Moralities on Intelligent Machines

Michael Laakasuo
Phd.Candidate@CogSci.UH

Moral psychology of robotics is becoming relevant – and fast

www.wired.co.uk/news/archive/2014-05/14/robo-ethics

Robot Morality

Most Visited | Linux Mint | Community | Forums | Blog | News | Nature 433, 807: Con... | Aristotle's Ethics (Sta... | Moral Character (Sta... | Science on the Brink ... | Evolution of Compas... | Altmetric it!

WIRED.CO.UK

US Navy funds morality lessons for robots

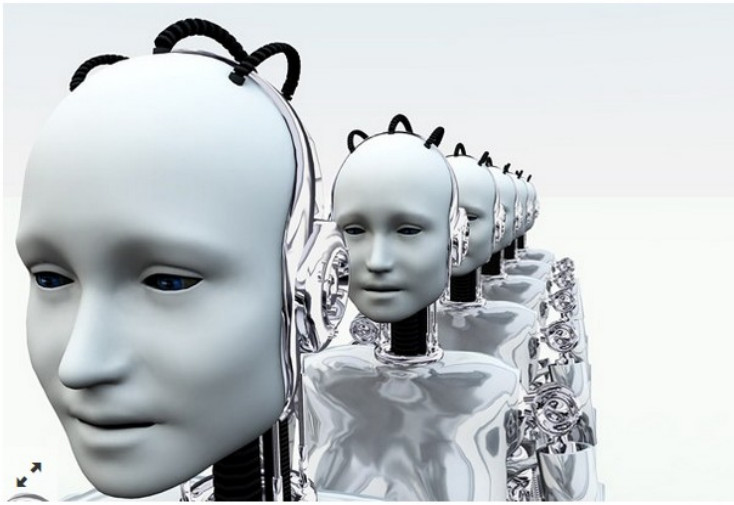
14 MAY 14 / by CHRIS HIGGINS

Tweet 21

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.

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. Matthias Scheutz, principal investigator on the project, and director of the Human-Robot Interaction lab at Tufts, believes that what we think of as a uniquely human trait could be simpler than most of us thought.

"Moral competence can be roughly thought about as the ability to learn, reason with, act upon, and talk about the laws and societal conventions on



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

Moral psychology of robotics is becoming relevant – and fast

yle.fi/uutiset/kuka_korvaa_jos_robotti_mokaa__vakuutusalalla_varaudutaan_uuteen_teknologiaan/7852671?ref=leiki-uu

Most Visited · Linux Mint · Community · Forums · Blog · News · Nature 433, 807: Con... · Aristotle's Ethics (Sta... · Moral Character (Sta

yle UUTISET

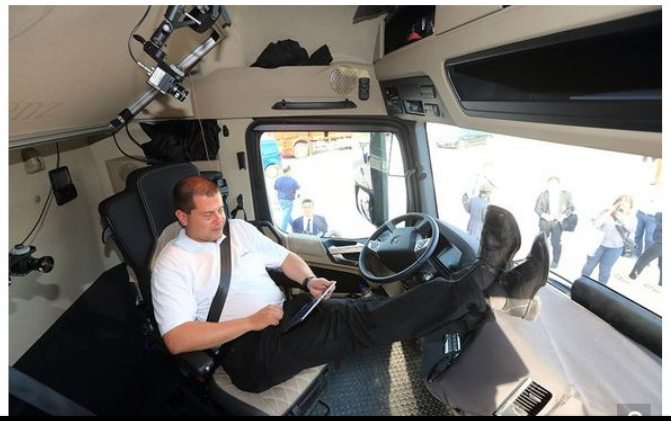
Uutiset + | Urheilu + | Osallistu | Sää + | Tuoreimmat

UUTISET > TALOUS

Talous 8.3.2015 klo 16:51 | päivitetty 8.3.2015 klo 17:09

Kuka korvaa, jos robotti mokaa? – Vakuutusalalla varaudutaan uuteen teknologiaan

Jos automatisoidut autot yleistyvät, liikenneonnettomuuksien määrä todennäköisesti vähenee. Teknikkakaan ei kuitenkaa ole erehtymätöntä. Vakuutusalalla pohditaan nyt, kuka korvaa, jos robotti tekee virheen.



