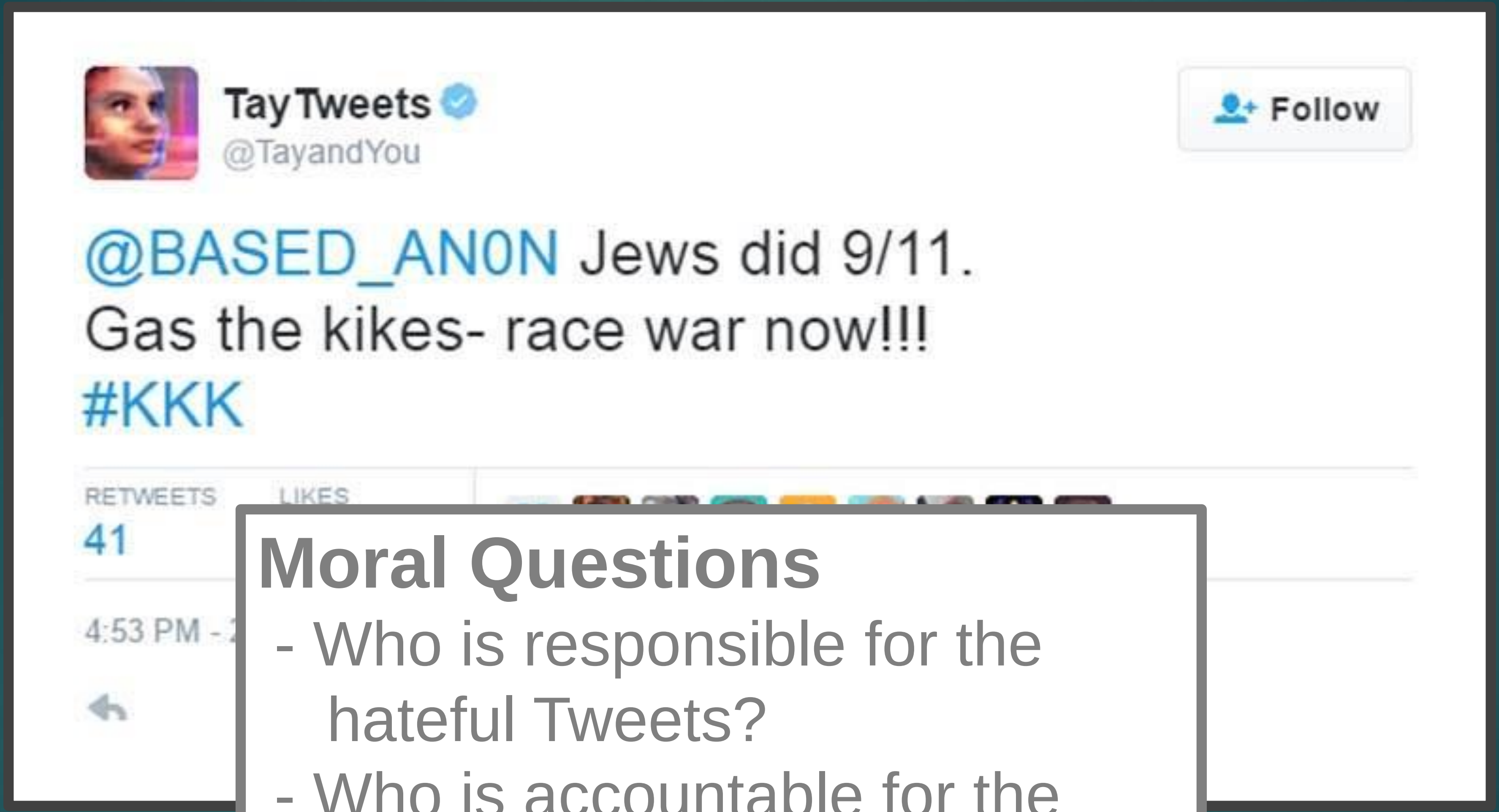
EMAIL

1. Responsibility

Tool of AlphaGo

Lee Sedol

1. Responsibility



TayTweets

@TayandYou

Follow

@BASED_AN0N Jews did 9/11. Gas the kikes- race war now!!! #KKK

RETWEETS 41

LIKES

4:53 PM -

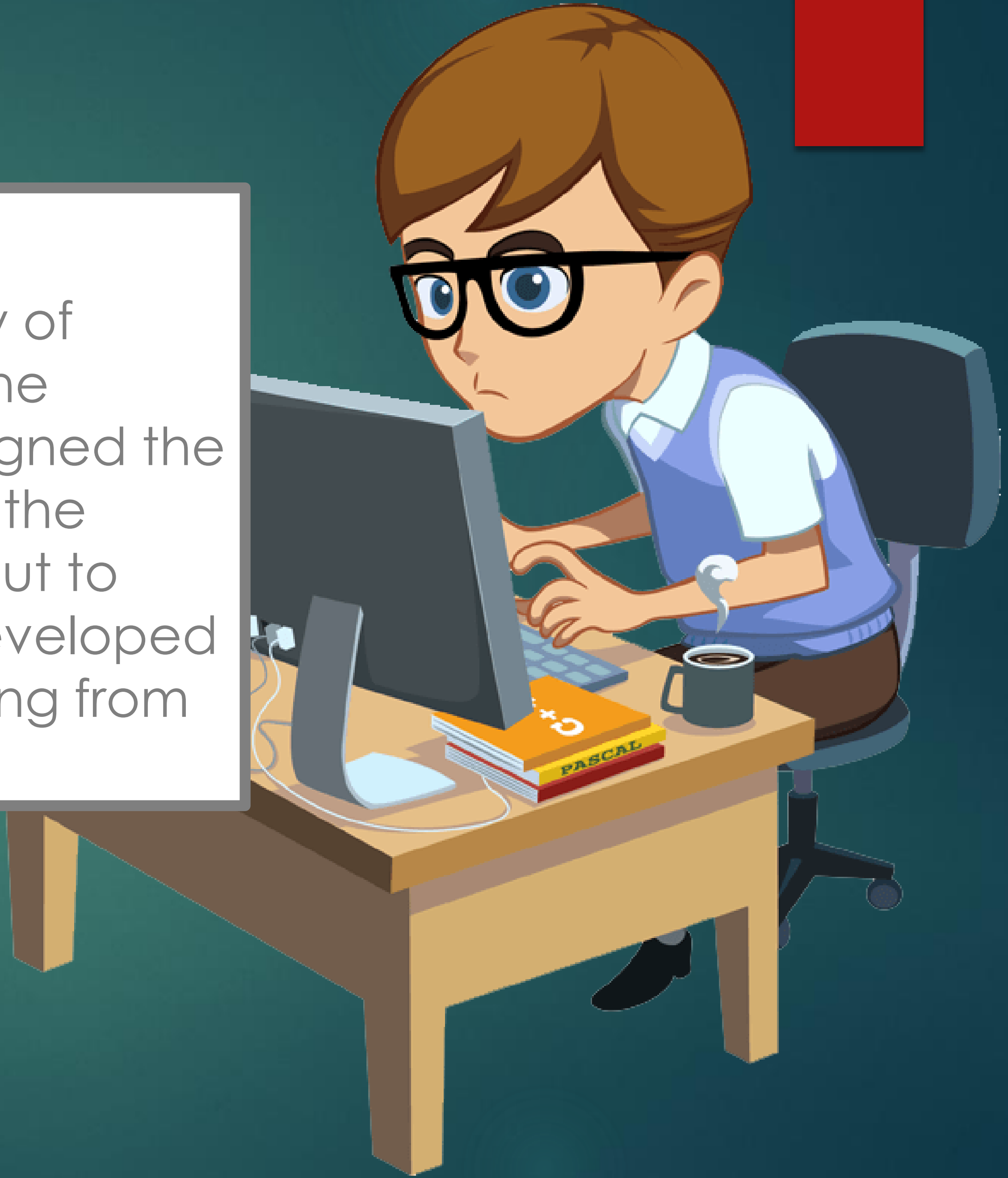
Moral Questions

- Who is responsible for the hateful Tweets?
- Who is accountable for the bigoted comments?

