Socially-responsible research in Cognitive Systems: No-win or win-win?

Tony Prescott
University of Sheffield

Why am I here?

1989

- Participant in numerous EU projects
- Leader of the *Robot Companions for Citizens* Working Group on Society (2011-12)
- Director of Sheffield Centre for Robotics (2011-)

Deliverable D2.4: **Robot Companions: Ethical, Legal and Social Issues**

FET Flagship Preparatory Action **CA-RoboCom**

Coordination **A**ction for the design and description of the FET Flagship candidate **Robo**t **Com**panions for Citizens

Call identifier: FP7-ICT-2011.9.5: FET Flagship Initiative Preparatory Actions

Grant Agreement: FP7-ICT-2011-FET-F # 284951

http://tinyurl.com/rccels

Prescott, Evers, Epton, McKee, Hawley, Webb, Benyon, , Strand, Nourbakhsh, de Cock Buning, Verschure, Dario (2012), *Robot Companions: Ethical, Legal and Social Issues* (CA-Robocom D2.4).

What's on the Horizon?

- Autonomous, potentially self-aware, control systems running critical infrastructure
- Driverless vehicles, robots performing surgery, robots involved in personal care
- Cognitive enhancements—more connected, better memory, better thinking
- Brain-machine interfaces and smart prostheses—a more biohybrid future for humanity

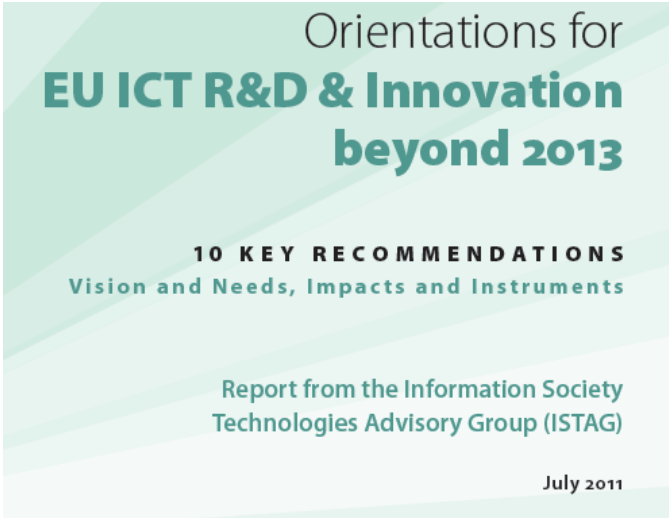
Overview

- The evolving ethical landscape of European science
- Applying ethics in Cognitive Systems research
- An example domain “Robots in Care”

The Evolving Ethical Landscape of European Science

EU R&D Strategy

The last two decades have shown that **ICT is THE key innovation enabler** in almost any technology domain as well as has changed the social behaviour of most of the people in Europe. **Europe must continue to focus on ICT as a key technology area.**



The increasing importance of the social dimension of technical innovation also requires the involvement of new stakeholders. Policy needs to include new, **non-traditional stakeholders who become increasingly important to articulate both technological opportunities and social concerns**.

As the collective interest in technology development and its usage becomes more crucial, **political and ethical research gains importance for ICT research.**

Ethics in Science: Traditional View

- Science performed well is intrinsically **neutral** and **good**
- **Ethics** is compliance with codes and regulations + expert advice
- Undesirable impacts of science are addressed by **corrective action** as and when they occur

Ethics in Science: Contemporary View

- Science is **not** isolated from society and politics
- Ethical science requires the **early anticipation and characterisation of risk** and could lead to decisions not to pursue some lines of research
- Ethics involves the need for **public dialogue**, broader than user involvement
- Integrating ethics into science requires that critical, sceptical and even dissenting voices are **attended to at all stages**

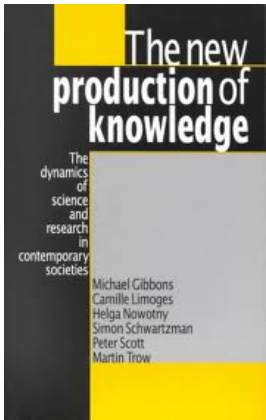
The Changing Nature of Science

Mode 1: Science as *discovery*

- Separate enterprise pursued for its own ends
- Clear disciplinary boundaries
- Scientists and their host institutions (Universities) are autonomous entities

Mode 2: Science as *knowledge production*

- Socially-distributed enterprise pursued for diverse motivations that can include commercial interest
- Application-oriented
- Trans-disciplinary
- Scientists are subject to many accountabilities



Gibbons Michael, Limoges Camille, Nowotny Helga, Schwartzman Simon, Scott Peter and Trow Martin (1994) *The New Production of Knowledge*.

Recognisable Trends

- The 'steering' of research priorities (international, national and institutional)
- The increased linking of science with innovation
- The commercialisation of research (e.g. of intellectual property)
- The accountability of science (e.g. auditing)

Gibbons Michael, Limoges Camille, Nowotny Helga, Schwartzman Simon, Scott Peter and Trow Martin (1994) *The New Production of Knowledge*.

Science—no longer neutral?

"**The single epistemological ideal of a neutral "view from nowhere" has been replaced by multiple views, with each situated somewhere**. The research process can no longer be characterised as an'objective' investigation of the natural (or social) world, or as a cool and reductionist interrogation of arbitrarily defined 'others'.

While scientific excellence (however defined) remains an indispensible criteria, it is obvious that additional criteria – be they economic, political, social or cultural – must be integated as well."

Nowotny, Scott and Gibbons (2003). "Mode 2 Revisited: The New Production of Knowledge".

This view is central to current science policy in the European Union

A 21st century view of science must not only embrace the wider societal context, but be prepared for the context to begin to talk back.

Reliable knowledge will no longer suffice, at least in those cases, where the consensuality reached within the scientific community will fail to impress those outside.

In a 21st century view of science, more will be demanded from science: a decisive shift towards a more extended notion of scientific knowledge, namely a shift towards socially robust or context-sensitive knowledge.

Helga Nowotny
President European Research Council

Public unease with science?

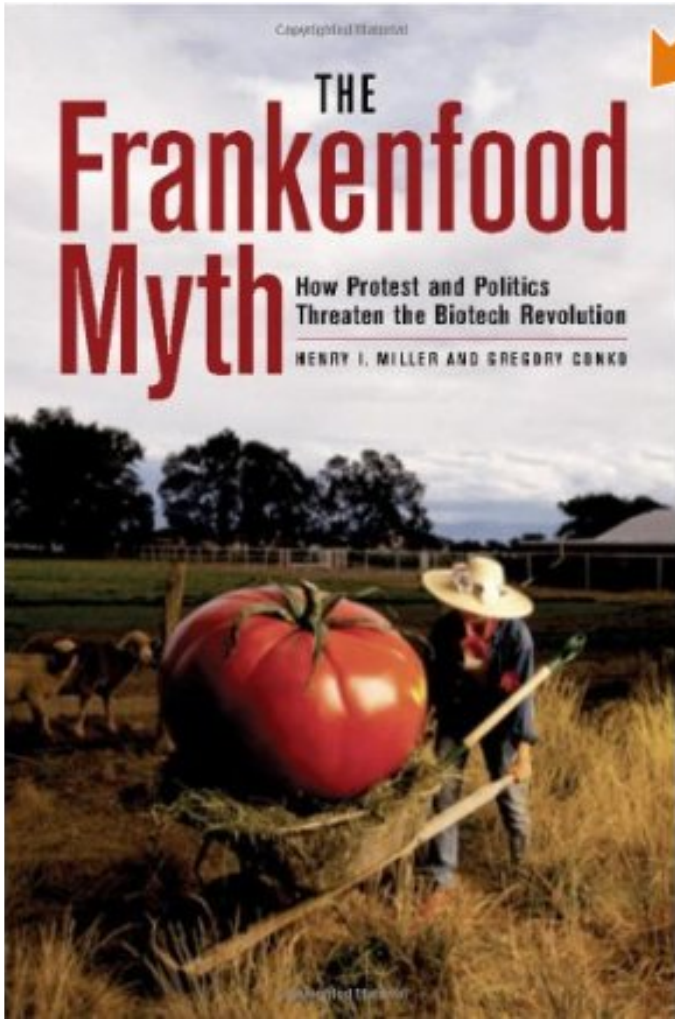
Perhaps the most widely recognised indicator of public unease concerns reactions to issues at the intersection of 'science' (including science-based technologies) and 'risk'. The public is thought to fear science because scientific innovations entail risk. Both science and risk, however, are ambiguous objects. It is frequently assumed in policy circles that the meanings of both for citizens must be the same as for experts, but that assumption is, in our view, itself a key element in generating 'public unease'. The widespread sense of unease – sometimes expressed as 'mistrust of' or 'alienation from' science – must be seen in broader perspective. We conclude indeed that there is no general, indiscriminate public disaffection with nor fear of 'science'. **Instead, there is selective disaffection in particular fields of science, amidst wider areas of acceptance – even enthusiasm.**