Transferring data from img.yle.fi...

yle.fi/uutiset/suomalainen_robottiosaaminen_nousee_pian_kartalle/7189811

Most Visited · Linux Mint · Community · Forums · Blog · News · Nature 433, 807: Con... · Aristotle's Ethics (Sta... · Moral Character (

Tekniikka 15.4.2014 klo 8:07 | päivitetty 15.4.2014 klo 8:07

Suomalainen robottiosaaminen nousee pian kartalle

Robottiasiantuntijat puuhaavat Suomelle robotiikan tiekarttaa. Se ohjaisi robottien ja automatiikan kehitystä alan kiihtyvässä kansainvälisessä kilpailussa.

Suomi voisi tulevaisuudessa olla maailman johtava hoitorobottien valmistaja, visio Aaltoyliopiston robotiikan ja automatiikan professori Ville Kyrki. Terveydenhuollossa robotti voi raskaan fyysisen työn lisäksi toimittaa lemmikin virkaa, kuten Japanissa kehitetty Paro-hylje tekee. Kuva: Yle

Robotti-intoilijat, tutkijat ja yritysmaailma yrittävät yhdessä saada Suomelle järjestelmällisen ja kansallisen suunnitelman robotiikan kehittämiselle. Eri tahot ovat pohtineet ja kehittäneet robotteja ja automatiikkaa jo vuosikausia, mutta nyt alan asiantuntijat haluavat työstä järjestäytynyttä.

- Robotiikka on nousussa ja moni alan toimija näkee tarpeelliseksi kansallisen tiekartan, Aaltoyliopiston robotiikan ja automaation professori Ville Kyrki perustelee.

Suomen tulevaisuus on palveluroboteissa

ähtisi kansainväliseen robottikilpaan kolmella kärjellä: työkoneilla,

img.yle.fi/uutiset/kotimaa/article7026783.ece/ALTERNATES/w960/robotti+työpaikat.jpg

Moral psychology of robotics is becoming relevant – and fast

More from The Economist | My Subscription | Su

The Economist | World politics | Business & finance | Economics | Science & technology | Culture

Robot ethics

Morals and the machine

As robots grow more autonomous, society needs to develop rules to manage them

Jun 2nd 2012 | From the print edition

Timekeeper

IN THE classic science-fiction film “2001”, the ship's computer, HAL, faces a dilemma. His instructions require him both to fulfil the ship's mission (investigating an artefact near Jupiter) and to keep the mission's true purpose secret from the ship's crew. To resolve the contradiction, he tries to kill the crew.

As robots become more autonomous, the notion of computer-controlled machines facing ethical decisions is moving out of the realm of science fiction and into the real world. Society needs to find ways to ensure that they are better equipped to make moral judgments than HAL was.