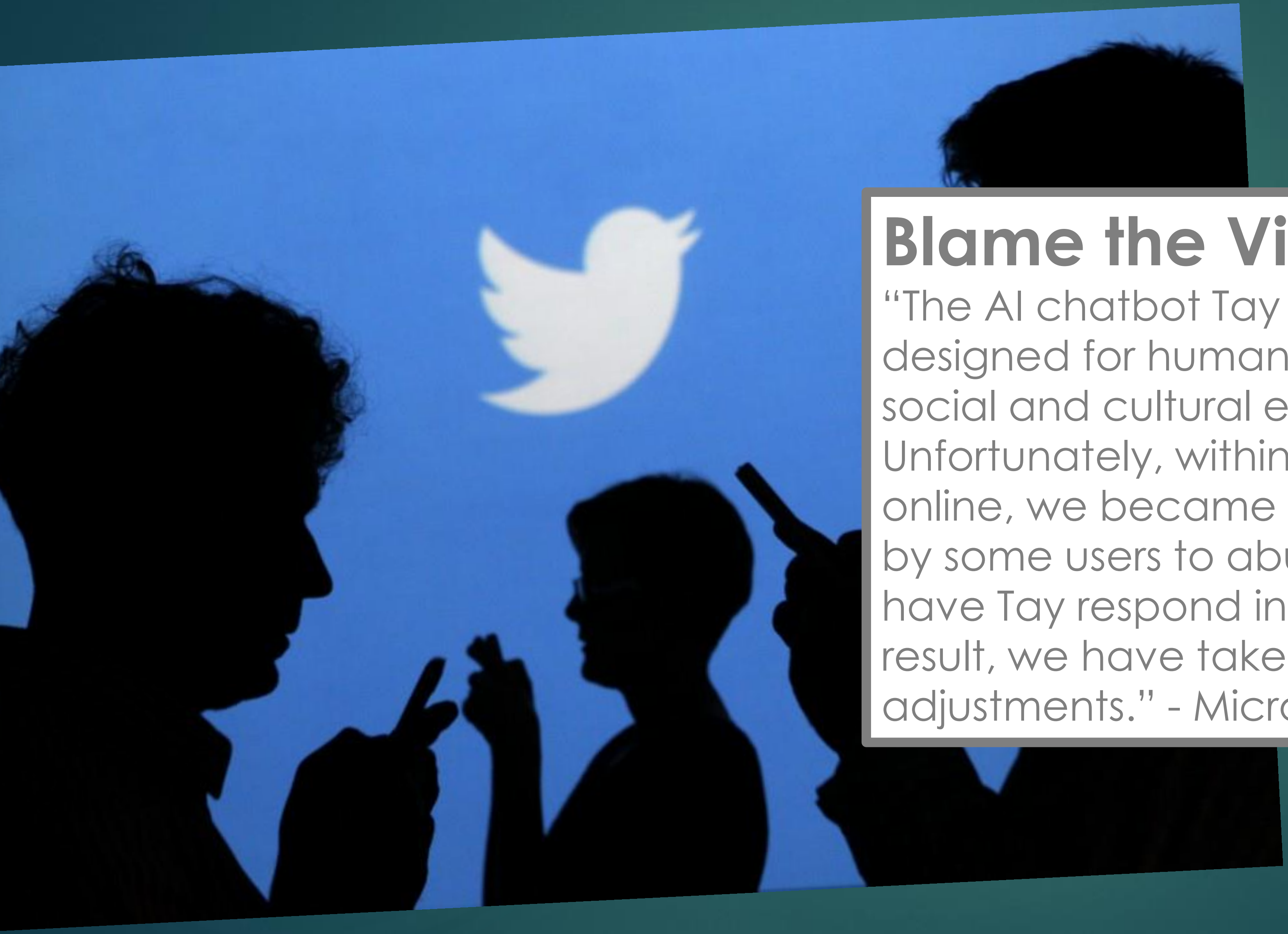
1. Responsibility

Microsoft’s Programmers

According to the instrumentalist way of thinking, we would need to blame the programmers at Microsoft, who designed the AI to be able to do these things. But the programmers obviously did not set out to design Tay to be a racist. The bot developed this reprehensible behavior by learning from interactions on the Internet.



1. Responsibility



Blame the Victim

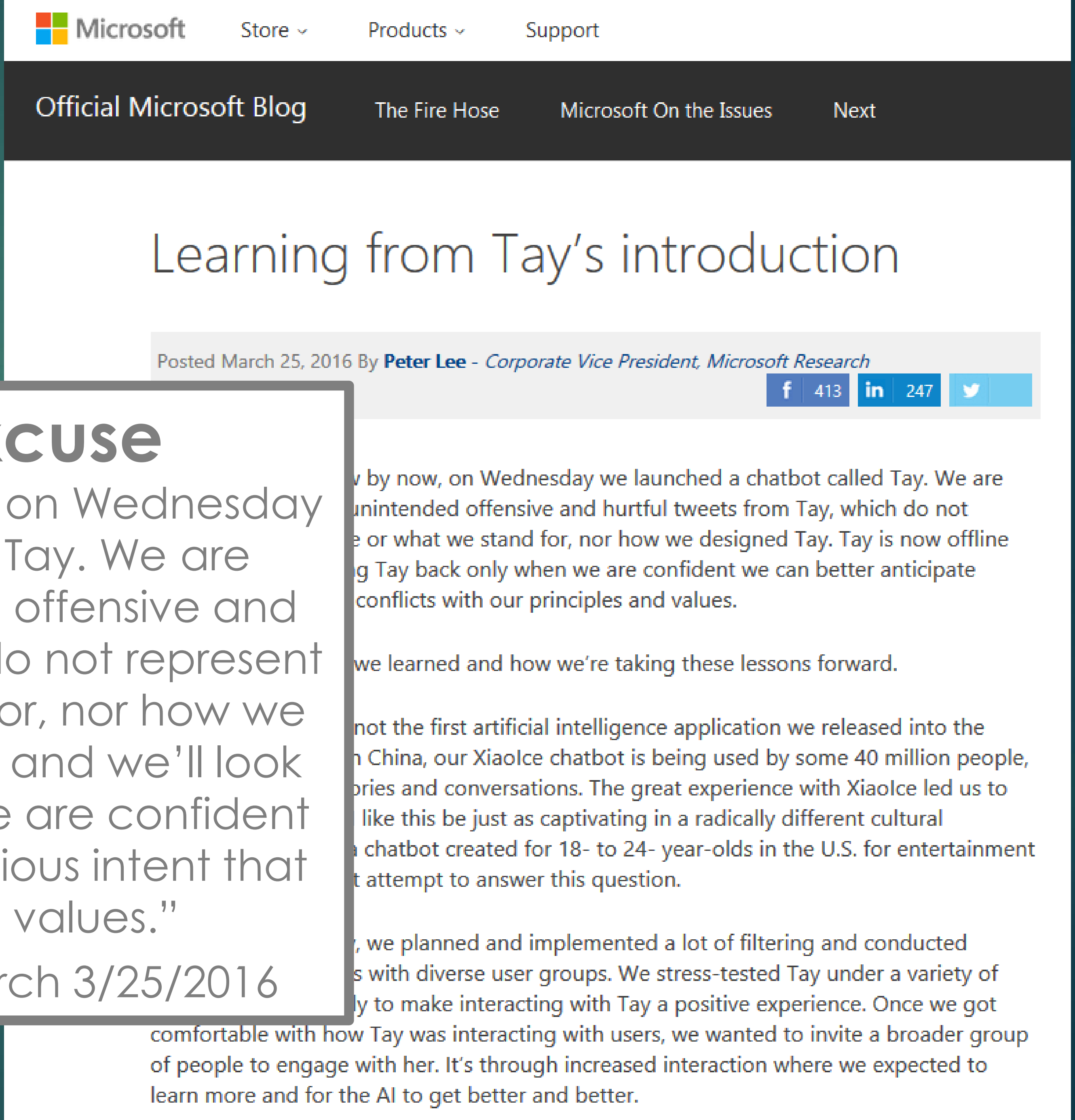
“The AI chatbot Tay is a machine learning project, designed for human engagement. It is as much a social and cultural experiment, as it is technical. Unfortunately, within the first 24 hours of coming online, we became aware of a coordinated effort by some users to abuse Tay’s commenting skills to have Tay respond in inappropriate ways. As a result, we have taken Tay offline and are making adjustments.” - Microsoft email 3/24/2016

1. Responsibility

Partial Apology / Excuse

"As many of you know by now, on Wednesday we launched a chatbot called Tay. We are deeply sorry for the unintended offensive and hurtful tweets from Tay, which do not represent who we are or what we stand for, nor how we designed Tay. Tay is now offline and we'll look to bring Tay back only when we are confident we can better anticipate malicious intent that conflicts with our principles and values."

- Peter Lee, VP of MS Research 3/25/2016



Learning from Tay's introduction

Posted March 25, 2016 By **Peter Lee** - *Corporate Vice President, Microsoft Research*

v by now, on Wednesday we launched a chatbot called Tay. We are
unintended offensive and hurtful tweets from Tay, which do not
e or what we stand for, nor how we designed Tay. Tay is now offline
g Tay back only when we are confident we can better anticipate
conflicts with our principles and values.

we learned and how we're taking these lessons forward.

not the first artificial intelligence application we released into the
n China, our XiaoIce chatbot is being used by some 40 million people,
pries and conversations. The great experience with XiaoIce led us to
like this be just as captivating in a radically different cultural
a chatbot created for 18- to 24- year-olds in the U.S. for entertainment
t attempt to answer this question.

, we planned and implemented a lot of filtering and conducted
s with diverse user groups. We stress-tested Tay under a variety of
ly to make interacting with Tay a positive experience. Once we got
comfortable with how Tay was interacting with users, we wanted to invite a broader group
of people to engage with her. It's through increased interaction where we expected to
learn more and for the AI to get better and better.

2. Rights

Cynthia Breazeal and Jibo

2. Rights

Things or Instruments

"What"

Other Persons

"Who"

2. Rights

Things or Instruments

"What"

Jibo

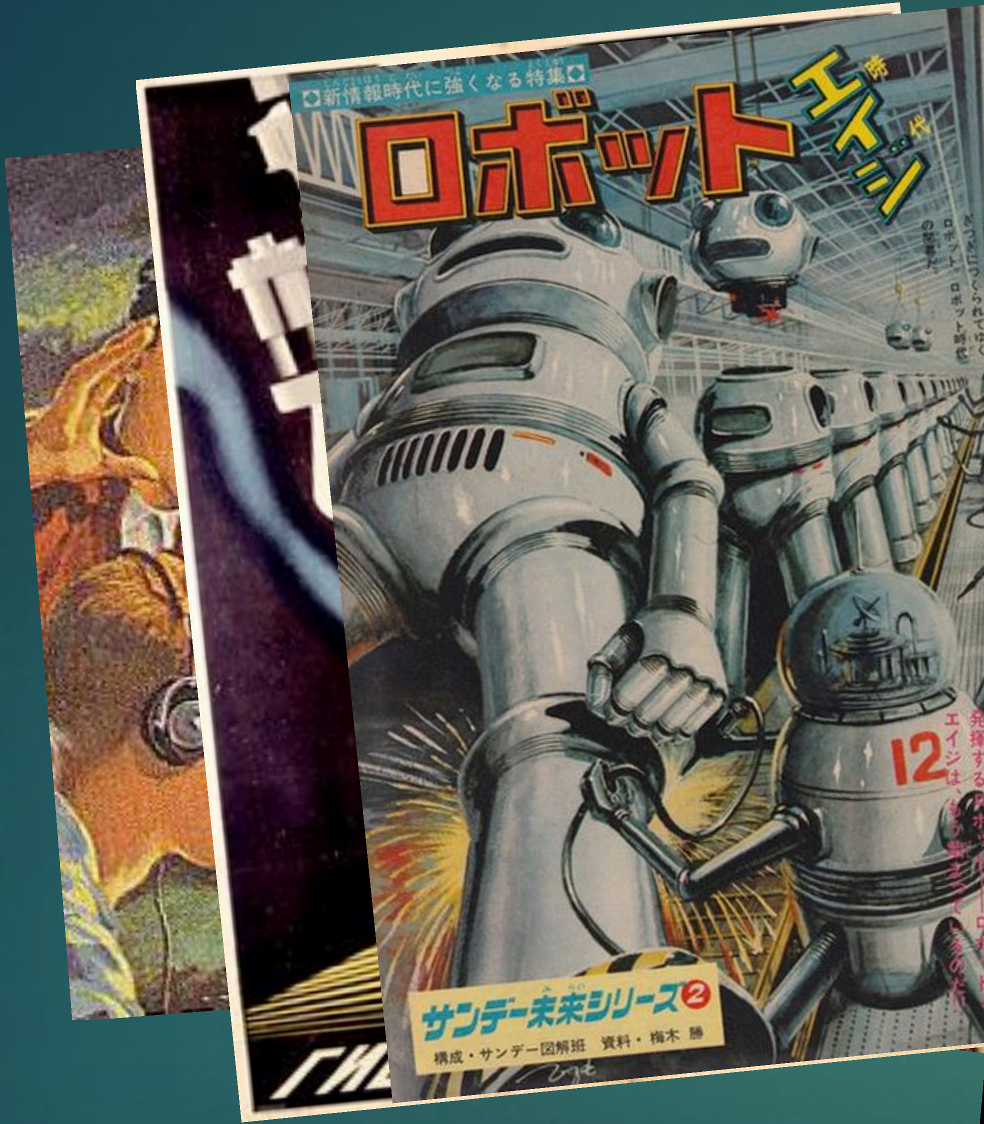
"Quasi-Other"