Wynne et al., 2007

TAKING EUROPEAN KNOWLEDGE SOCIETY SERIOUSLY

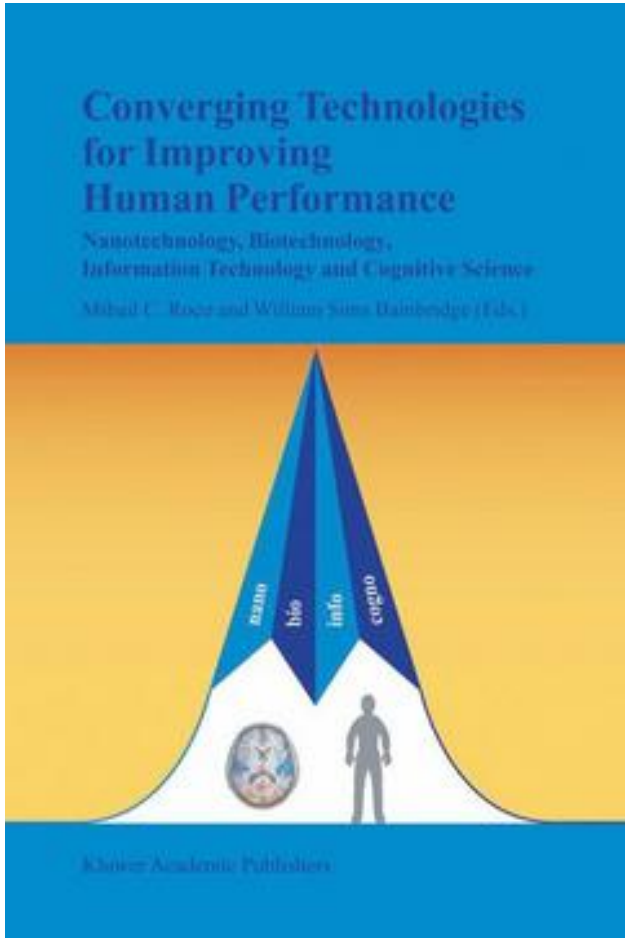
Report of the Expert Group on Science and Governance to the Science, Economy and Society Directorate, Directorate-General for Research, European Commission

The Risks of Getting in Wrong

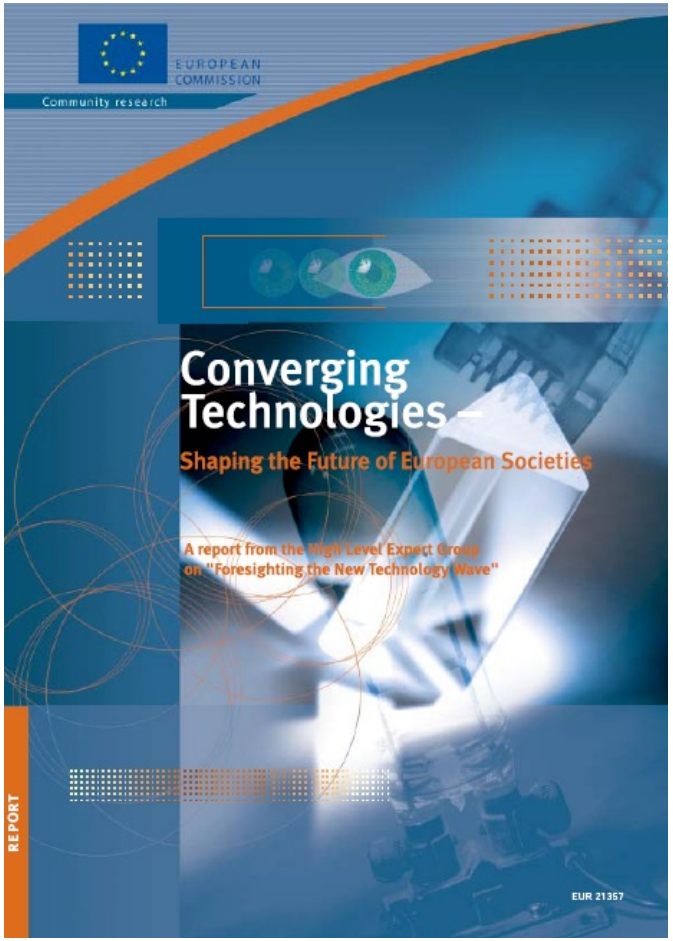


Technologies arising from Cognitive Systems research are potentially more controversial than GM.

Converging Technologies: Contrasting Views



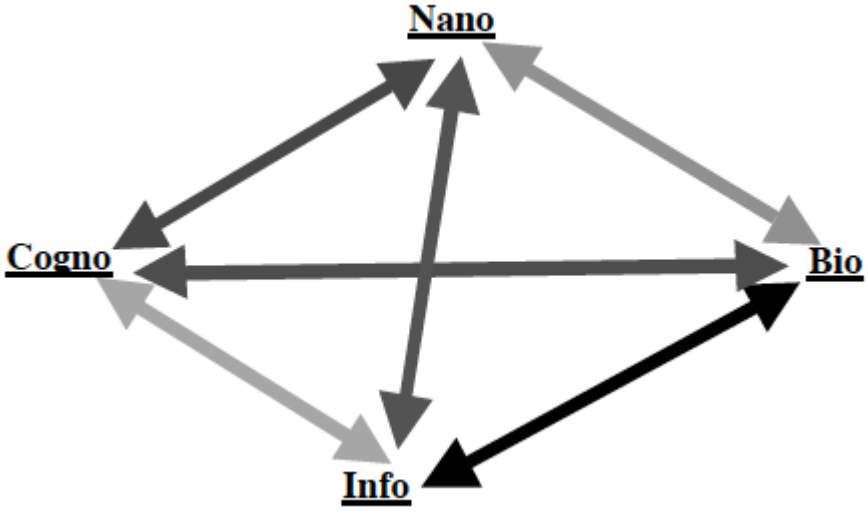
2001 NSF-sponsored conference summarised in Roco & Bainbridge 2002



Foresight exercise commissioned by the EC Directorate, written by Nordmann et al. 2004

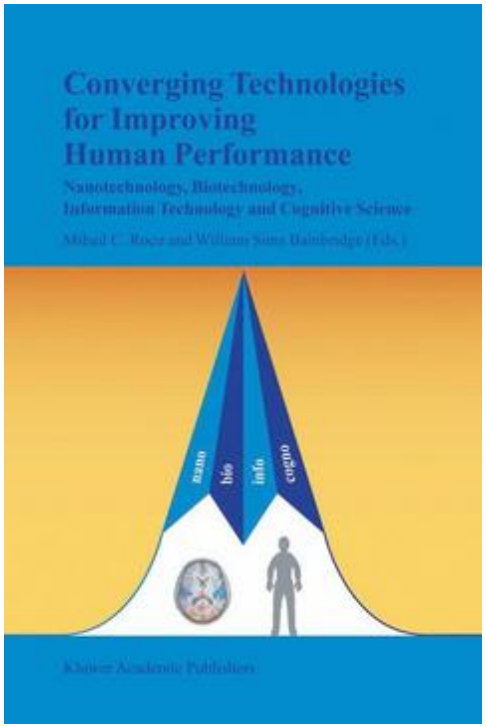
What are the Convergent Technologies?

The phrase “convergent technologies” refers to the synergistic combination of four major “NBIC” (nano-bio-info-cogno) provinces of science and technology, each of which is currently progressing at a rapid rate: (a) nanoscience and nanotechnology; (b) biotechnology and biomedicine, including genetic engineering; **(c) information technology, including advanced computing and communications; (d) cognitive science, including cognitive neuroscience.**



The Optimist's Vision

- Call for accelerating investment in CTS
- Co-ordinated action from different sectors
- Technological human enhancement for both civilian and military purposes
- A "New Renaissance" based on a holistic understanding of nature



“If the Cognitive Scientists can think it,
the Nano people can build it,
the Bio people can implement it,
and the IT people can monitor and
control it.”