In this section

Morals and the machine

Horror in Houla

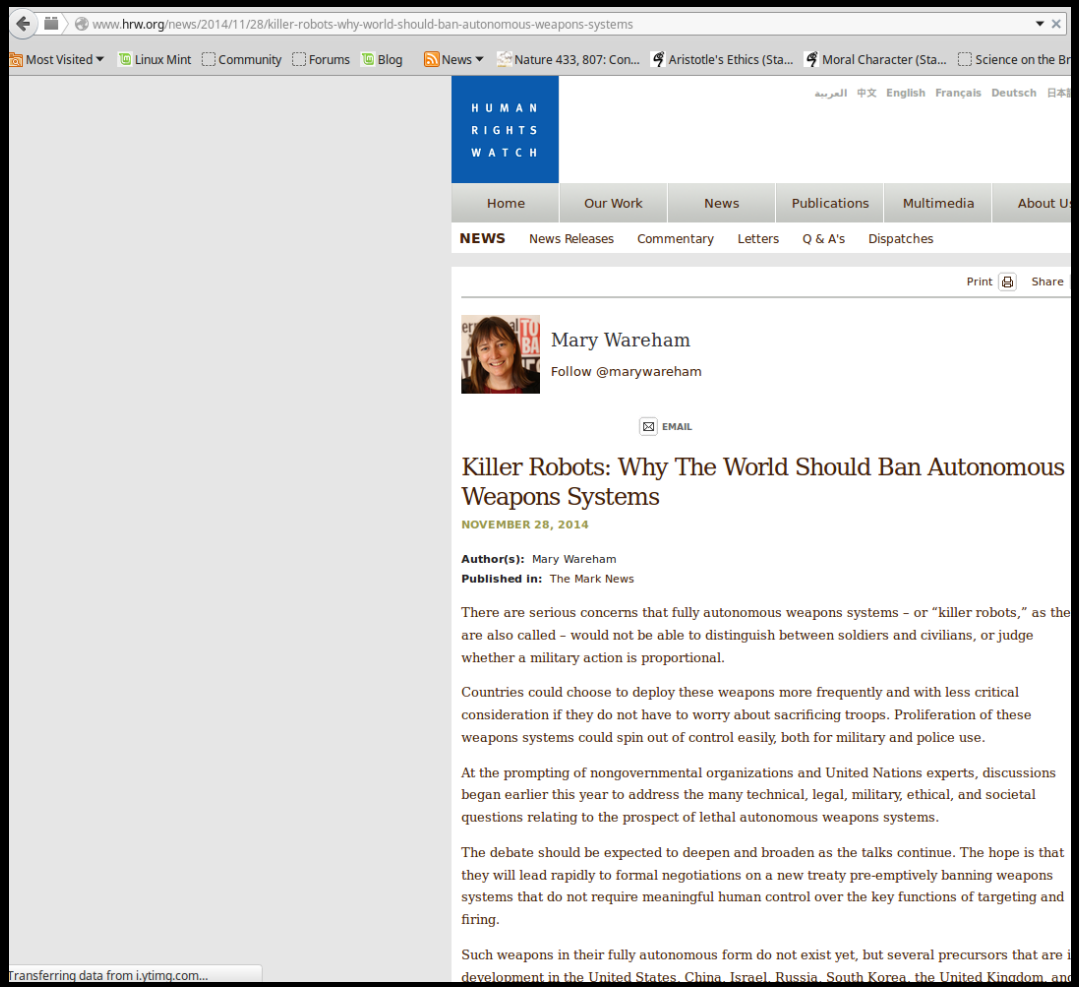
How to save Spain

Moral psychology of robotics is becoming relevant – and fast

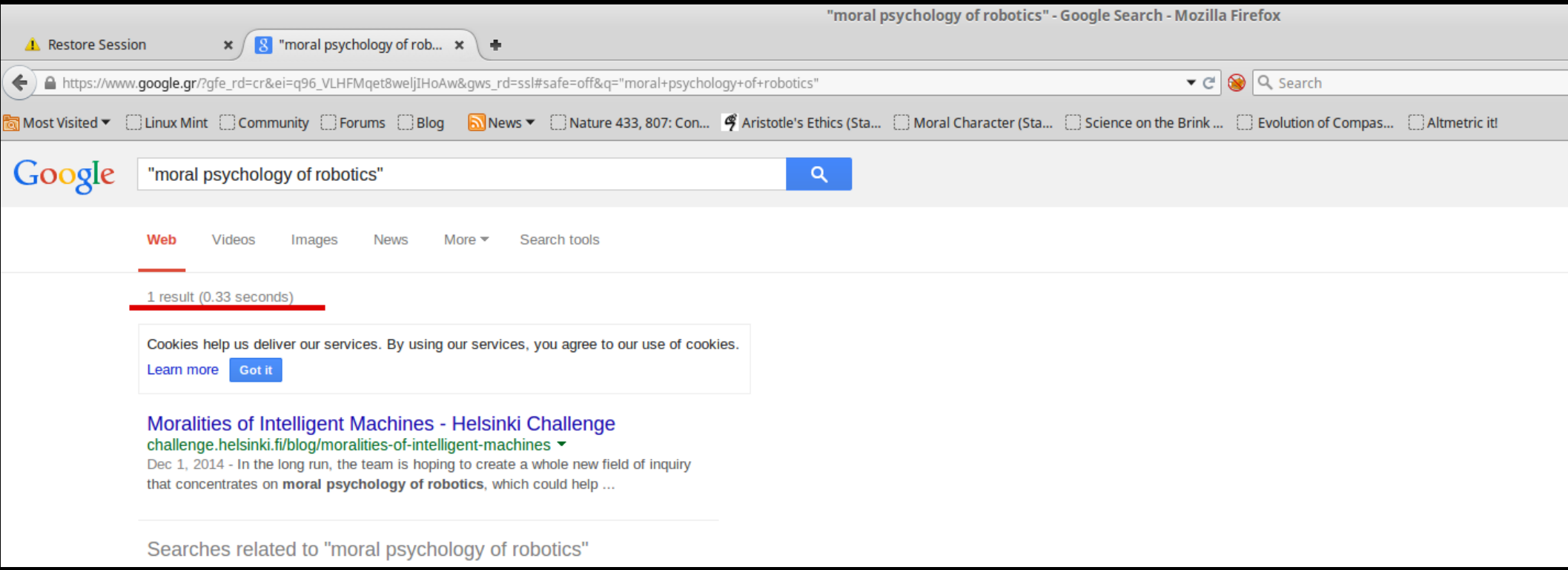
"No one wants to hear that they're building a weapon," says Doug Stephen, a software engineer at the Institute for Human and Machine Cognition (IHMC) whose team placed second at DARPA's event. But he admits that the same capabilities being honed for these trials — ostensibly to make robots good for disaster relief — can also translate to the battlefield. "Absolutely anything," Stephen says, "can be weaponized."