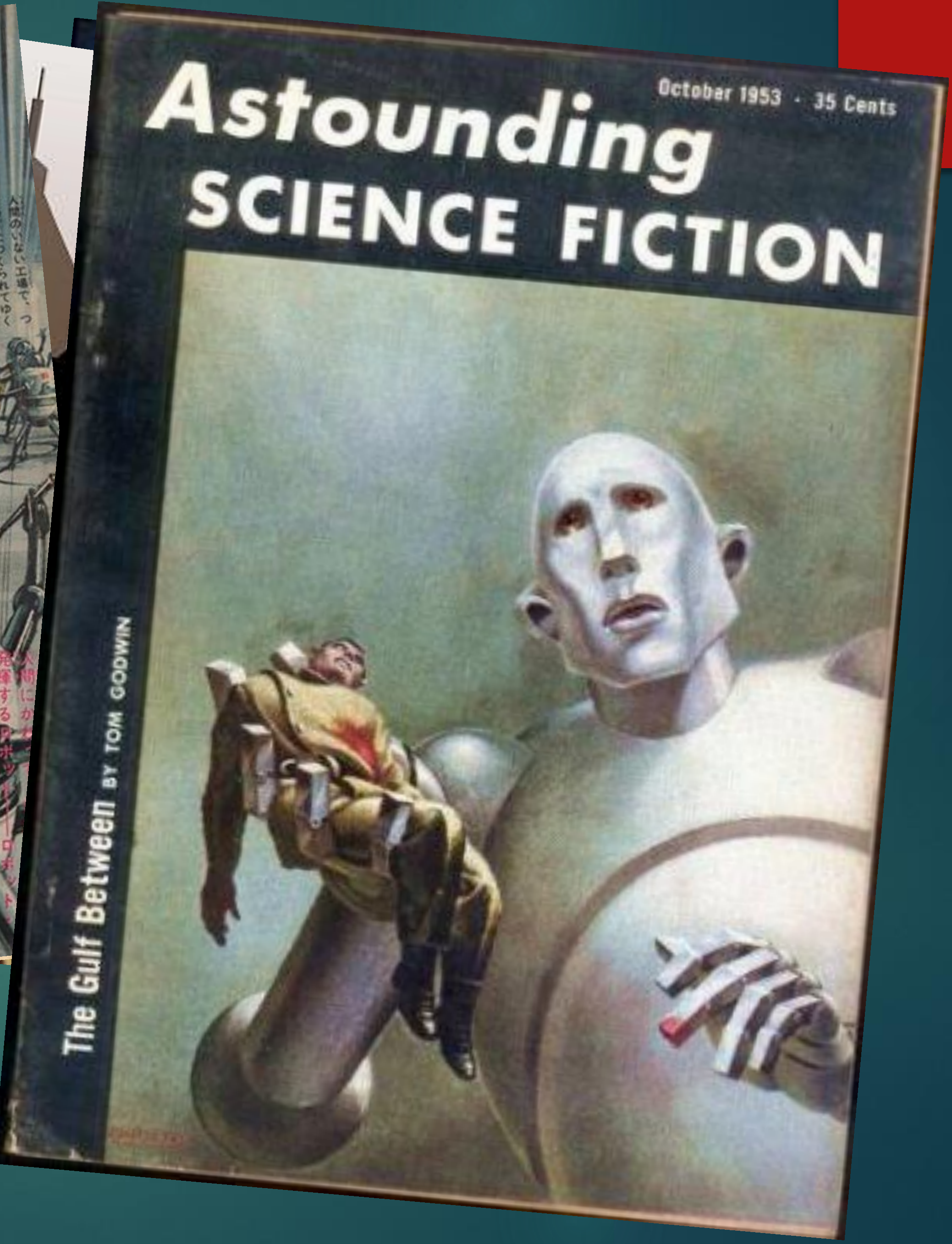
Other Persons

"Who"