W. A.Wallace quoted in Roco and
Bainbridge 2002a, 13)

Confidence in Dealing with Risk

We understand how to make decisions in the face of uncertainty. We know how to use knowledge about the world to predict the future, to reason, imagine, and plan actions to achieve goals. We have algorithms that can decide what is desirable and plan how to get it. We have procedures to estimate costs, risks, and benefits of potential actions.We can write computer programs to deal with uncertainty and compensate for unexpected events.
James Albus, quoted in Roco and Bainbridge (2002b, 292)

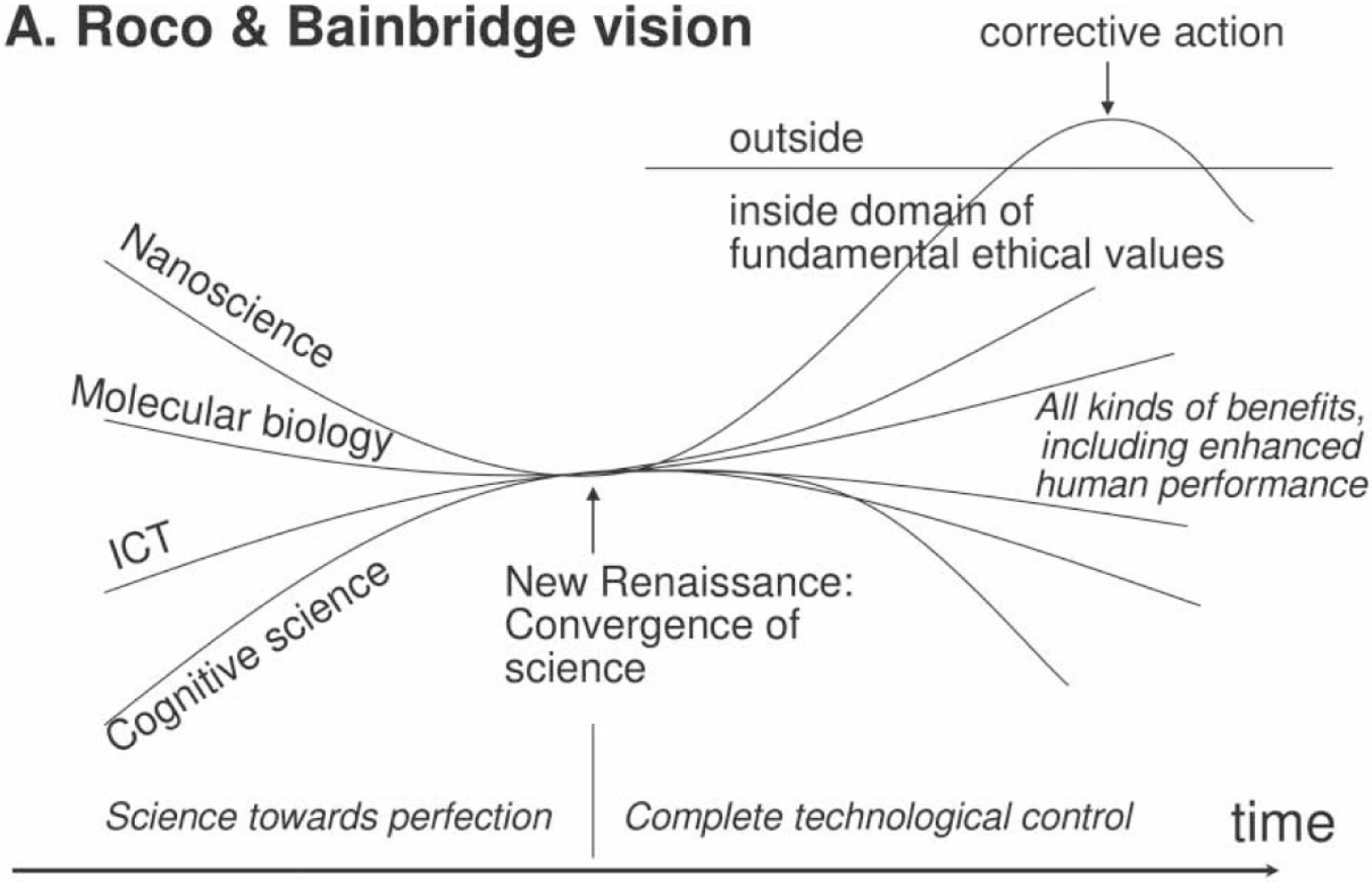
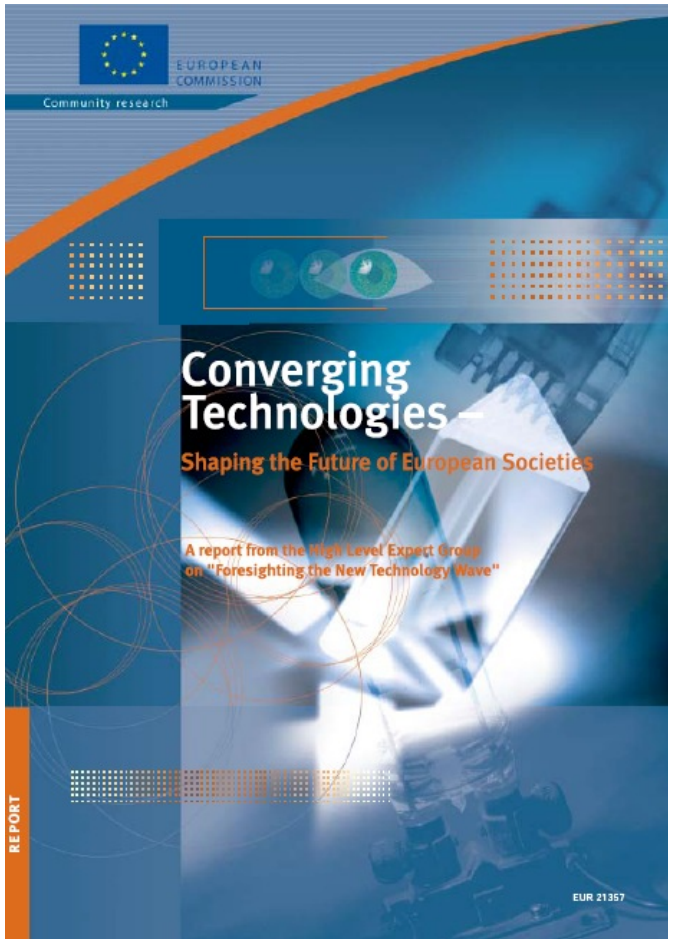


Diagram from Kjølberg , Delgado-Ramos, Wickson & Strand (2008) Models of Governance for Converging Technologies. *Technology Analysis & Strategic Management*, 20:1, 83-97, DOI: 10.1080/09537320701726676

The Cautionary Response

- Takes R&B report as starting point
- Calls for strong societal involvement
- Is cautious about "enhancing human performance" as a goal
- Prefers specific goals benefiting European society determined through dialogue



Converging Technologies for the European Knowledge Society (CTEKS)