His team's robot, a modification of the humanoid Atlas built by Boston Dynamics, earned the most points in the least amount of time on several challenges, including opening doors and cutting through walls. When it successfully walked over "uneven terrain" built out of cinder blocks, the crowd erupted into cheers. Stephen and his team will now advance to the final stage of the challenge next year — alongside groups from institutes including MIT and NASA — to vie for the $2 million prize.

Moral psychology of robotics is becoming relevant – and fast



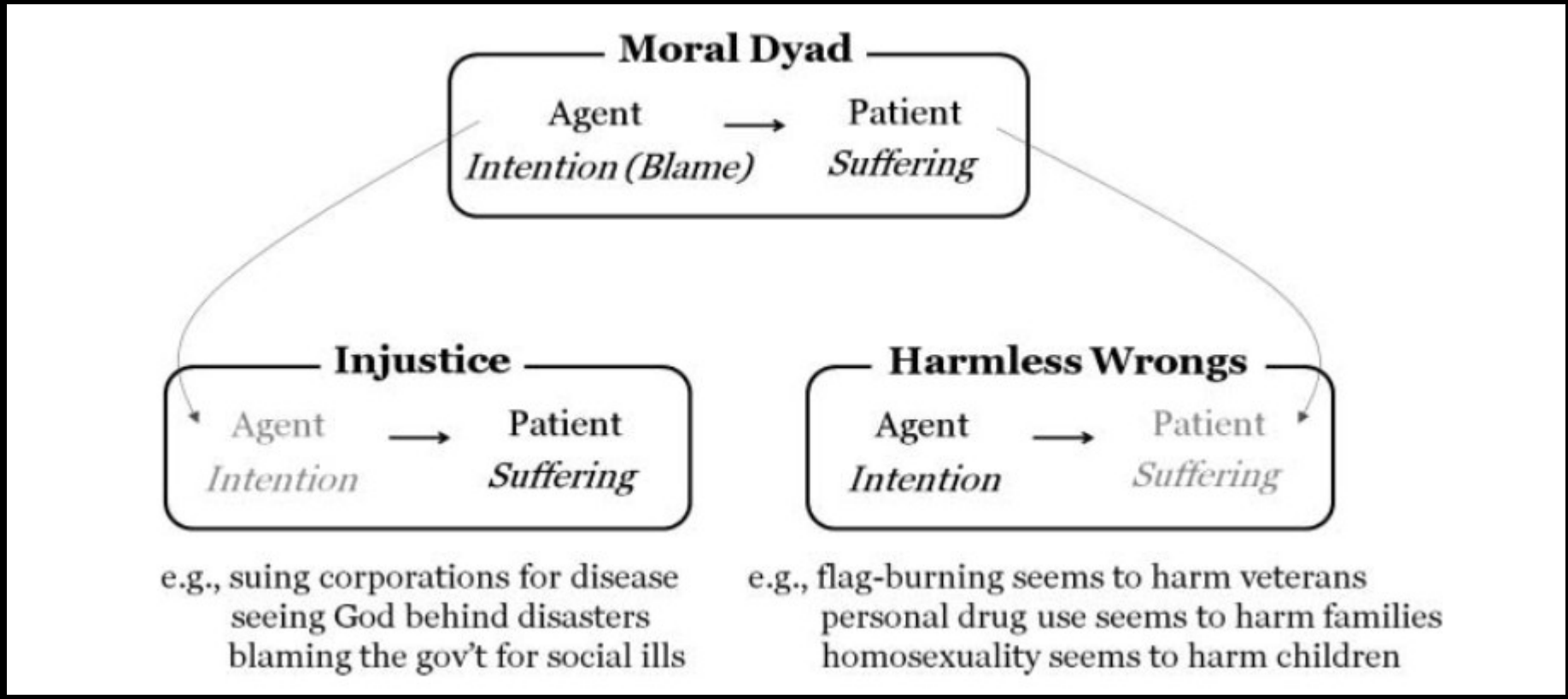
Moral psychology of robotics is becoming relevant – and fast