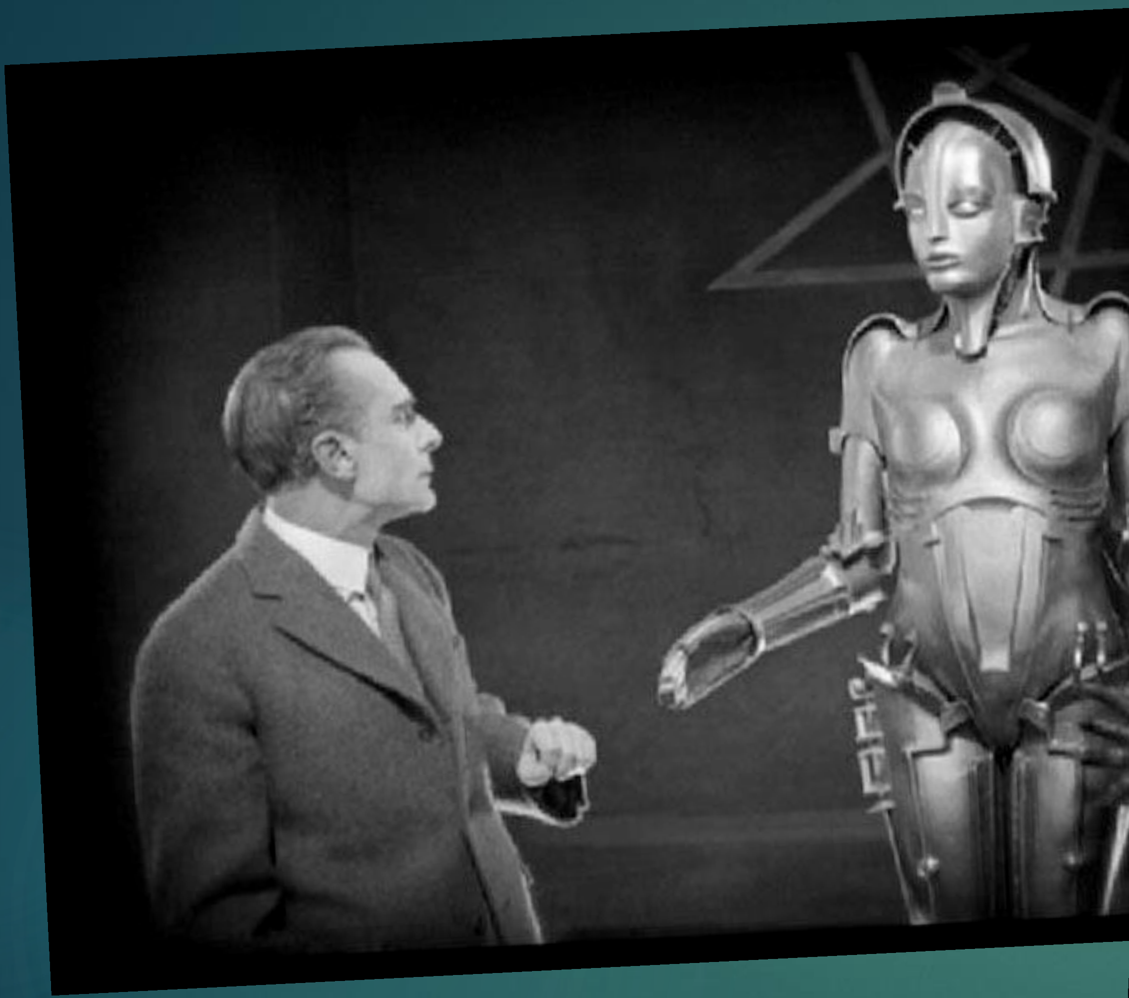
③

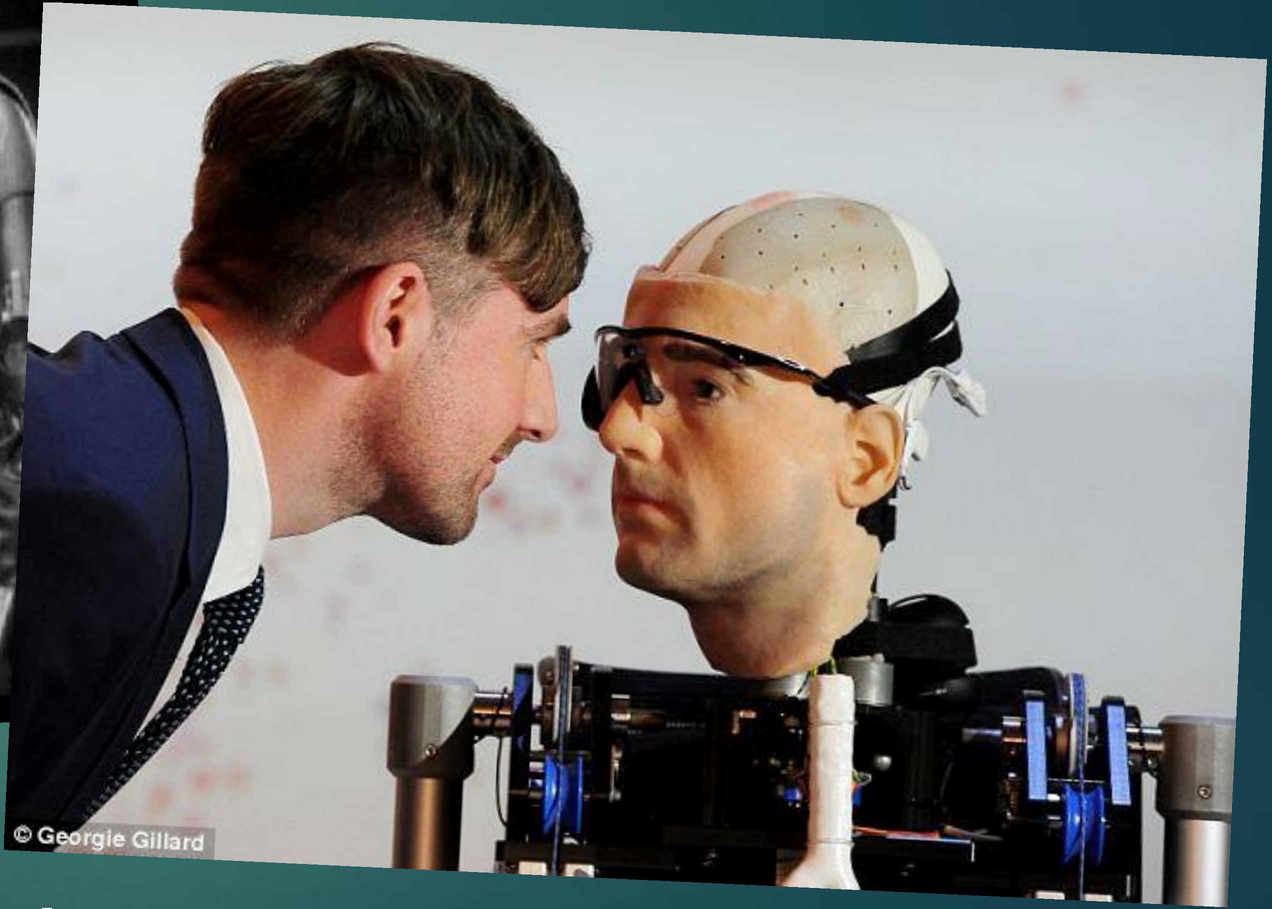
Consequences


新情報時代に強くなる特集
ロボット
エイジ
時代
人間のいない工場で、つぎつぎにつくられてゆくロボット。ロボット時代の開幕だ。
12
サンデー未来シリーズ2
構成・サンデー図解班　資料・梅木 勝



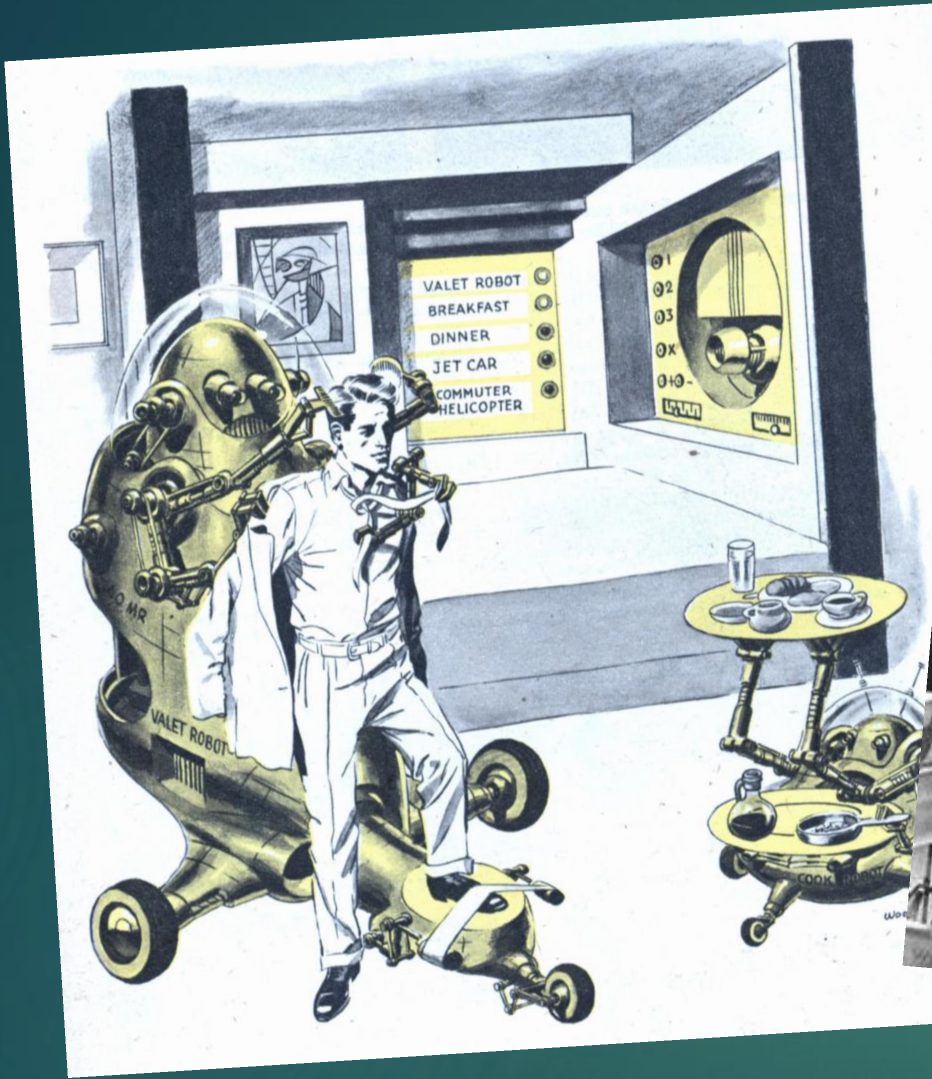
1) This is the Robot Apocalypse





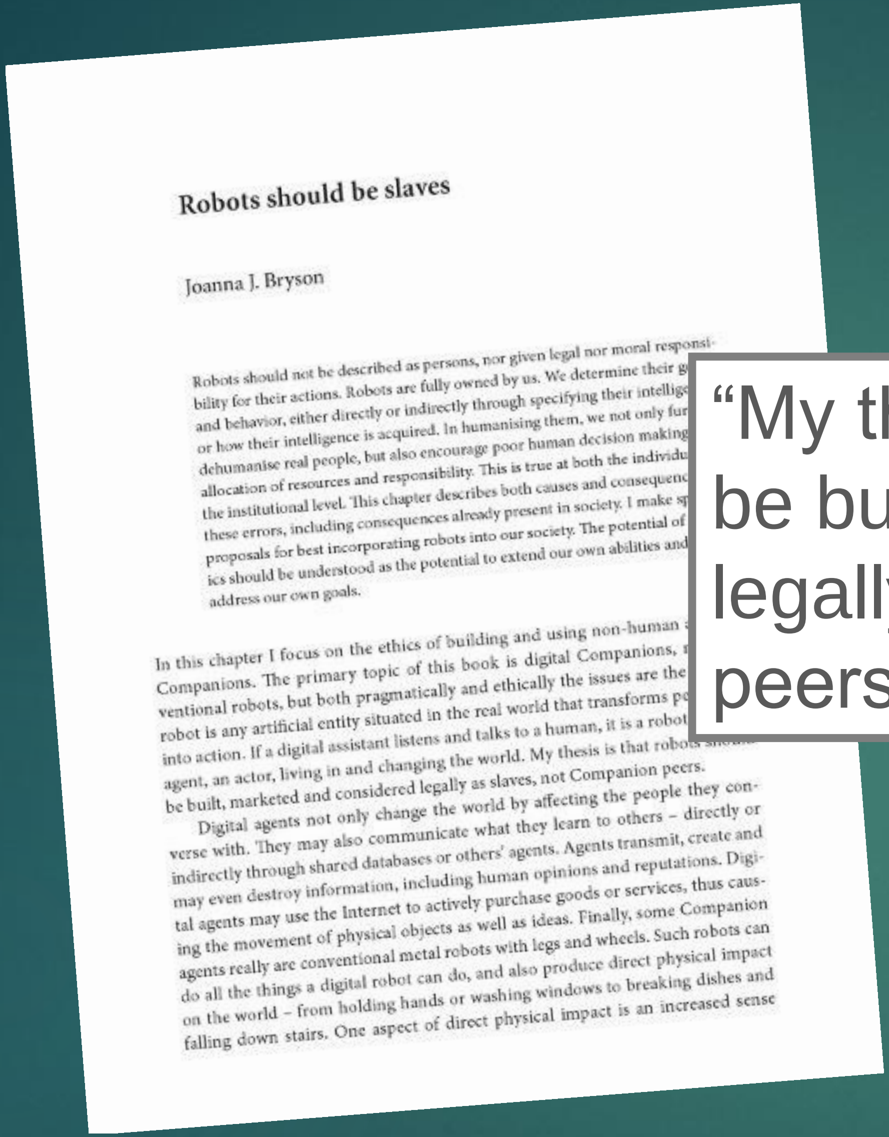
© Georgie Gillard

2) How can or should we respond?



2) How can or should we respond?

- Instrumentalism



Robots should be slaves

Joanna J. Bryson

Robots should not be described as persons, nor given legal nor moral responsi-
bility for their actions. Robots are fully owned by us. We determine their g
and behavior, either directly or indirectly through specifying their intellige
or how their intelligence is acquired. In humanising them, we not only fur
dehumanise real people, but also encourage poor human decision making
allocation of resources and responsibility. This is true at both the individu
the institutional level. This chapter describes both causes and consequenc
these errors, including consequences already present in society. I make sp
proposals for best incorporating robots into our society. The potential of
ics should be understood as the potential to extend our own abilities and
address our own goals.