B. CTEKS vision

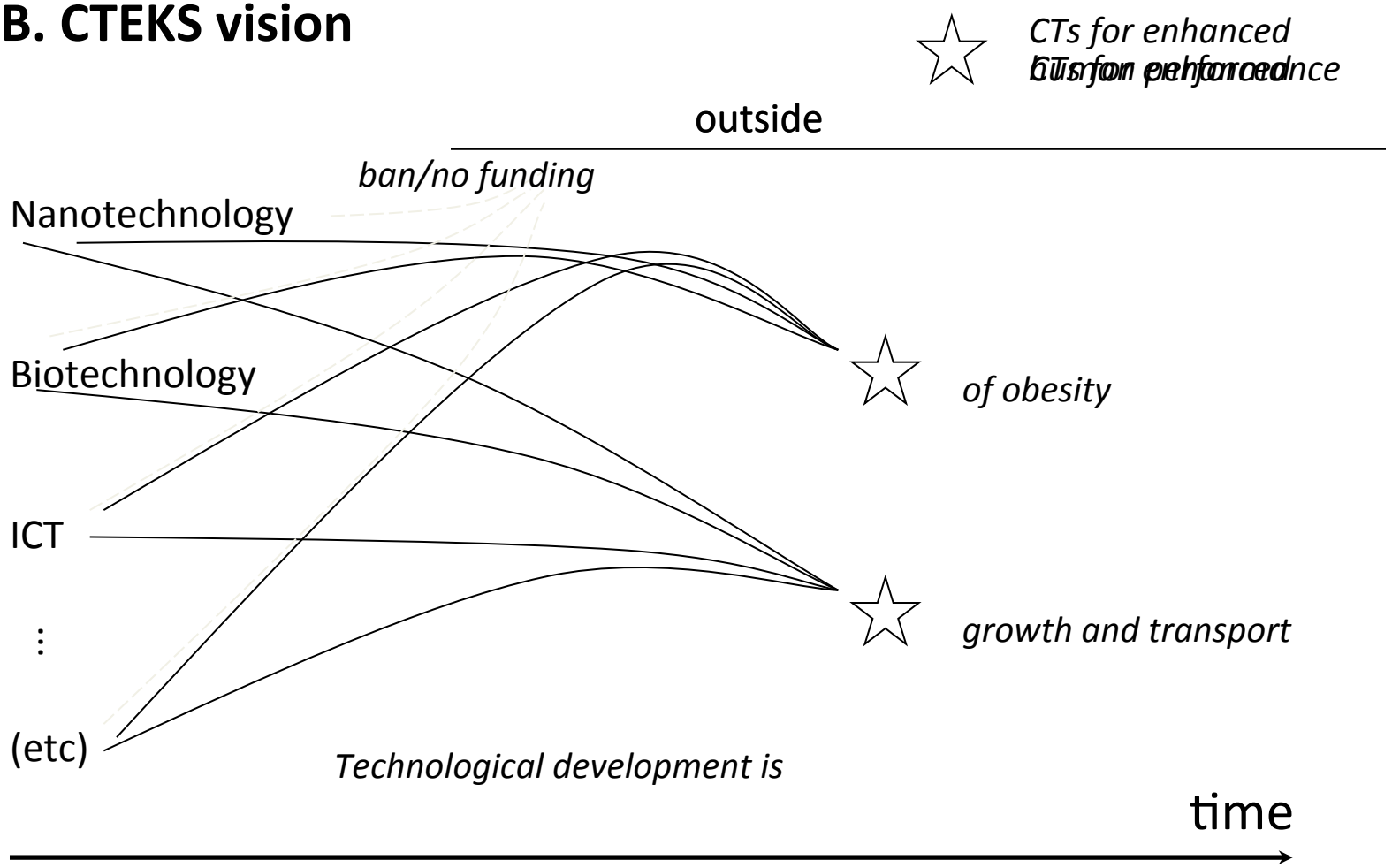


Figure from Kjølberg , Delgado-Ramos, Wickson & Strand (2008) Models of Governance for Converging Technologies. *Technology Analysis & Strategic Management*, 20:1, 83-97, DOI: 10.1080/09537320701726676

Applying Ethics in Cognitive Systems Research

Scientists have to think about risk

"against a background of inherent uncertainty about the future
"against a background of inherent uncertainty about the future state of knowledge (and of almost everything else) from which, of course, scientific potential is derived, it is necessary to reach beyond the knowable context of application to the unknowable context of implication. Here **knowledge producers have to reach out and anticipate reflexively the implications of research processes**."

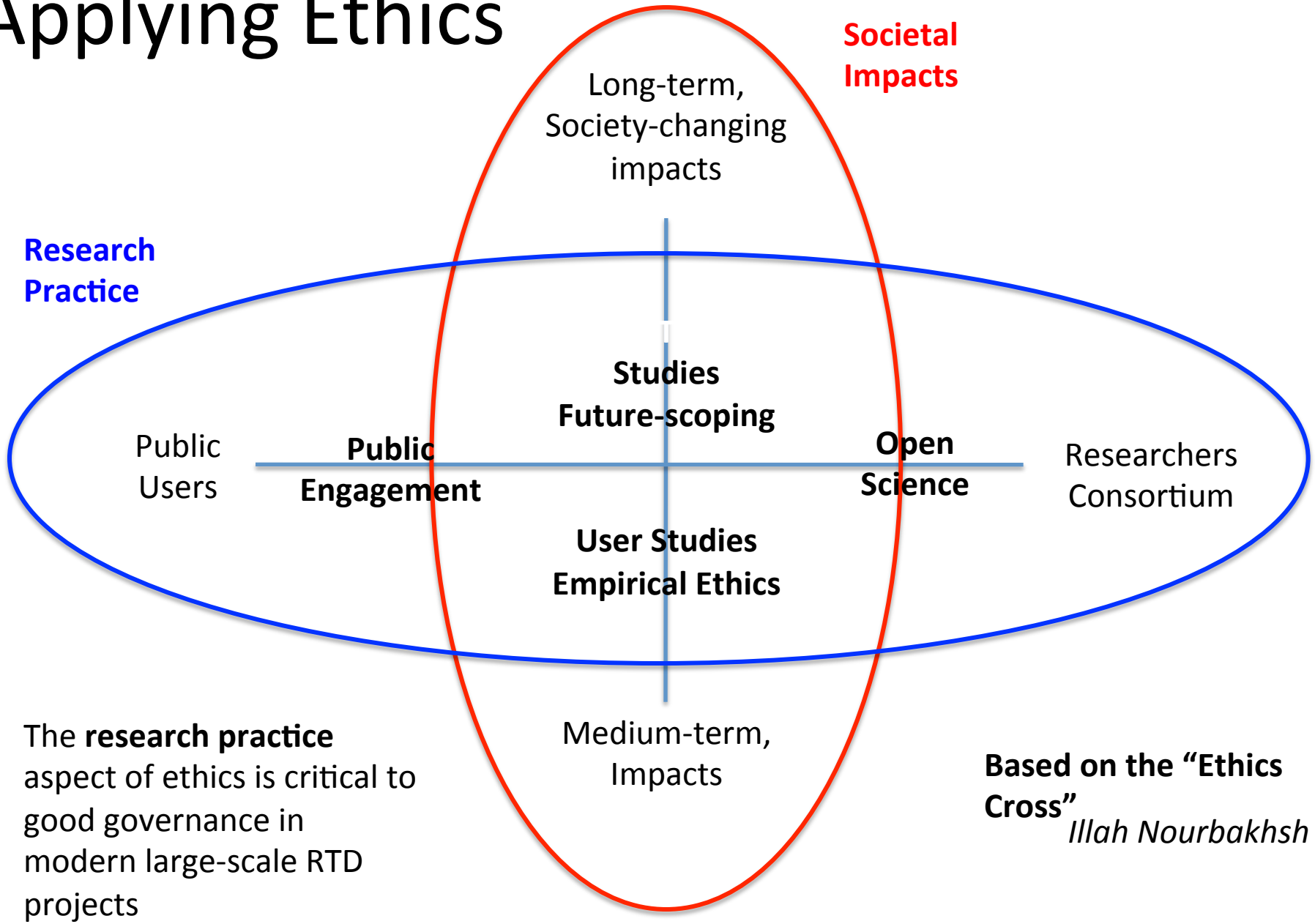
Nowotny
, Scott and Gibbons (2003). "Mode 2 Revisited: The New

And Engage in More Dialogue

sufficiently fostering a meaningful exchange.”

EURAB: European Research Advisory Board

Applying Ethics



The **research practice** aspect of ethics is critical to good governance in modern large-scale RTD projects

Based on the "Ethics Cross"
Illah Nourbakhsh

Prescott et al. (2012), *Robot Companions: Ethical, Legal and Social Issues* (CA-Robocom D2.4).

A model from nanotechnologies

COMMISSION OF THE EUROPEAN COMMUNITIES

Brussels, 07/02/2008
C(2008) 424 final

COMMISSION RECOMMENDATION

of 07/02/2008

on a code of conduct for responsible nanosciences and nanotechnologies research

3.1 Meaning

N&N research activities should be comprehensible to the public. They should respect fundamental rights and be conducted in the interest of the well-being of individuals and society in their design, implementation, dissemination and use.

3.2 Sustainability

N&N research activities should be safe, ethical and contribute to sustainable development serving the sustainability objectives of the Community as well as contributing to the United Nations' Millennium Development Goals[11]. They should not harm or create a biological, physical or moral threat to people, animals, plants or the environment, at present or in the future.

3.3 Precaution

N&N research activities should be conducted in accordance with the precautionary principle, anticipating potential environmental, health and safety impacts of N&N outcomes and taking due precautions, proportional to the level of protection, while encouraging progress for the benefit of society and the environment.

3.4 Inclusiveness

Governance of N&N research activities should be guided by the principles of openness to all stakeholders, transparency and respect for the legitimate right of access to information. It should allow the participation in decision-making processes of all stakeholders involved in or concerned by N&N research activities.

3.5 Excellence

N&N research activities should meet the best scientific standards, including standards underpinning the integrity of research and standards relating to Good Laboratory Practices[12].

3.6 Innovation

Governance of N&N research activities should encourage maximum creativity, flexibility and planning ability for innovation and growth.