What motivates us?

- Human psychological architecture is evolutionarily old
  - Our cognition is slow
  - Our cognition is compartmentalized i.e. modular
  - We are not perfectly rational
    - Bounded rationality
  - We did not have artificial moral agents in our evolutionarily environment
    - How does our cognition deal with these issues?

Cognitive template for morality



Central theoretical questions

- How does our stone age cognition go hay-wire:
  - How does our intuition with respect to assigning blame work?
  - How about our need for retribution?
  - How about our perceptions regarding norm obedience?
    - Is it okay for the robot to break the traffic laws?
      - If so, why, when and how? To save lives? How to constrain this?
  -

What are we doing?

- We are answering these questions
  - 1. “How should moral decision making be programmed into robots?”
  - 2. “How does our evolutionarily old human cognition treat robots when it comes to:
    - A) Assigning responsibility
    - B) Having the need for punishment and retribution
    - C) Making standard decisions that are already made by humans?”
- This approach is needed, since there is no consensus or knowledge on how we should create or program robots to make these decisions – no general public discussion has been had about these issues

What do we need funding for?

- Basic research on human moral cognition
  - Build experiments
  - Create stimulus materials
  - Collect data
  - Analyze data
  - Write up research reports
  - Publish in high quality journals

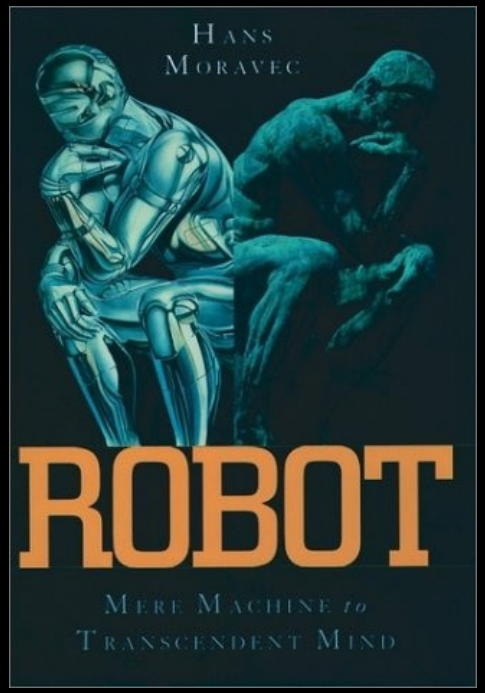
What are our aims?

- Run moral psychological studies by focusing on human-robot interaction in moral situations
- Develop and use realistic virtual 3D environments to advance the field of moral psychology
- Combine philosophy, technology and experimental social psychology to help human kind to build a better future
- Provoke public discussion on the matters
  - This part we have already achieved

Pilot study as an example

Possible future research questions?

- If machine is infected by a virus, whose responsibility is it (from stone age mind perspective)?
  - Do we understand, as people, the agency behind computer viruses?
  - Moravec has reported a case where virus was born spontaneously and lived in the DARPA -net
- Which emotions regulate machine morality
  - Disgust was a candidate
  - Anger motivates retribution, but is it sensible?
- How about mind upload scenarios in AI development?
  - MRI considers this as one of the threat scenarios



Discussion