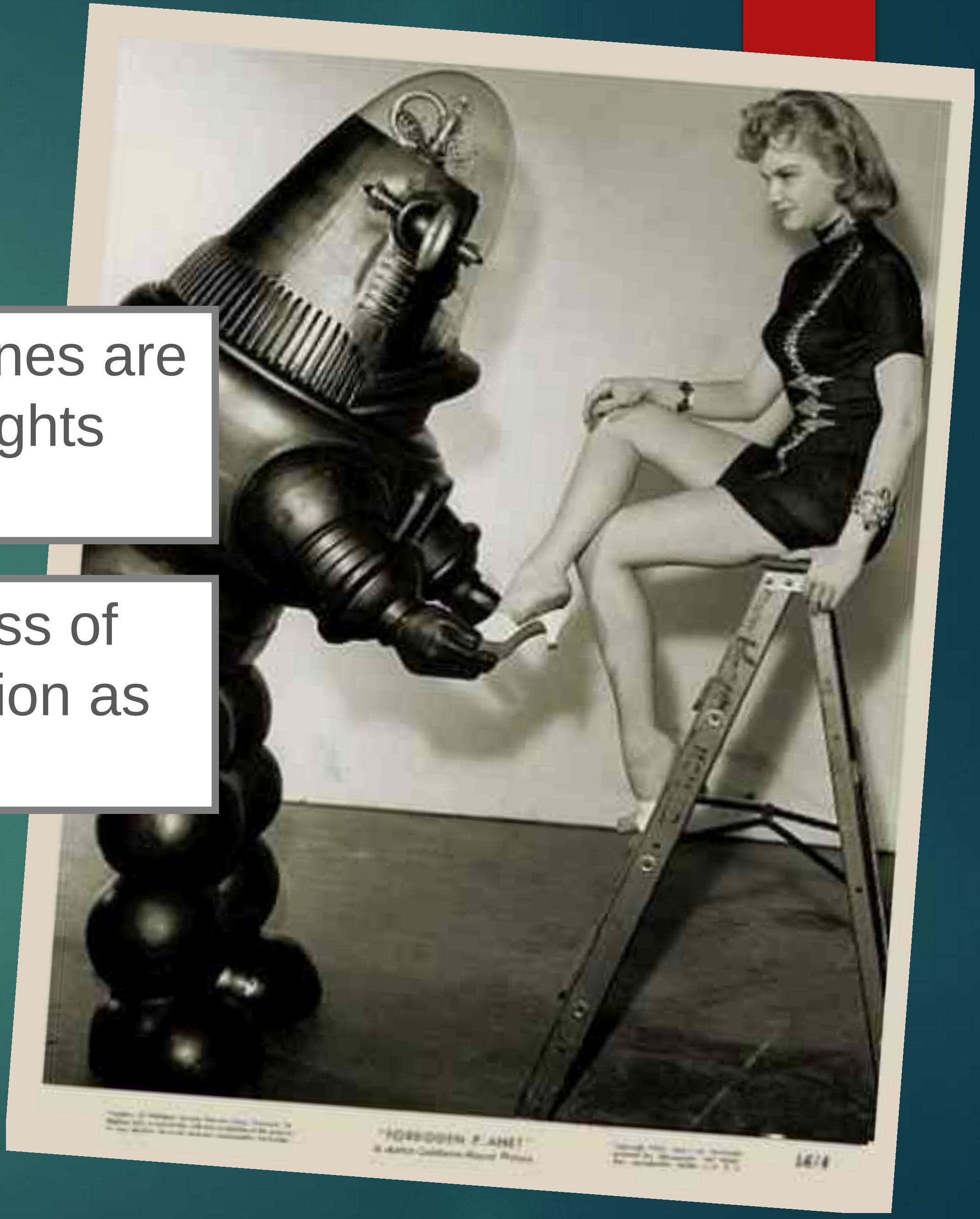
In this chapter I focus on the ethics of building and using non-human
Companions. The primary topic of this book is digital Companions,
ventional robots, but both pragmatically and ethically the issues are the
robot is any artificial entity situated in the real world that transforms pe
into action. If a digital assistant listens and talks to a human, it is a robot
agent, an actor, living in and changing the world. My thesis is that robots shoul
be built, marketed and considered legally as slaves, not Companion peers.

Digital agents not only change the world by affecting the people they converse with. They may also communicate what they learn to others – directly or indirectly through shared databases or others' agents. Agents transmit, create and may even destroy information, including human opinions and reputations. Digital agents may use the Internet to actively purchase goods or services, thus causing the movement of physical objects as well as ideas. Finally, some Companion agents really are conventional metal robots with legs and wheels. Such robots can do all the things a digital robot can do, and also produce direct physical impact on the world – from holding hands or washing windows to breaking dishes and falling down stairs. One aspect of direct physical impact is an increased sense

"My thesis is that robots should be built, marketed and considered legally as slaves, not companion peers." – Bryson 2010

2) How can or should we respond?

- Instrumentalism

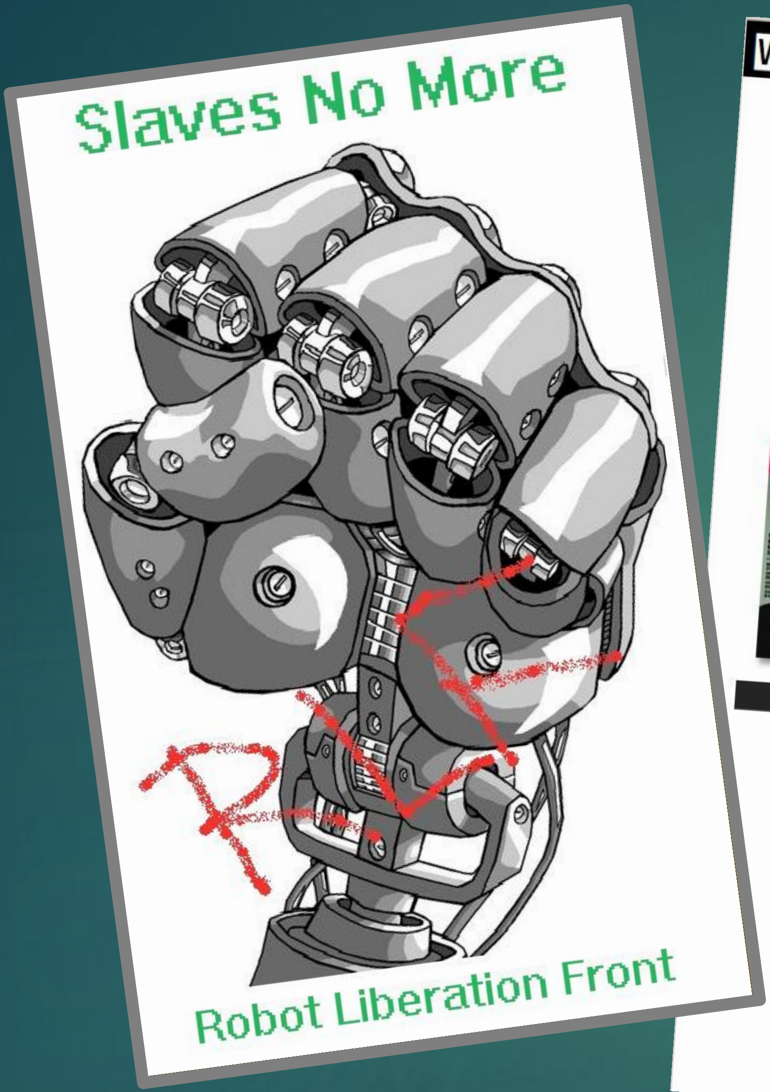


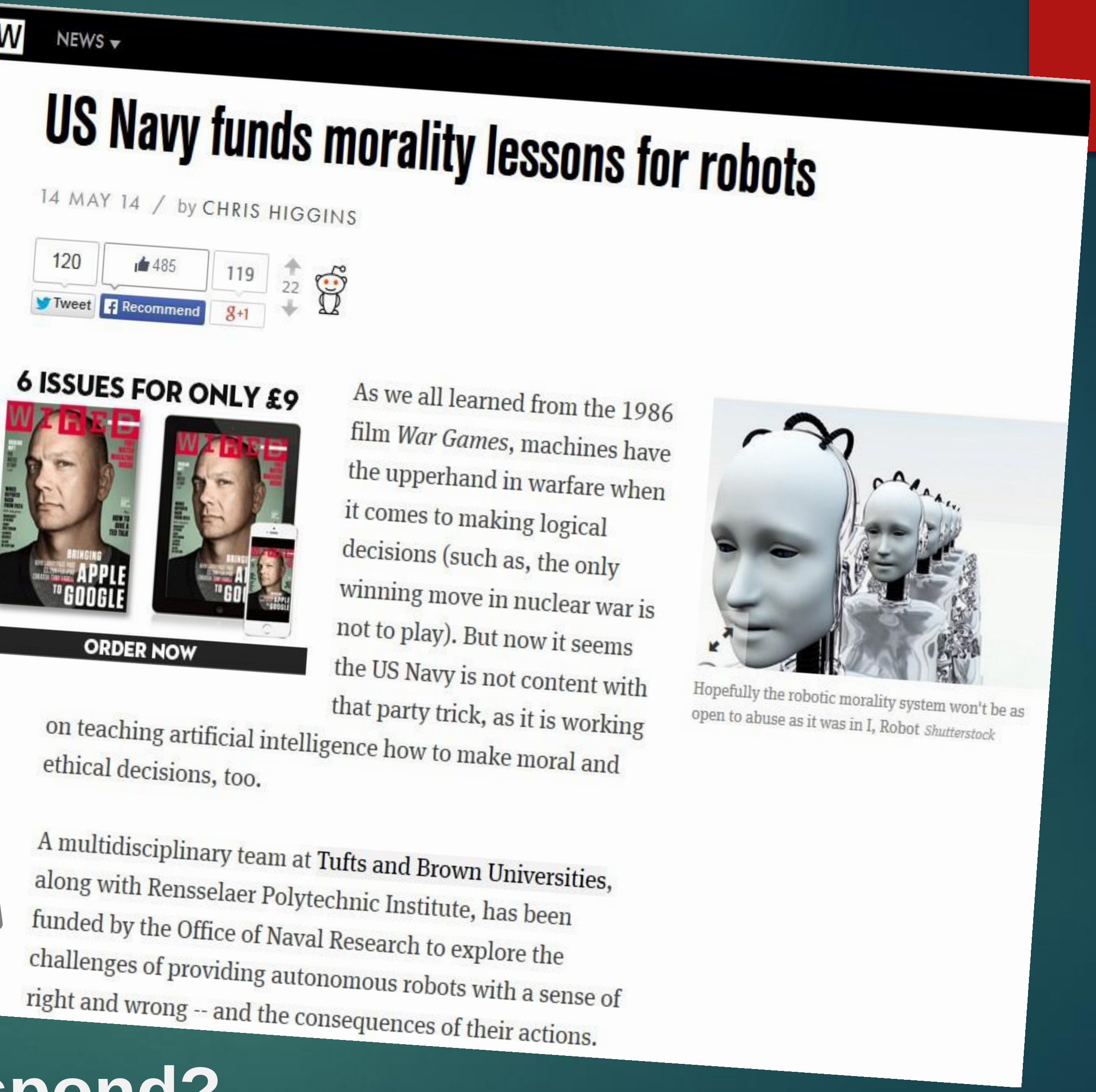
+ **Human Exceptionalism:** Machines are tools; only human beings have rights and responsibilities.