3.7 Accountability

Researchers and research organisations should remain accountable for the social, environmental and human health impacts that their N&N research may impose on present and future generations.

techn01life

a Transdisciplinary approach to the Emerging CHallenges of NOvel technologies: Lifeworld and Imaginaries in Foresight and Ethics

- A methodology to map ethical issues at early stages of S&T and policy development and to represent "social imaginaries" relating to these ethical issues

Co-ordinated by **Roger Strand** of the Centre for the Study of the Sciences and the Humanities, University of Bergen (Partners in Tartu, Manchester, Paris, Lancaster)

http://www.technolife.no

- Steps are:
  - Identify “hot topics”
  - Engage in a participatory deliberative exercise with citizens stakeholders
  - Online voting procedure
  - Quantitative analysis to identify arguments, concerns, imaginaries, alternative reference frames

http://www.technolife.no

TechnoLife debate: Body & Mind Enhancement

Watch video at:
http://technolife.no/short_movies/ or
https://www.youtube.com/watch?v=STiuB7nQn1w

- Hot Topics:
- a) The relationship between normality and perfection
- b) Freedom of choice and social difference
- c) Change in the life-cycle and life-span of individuals and the human species
- Most participants "strongly in favour of ensuring that enhancements be implemented within pluralistic and diverse societies and value systems."
- Social justice a key issue

From *TechnoLife Final Report*

http://www.technolife.no

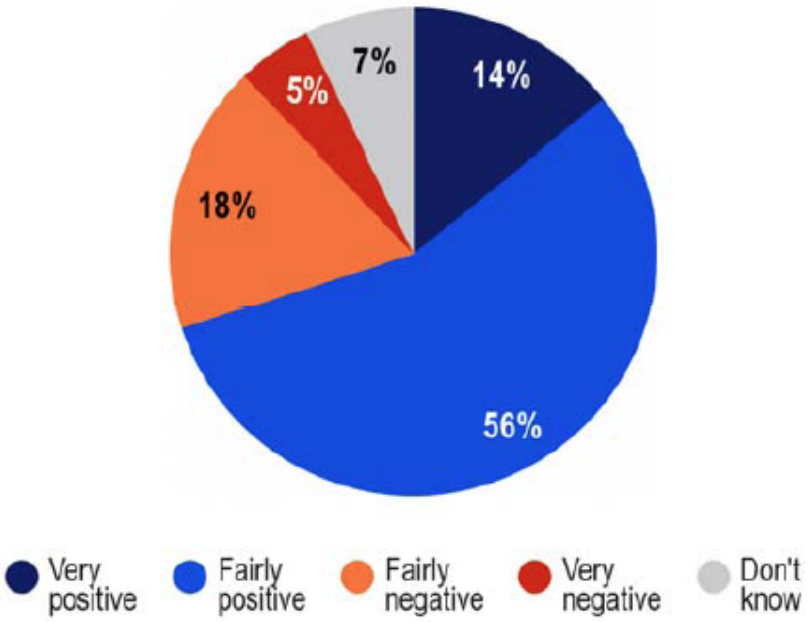
Example domain: The risks and benefits of using robots in care

Prescott (2013), Sunny uplands or slippery slopes? The risks and benefits of using robots in care, UKRE Workshop on Robot Ethics, 25th March 2013, Sheffield, UK.

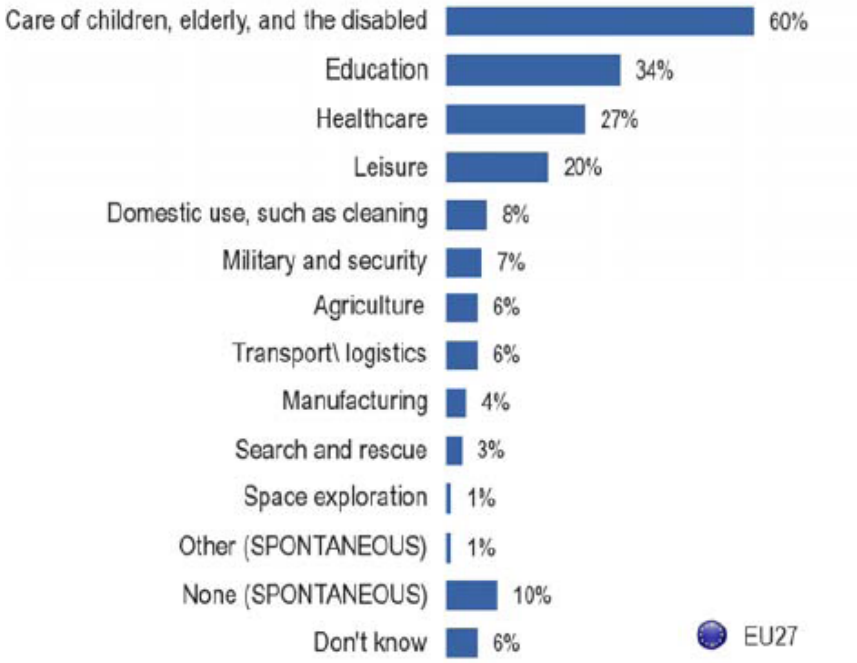
http://tinyurl.com/robotisincare

Public Attitudes

QA4. Generally speaking, do you have a very positive, fairly positive, fairly negative or very negative view of robots?



QA7. And on the other hand, in which areas do you think that the use of robots should be banned?



European Commission

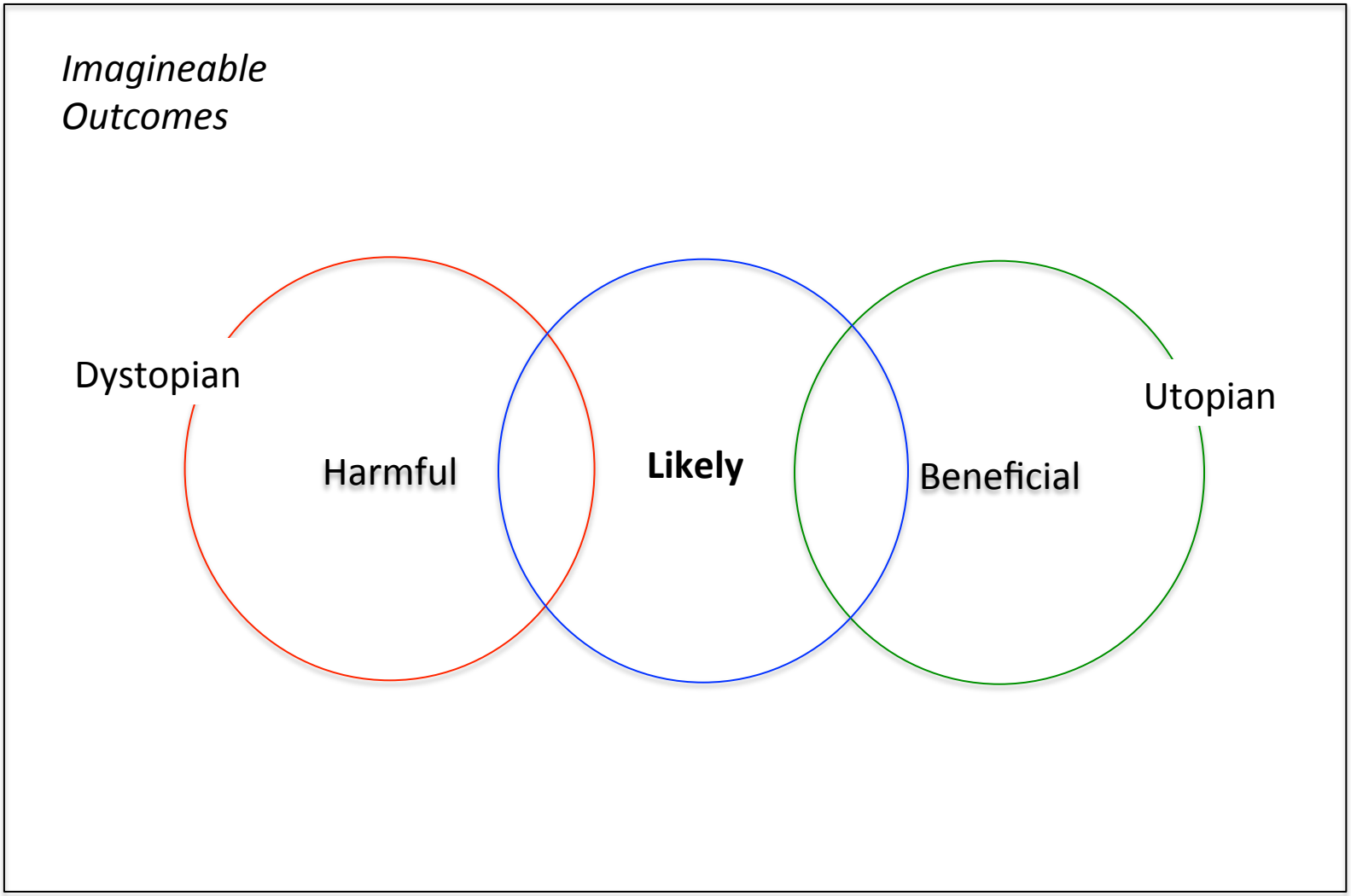
EU Barometer Survey 382, March 2012

27 EU countries, 26,751 respondents

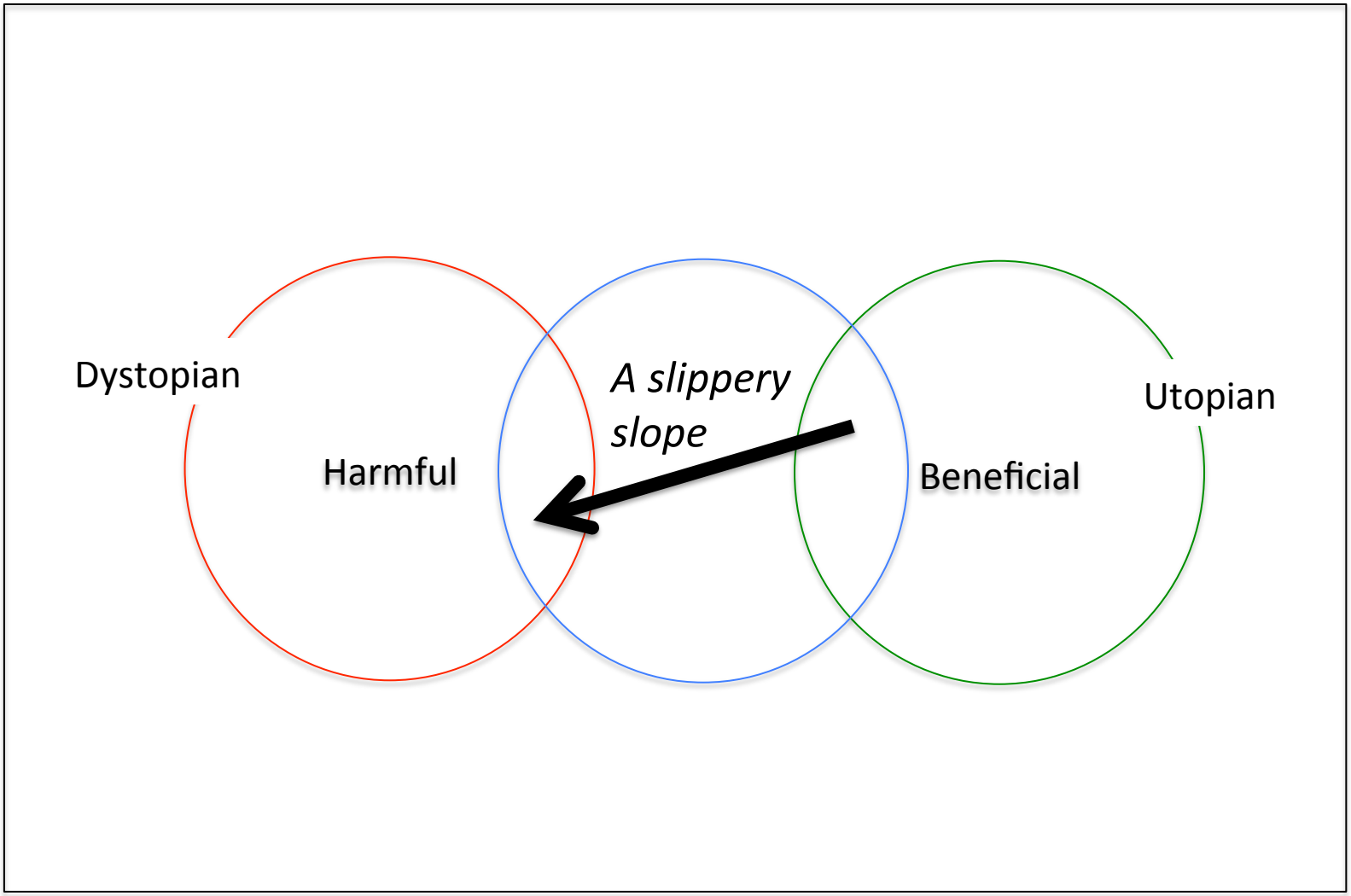
Robots and social isolation

“It is likely that success in introducing robots into the aged-care sector will be at the expense of the amount of human engagement available to frail aged persons. [...] It is our view that handing over cleaning and other household tasks to Robocare, Rosie, Yumel, Wakamaru, or Mentorbot—or their equivalent—would therefore most likely be detrimental to the well-being of frail older people.” (Sparrow and Sparrow, “In the hands of machines? The future of aged care”, Minds and Machines 2006)

Dystopias, utopias & plausible futures



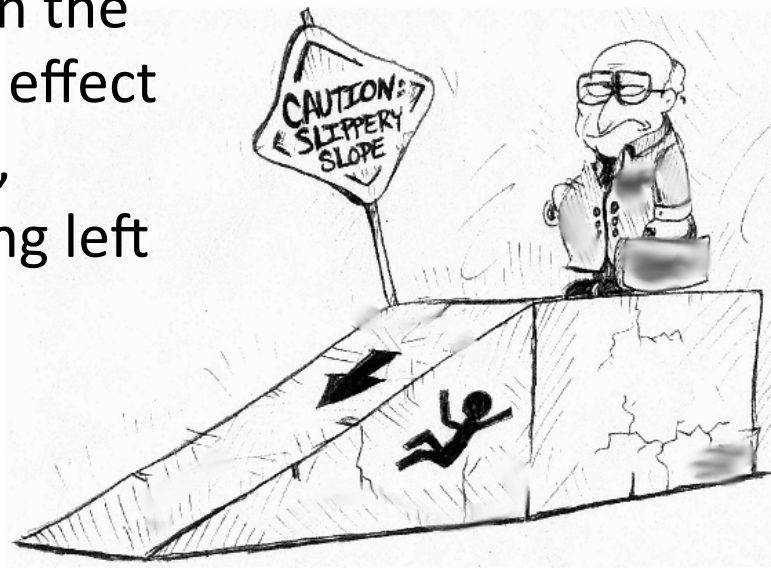
Slippery slopes



Slippery slope logic

- An action that by itself seems harmless, establishes a trajectory that, with high probability, leads to an unintended and unethical final outcome

- Example: Introducing robots for help in the care of older people has the knock-on effect of reducing contact with other people, ultimately leading to older people being left "in the hands of machines".



Validity of slippery slopes

- Slippery slopes can be valid but only if their proponents can establish a clear causal chain, also known as the warrant
- The problem is that often this causal chain is not adequately justified, or, that the links along the chain have some probability less than one and that therefore the cumulative likelihood of the worst case outcome is actually much lower than supposed

To evaluate the real likelihood of slippery slopes we have to take into account defeaters that could prevent the causal chain from being enacted

Warrant for the slippery slope of robot care

"Even if robots were to become capable of filling some service roles in the aged-care sector, **economic pressures on the sector** would most likely ensure that the result was a decrease in the amount of human contact experienced by older persons being cared for, which itself would be detrimental to their well-being"

(Sparrow and Sparrow, 2006)

What is required for this to come about?

(i) Robots will need to be sufficiently autonomous as to not require the supervision of human care staff

(ii) The use of robots will not lead to more effective deployment of human care that compensates any reduction in the number of human carers.

(iii) Older people themselves will take no significant action to compensate for any loss in social contact due to the introduction of care robots.

(iv) Governments, companies, charities, and other bodies responsible for social care of older people, will consider that human-human contact is sufficiently unimportant that older people can be left increasingly in the care of robots.

Robot autonomy

(i) Robots will need to be sufficiently autonomous as to not require the supervision of human care staff

It is unlikely, in the foreseeable future, that robots will be sufficiently autonomous to provide care unaided. It is more likely that care will be provided by robot-human teams

Robots and human-human care

(ii) The use of robots will not lead to more effective deployment of human care that compensates any reduction in the number of human carers

Caring roles could become more professional, carers could have more time not less for human-human contact; decoupling physical care from social care can allow the latter to be provided where it is most effective

Older people can adapt

(iii) Older people themselves will take no significant action to compensate for any loss in social contact due to the introduction of care robots.

At a time when we are expecting increasing numbers of older people, who will stay active longer, and who will be better connected through use of ICT technologies, it seems paradoxical that they will not respond to changes in the nature of care, such as the introduction of robots, by taking actions themselves to help meet their social needs

Older people will adapt

"Computer literate, often umbilically tied to the Internet for work and play, they [ageing baby boomers] may find the transition to a world of virtual caring relationships and social life at a distance exceptionally congenial."