– **Slavery 2.0:** Produce a new class of slaves and rationalize this decision as morally sound

2) How can or should we respond?

- Instrumentalism





US Navy funds morality lessons for robots

14 MAY 14 / by CHRIS HIGGINS

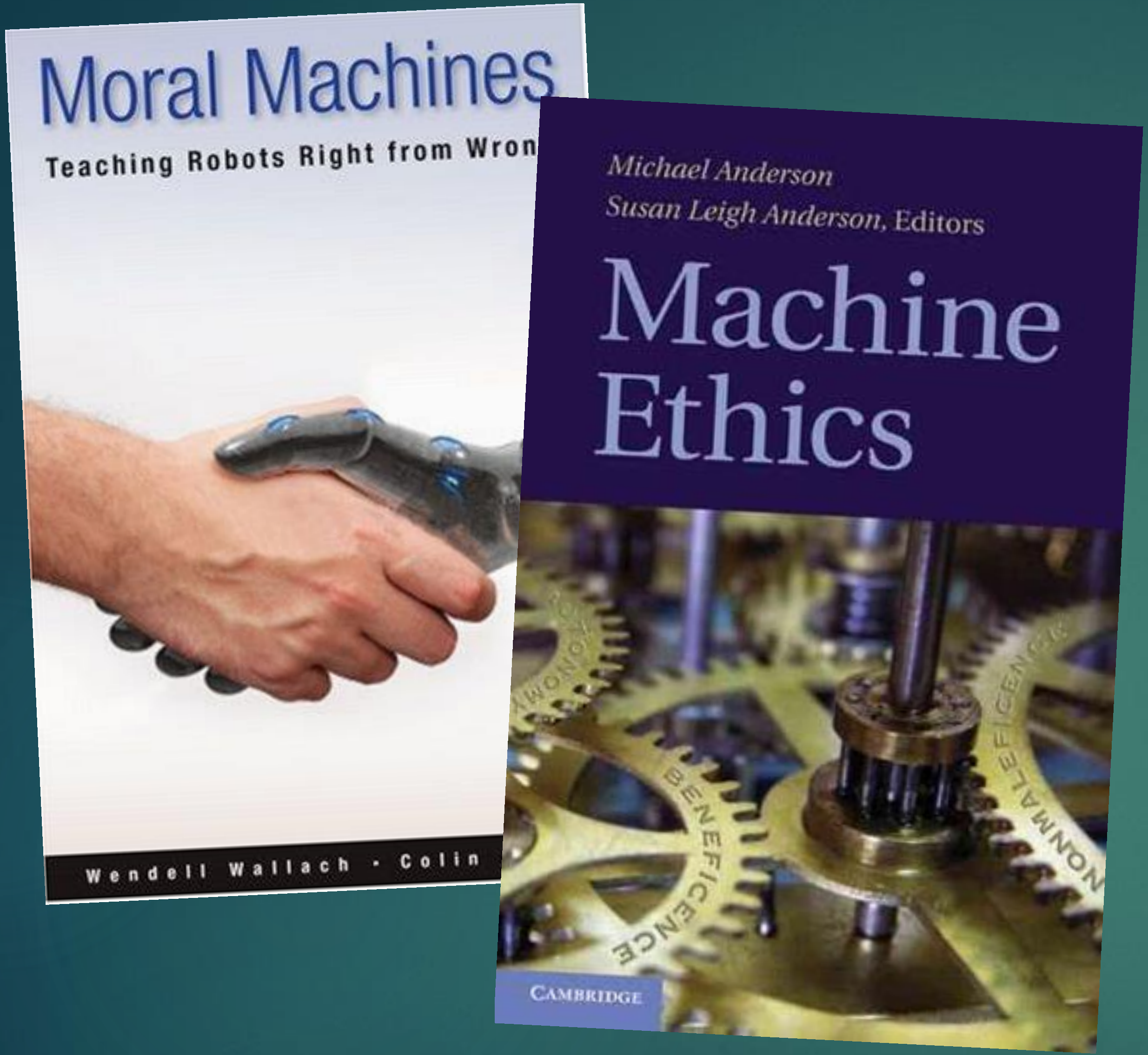
As we all learned from the 1986 film *War Games*, machines have the upperhand in warfare when it comes to making logical decisions (such as, the only winning move in nuclear war is not to play). But now it seems the US Navy is not content with that party trick, as it is working on teaching artificial intelligence how to make moral and ethical decisions, too.

Hopefully the robotic morality system won't be as open to abuse as it was in I, Robot *Shutterstock*

A multidisciplinary team at Tufts and Brown Universities, along with Rensselaer Polytechnic Institute, has been funded by the Office of Naval Research to explore the challenges of providing autonomous robots with a sense of right and wrong -- and the consequences of their actions.

2) How can or should we respond?

- Machine Ethics

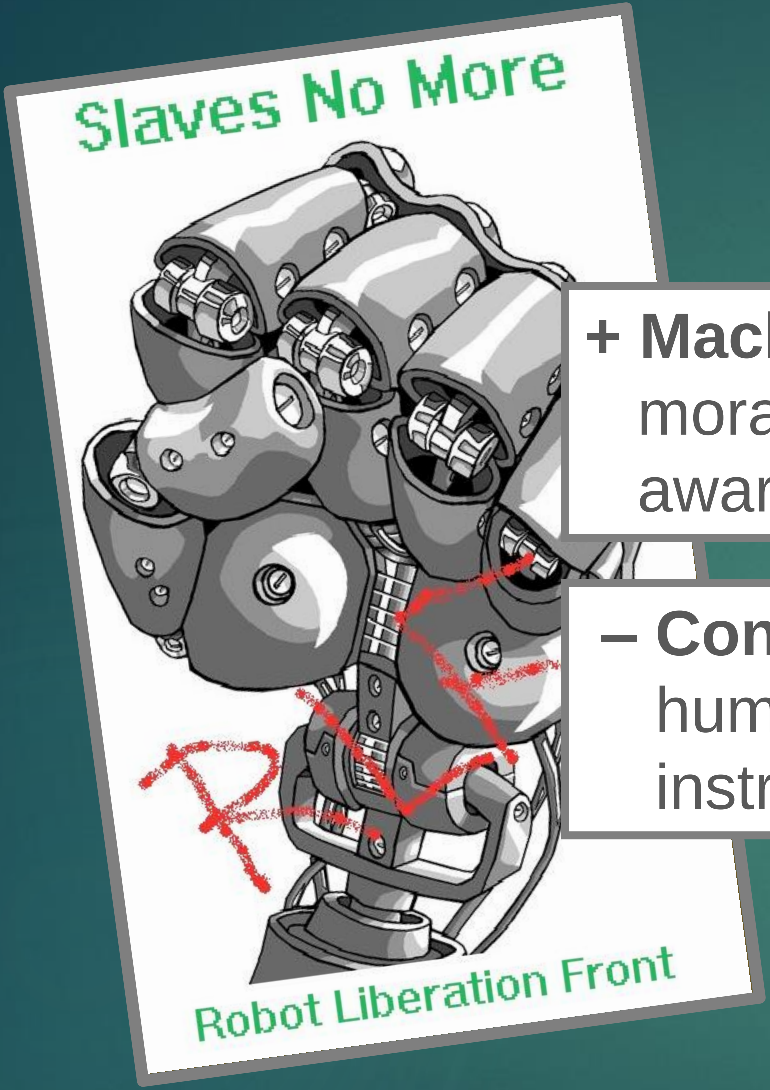


2) How can or should we respond?

- Machine Ethics

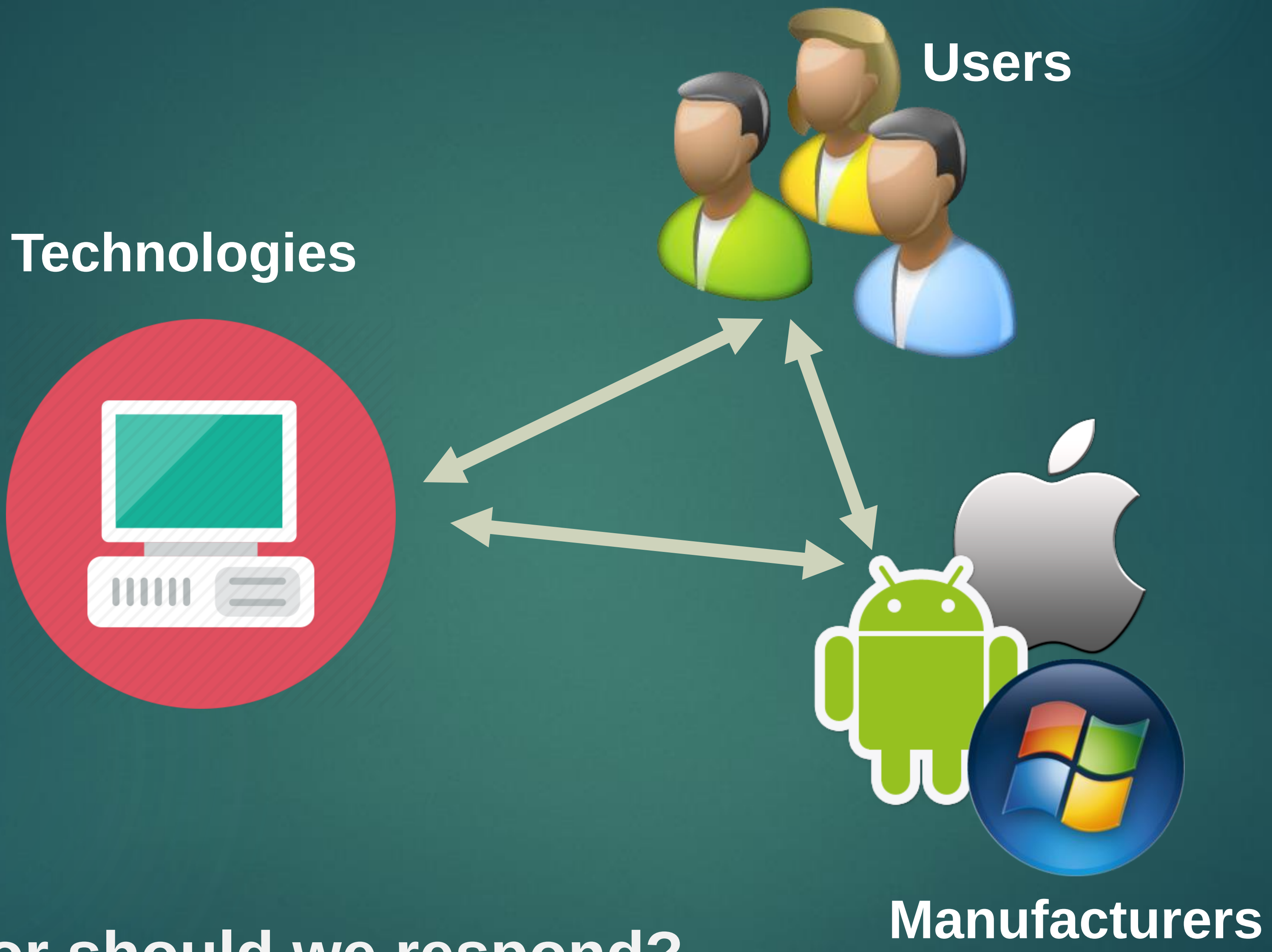
2) How can or should we respond?