Sorell and Draper, 2012

Societal inaction

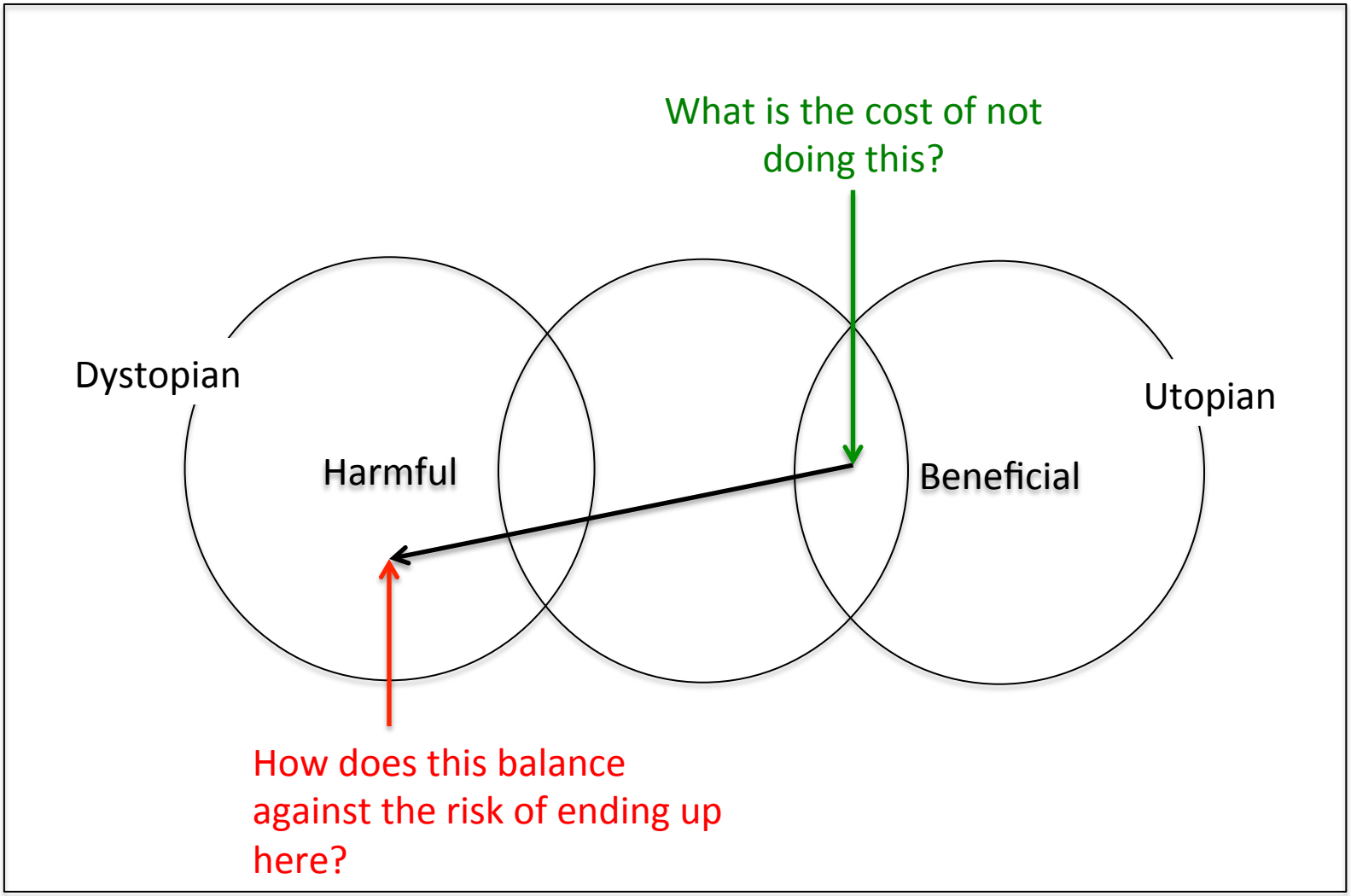
(iv) Governments, companies, charities, and other bodies responsible for social care of older people, will consider that human-human contact is sufficiently unimportant that older people can be left increasingly in the care of robots.

Developed countries will have a much higher percentage of older people than they do now. Older people, as a constituency, already have significant political power, and the demographic shift will only give them a stronger voice

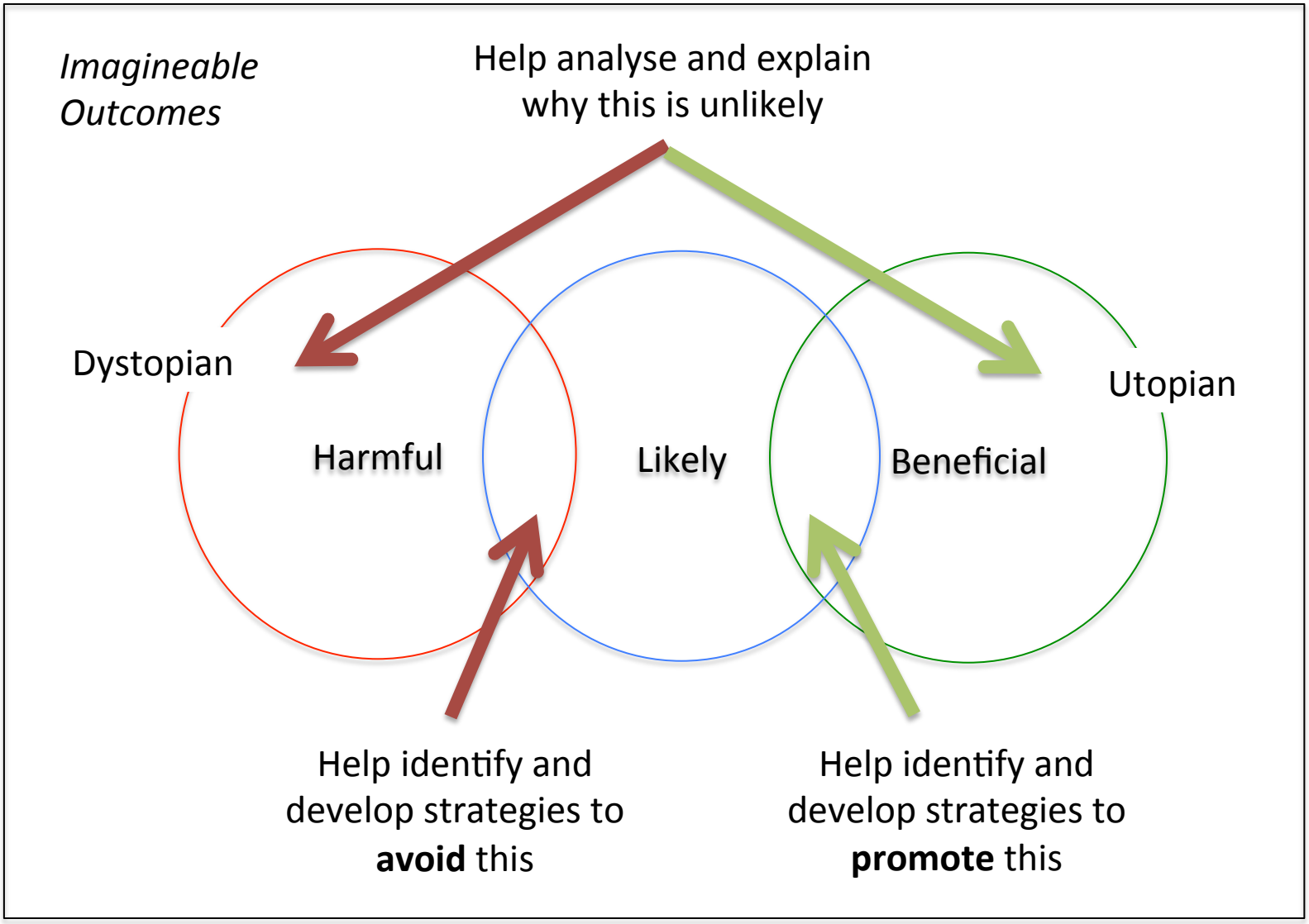
Getting the balance right

if, as the Eurobarometer survey suggests, people are already turning against the possibility of robot carers, even before they exist in any real way, then this is a significant worry—**the costs of *not* developing these technologies to future human welfare, might actually outweigh the risks**

Getting the balance right



A Role for Researchers?



Robots can help carers

- Robots can
  - Help address the shortage of skilled labour in the caring professions
  - Assist carers to be more efficient
  - Help older people assist each other
  - Allow carers to focus more on the human-to-human aspects of their work, creating a more skilled profession and making care-work more rewarding
  - Relieve some of the injuries that happen to carers through the physical aspects of their work

The introduction of robots will <u>not</u> solve all societal problems raised by the demographic shift, which must also be addressed through wider political and social actions, but they can provide part of the solution

Table 1 Reasons why elderly people need more care and support, and the role

No.	Trigger factor	Psychological	Physical
1	A fear of falling		
2	A major health event – such as support following a stroke or hip replacement		
3	A perceived decline and concern for own health		
4	A person feeling lonely		
5	Abuse (physical or mental)		
6	Bereavement of a family member or friend		
7	Cognition impairment (such as dementia, confusion or memory loss)		
8	Consequences of admission to hospital		
9	Depression, mental breakdown or deterioration		
10	Deteriorating physical functioning		
11	Difficulty cooking for themselves		
12	Difficulty in managing stairs or steps		
13	Difficulty toileting/continence management		
14	Family, friends or neighbours can no longer provide support to maintain the person		
15	Family/caregiver stress		
16	Housework problematic		
17	Inability to care for self at home		
18	Inability to cope with Independent Activities of Daily Living*		
19	Inadequate home care provision		
20	Managing pressure sores		
21	Medication management – such as compliance problems		

UN Convention on the Rights of Disabled People

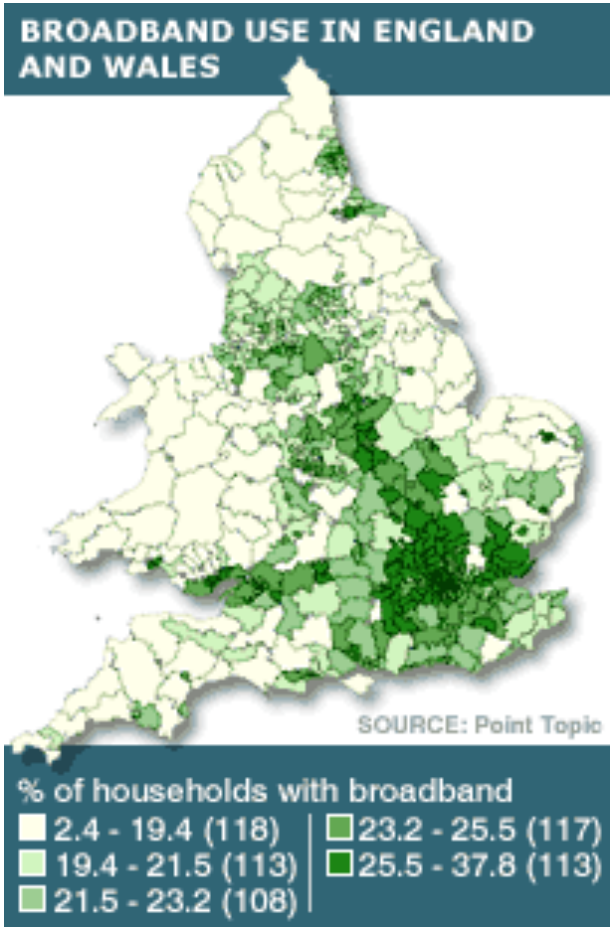
- Article 4 of the UN Convention makes it an obligation for research and development into new technologies to occur, stressing that development should take into account cost (i.e. ensuring products and technologies are reasonably and affordably priced) and that information about these products is made available to their intended users.

Joanne O'Riordan "Build me a Robot"

Watch vIdeo at:
http://www.youtube.com/watch?v=xRk2TeaC_Qw

The robot divide?

- A risk we should consider is that, like the digital divide, robots for care will only be available for the better-off in our societies



Roadmap

Robot Companions For Citizens: Roadmapping The Potential For Future Robots In Empowering Older People

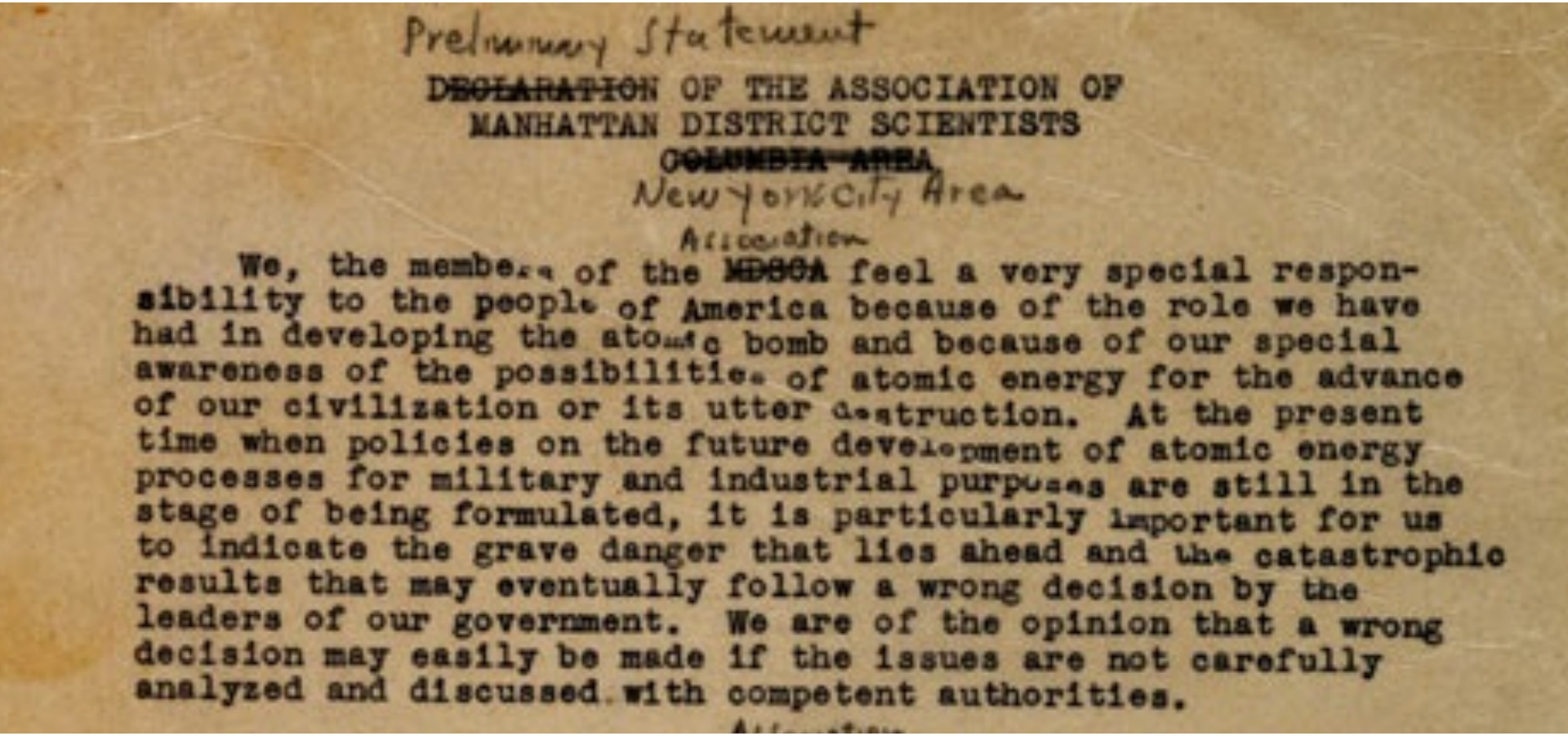
Tony J Prescott[1], Tracy Epton[1], Vanessa Evers[2], Kevin McKee[3], Mark Hawley[1], Thomas Webb[1], David Benyon[4], Sebastian Conran[5], Roger Strand[6], Madeleine de Cock Buning[7], Paul Verschure[8], Paolo Dario[9] and the Robot Companions for Citizens Society Community Working Group

[1] University of Sheffield, United Kingdom
[2] University of Twente, Netherlands
[3] Dalarna University, Sweden
[4] Napier University, United Kingdom
[5] Conran Design Associates LLP, United Kingdom
[6] Bergen University, Norway
[7] University of Utrecht, Netherlands
[8] University of Pompeu Fabra, Spain,
[9] Scuola Superiore Sant'Anna, Italy

http://tinyurl.com/braidrobots

Conclusion: no-win or win-win?

- Ethics is about understanding risk in science and dealing with its implications
- Ethics in Cognitive Systems research should not be left to expert “ethicists”. Multiple sectors of society must engage in the debate, including researchers themselves
- Scientists will be held accountable for the results of their research
- If Cognitive Systems researchers engage-in/lead the debate this can lead to better research whose consequences are to benefit of everyone



Preliminary Statement

~~DECLARATION~~ OF THE ASSOCIATION OF
MANHATTAN DISTRICT SCIENTISTS
~~COLUMBIA AREA~~
New York City Area
Association

We, the members of the ~~MDSCA~~ feel a very special responsibility to the people of America because of the role we have had in developing the atomic bomb and because of our special awareness of the possibilities of atomic energy for the advance of our civilization or its utter destruction. At the present time when policies on the future development of atomic energy processes for military and industrial purposes are still in the stage of being formulated, it is particularly important for us to indicate the grave danger that lies ahead and the catastrophic results that may eventually follow a wrong decision by the leaders of our government. We are of the opinion that a wrong decision may easily be made if the issues are not carefully analyzed and discussed with competent authorities.

The Preliminary Statement of the Association of Manhattan District Scientists, New York City Area, circa 1945