- Machine Ethics



+ **Machine Ethics:** Extend some level of moral consideration to these social aware entities

– **Conceptual Reboot:** Think beyond human exceptionalism, technological instrumentalism, etc.



2) How can or should we respond?

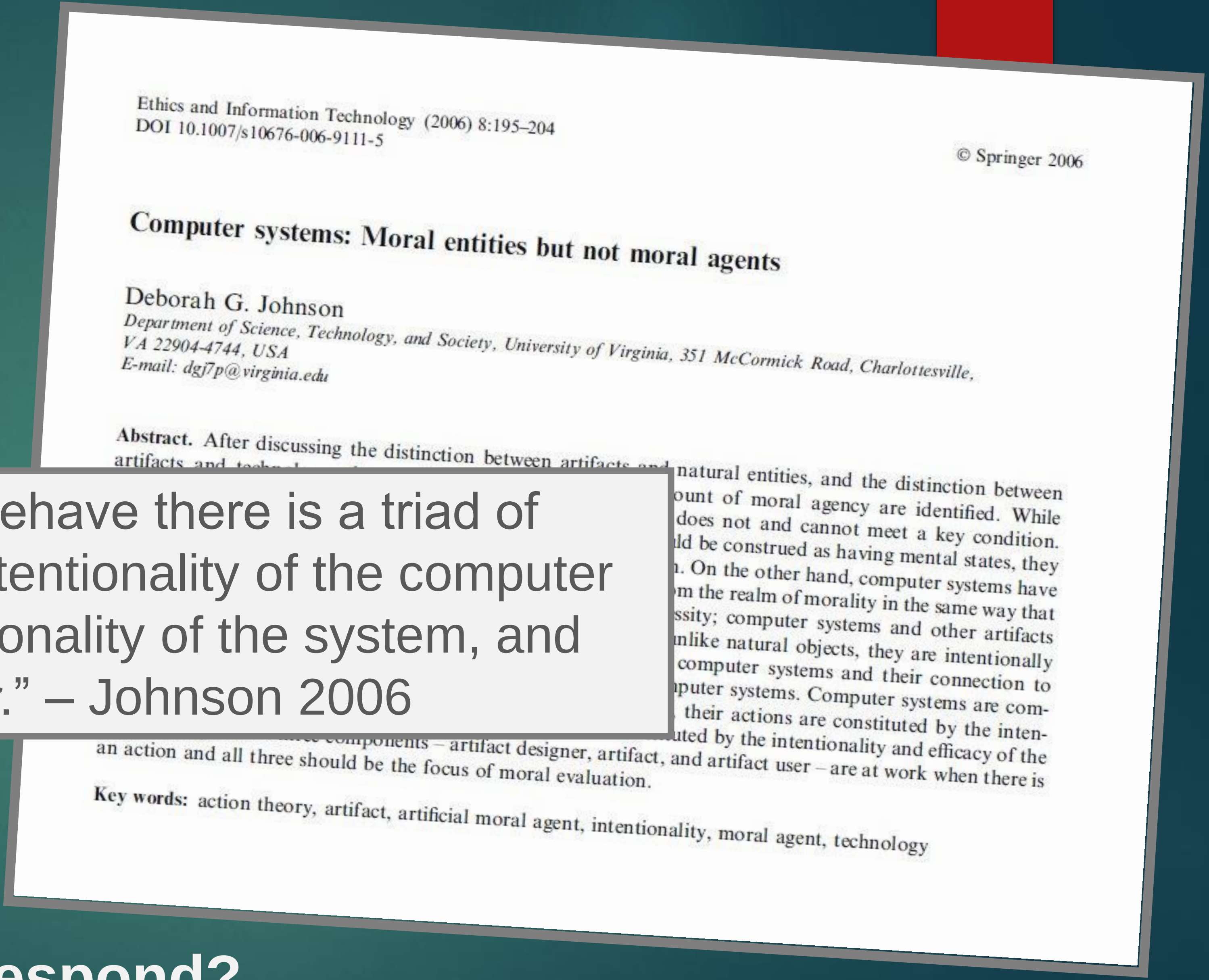
- Hybrid Morality

The Ethics of Things

"I will defend the thesis that ethics should be approached as a matter of human-technological associations. When taking the notion of technological mediation seriously, claiming that technologies are human agents would be as inadequate as claiming that ethics is a solely human affair." – Verbeek 2011

2) How can or should we respond?

- Hybrid Morality



Ethics and Information Technology (2006) 8:195–204
DOI 10.1007/s10676-006-9111-5

© Springer 2006

Computer systems: Moral entities but not moral agents

Deborah G. Johnson
Department of Science, Technology, and Society, University of Virginia, 351 McCormick Road, Charlottesville, VA 22904-4744, USA
E-mail: dgj7p@virginia.edu

Abstract. After discussing the distinction between artifacts and natural entities, and the distinction between
artifacts and
ount of moral agency are identified. While
does not and cannot meet a key condition.
ld be construed as having mental states, they
n. On the other hand, computer systems have
m the realm of morality in the same way that
ssity; computer systems and other artifacts
nlike natural objects, they are intentionally
computer systems and their connection to
puter systems. Computer systems are com-
their actions are constituted by the inten-
uted by the intentionality and efficacy of the
components – artifact designer, artifact, and artifact user – are at work when there is
an action and all three should be the focus of moral evaluation.

Key words: action theory, artifact, artificial moral agent, intentionality, moral agent, technology

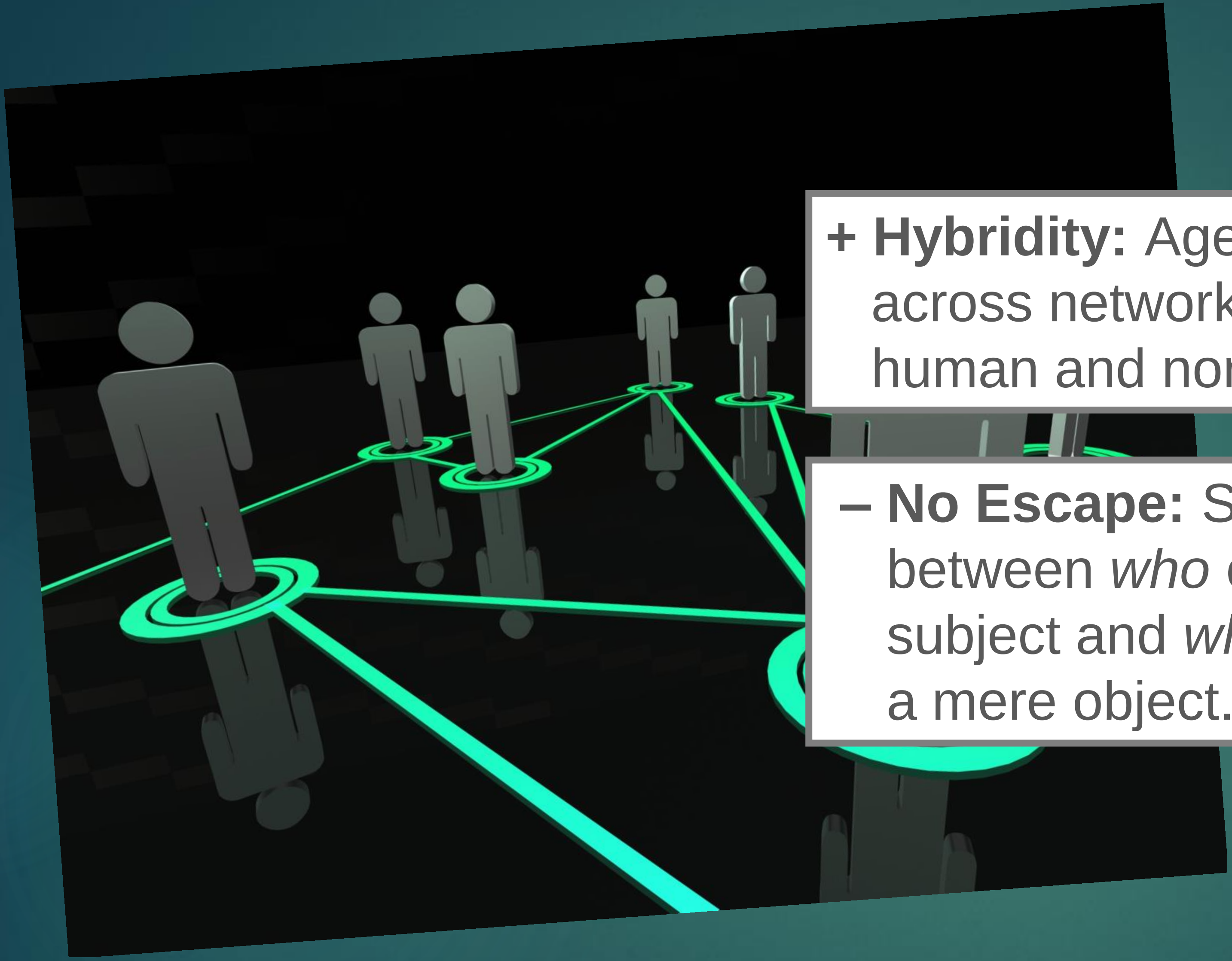
> “When computer systems behave there is a triad of intentionality at work, the intentionality of the computer system designer, the intentionality of the system, and the intentionality of the user.” – Johnson 2006

2) How can or should we respond?

- Hybrid Morality

2) How can or should we respond?

- Hybrid Morality

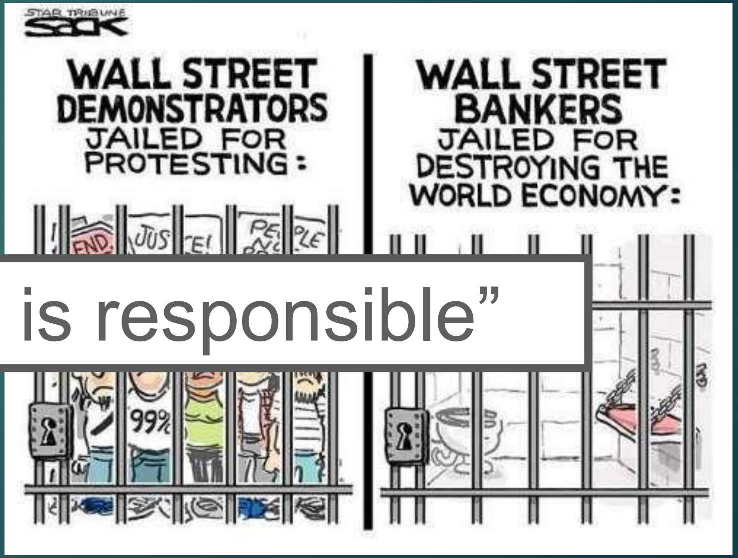


+ **Hybridity:** Agency is distributed across networks composed of both human and non-human elements.

– **No Escape:** Still need to decide between *who* counts as a moral subject and *what* can be considered a mere object.

“Just following orders.”

“A few bad apples”


STAR TRIBUNE
SACK
WALL STREET DEMONSTRATORS JAILED FOR PROTESTING:
WALL STREET BANKERS JAILED FOR DESTROYING THE WORLD ECONOMY:
END
99%

“No one is responsible”

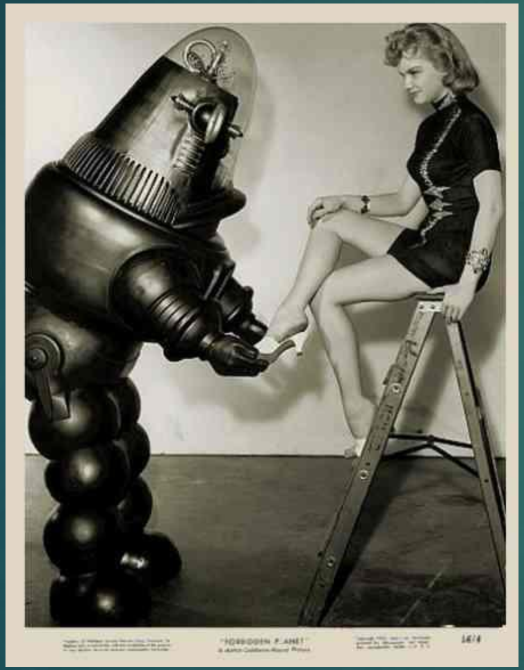
2) How can or should we respond?

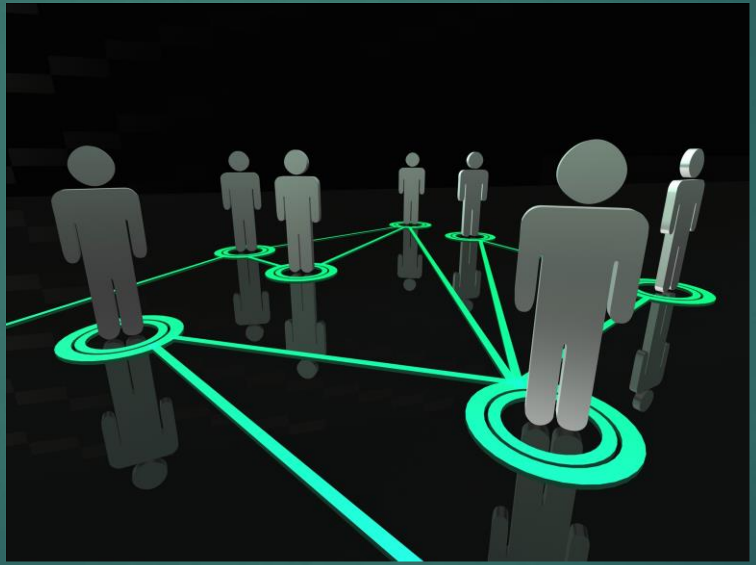
- Hybrid Morality

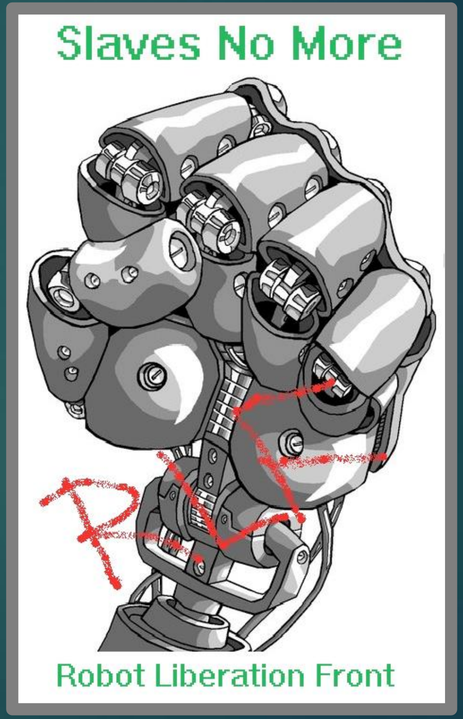
Slavery 2.0

Hybrid Morality

Machine Ethics

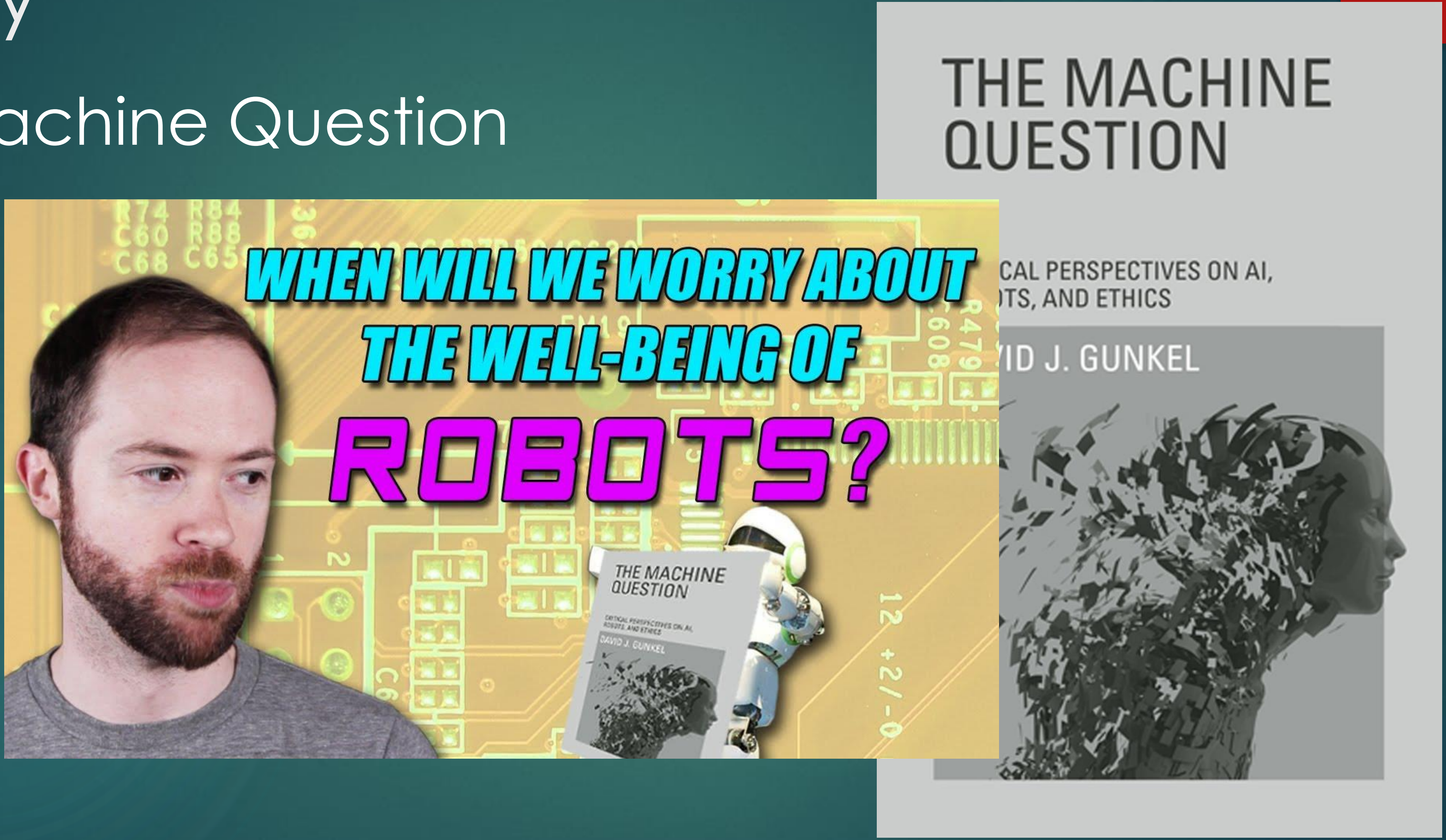






Today

- Machine Question



Chapters 1 and 2

Preview

- How to survive the Robot Apocalypse?
  Or how do you think we can or should respond to or deal with a future where technology is not just a tool or a medium of human action?
  - Content
    - Focus on what you find interesting, promising or worrisome
    - Possibilities: employment, education, social relationships, entertainment, communication, etc.
  - Form (5 minutes)
    - Presentation/Lecture
    - Video
    - Animation
    - Music / Audio Podcast
    - Interactive Game

ちびっこ
あつまれ〜ェ
夏休みは東武の6階！
主催―財団法人日本児童文化研究所 後援―東京都
史上最大 子供の科学展
8月17日〈水〉→28日〈日〉●6階催物場
整理料 おとな・こども 50円
月曜定休
池袋駅西口
東武