Future Computing and Robotics:
A Report from the HBP Foresight Lab

Authors:	Nikolas ROSE, KCL (P32);
	Christine AICARDI, KCL (P32);
	Michael REINSBOROUGH, KCL (P32)

Table of Contents

List of Figures

Executive Summary

The **Human Brain Project (HBP)** is one of the Future and Emerging Technology Flagship initiatives funded by the European Commission. It is a ten-year initiative in medicine, neuroscience and computing which brings together scientists and institutions from 20 nations across Europe, and has a strong element of international cooperation. The first 30 months of the Project, the so-called 'Ramp-Up' Phase, has run from October 2013 to March 2016. This report by the HBP Foresight Lab at King's College London (Work Package 12.1) is the last in a series of three foresight reports delivered in the Ramp-Up Phase, each dedicated to a domain to which the HBP aims to contribute: medicine, neuroscience, and now information & communication technology (ICT) and robotics.

As computing is increasingly integrated into all aspects of societies, economies and everyday life, and as robotic machines play an ever larger role in industrial production and elsewhere, carrying out many tasks that were previously only possible by skilled human labour, there is a growing public debate about the ways these developments are influencing different aspects of our societies, our economies and our lives. Lately, such developments have led to a growing amount of speculation about so called 'intelligent machines' with some suggesting that we will soon reach the 'singularity', the point at which machine intelligence overtakes human intelligence. The view that we take in this report is that such speculations are, at the very least, premature, and divert our attention from more pressing social and ethical issues that are arising in connection to the spread of not-so-intelligent machines. It is these less dramatic, and less speculative, implications that we focus on.

The HBP Subprojects can potentially contribute to many areas of ICT and robotics, leading to a variety of future applications and cutting across many domains. Instead of a piecemeal inventory of possible products, we adopt a holistic approach, looking at hardware and software, machines and humans, as parts of larger systems. Our goal is to identify social and ethical challenges related to the potential contributions of the Project to future ICT and robotics.

We first single out two cross-cutting topics: intelligent machines and human-machine integration. Then we look beyond the direct contributions of the HBP research and explore potential social and economic challenges that more general developments in ICT and robotics may bring, focusing in particular on affective computing and the impact on jobs and the economy.

On the basis of our Foresight work, we draw the following conclusions:

- We consider that it is misleading for participants in the HBP to insist on a restrictive definition of Artificial Intelligence and thus to claim that the HBP need not engage with the social and ethical issues raised by developments in Artificial Intelligence. Indeed a number of HBP Sub-Projects will collectively contribute to the broader field of machine intelligence. Hence there is a need for HBP researchers and directors to work with the Ethics and Society Subproject and to participate in current debates that address the need to make Artificial Intelligence ethical and socially beneficial.

- There is a need for short term social and ethical issues of specialised Artificial Intelligence to be demarcated from speculation about long term potential existential risks, and these short and medium term issues should be addressed as apriority, at national and trans-national levels. These debates should not be left to the private sector and private philanthropy, but should be addressed openly and democratically.

- The human should not be 'designed out' of technology, instead the human should be put at the centre. A human-centred design approach that does not narrowly focus on the individual but takes into account the wider socioeconomic context, can bring to

light a broader, and different, range of social and ethical issues. Our strong view is that such an approach should inform strategic choices and decisions driving research and innovation for future computing and robotics.

- No technology that uses the brain as its inspiration should neglect the capacity of the brain, in constant interaction with its interpersonal and physical environment, to develop and sustain a social and emotional mind.

- While there are obvious benefits of robots in the domain of care - especially for older people, those with disabilities, and perhaps for children - care is a human interaction involving genuine reciprocation of feelings and obligations, and these entail the kind of high level affective and interpersonal skills that are currently challenging for robots. It is important that policy makers recognise the limits of robots for such work, as well as their benefits.

- Given the contribution of the demand for ICT materials to geopolitical rivalries and of ICT hardware to electronic waste, those urging and funding the development of ICT and robotics should do so within a clear analysis of the life cycle of the material artefacts that are proliferating, and devote attention and funding to the development of more sustainable and ecologically desirable alternatives.

- Given the commitment of the European Union to enhancing the wealth and wellbeing of its citizens, it is necessary for policy makers at European level to pay attention to the medium term economic consequences of the kinds of developments that might result from the translation to applications of the research of the HBP and similar work in advanced ICT and robotics.

We recommend the following in relation to the HBP:

- Given the many interdependencies between different parts of the Project regarding potential contributions to future ICT and robotics, the Directorate and the Science and Infrastructure Board of the HBP should conduct a systematic, Project-wide review to take stock of all these synergetic potentials and devise a Responsible Research and Innovation roadmap for building on them.

- Given that a number of Subprojects will collectively contribute to the domain of machine intelligence, we recommend that HBP researchers engage with current initiatives and debates that consider ways to make Artificial Intelligence ethical and socially beneficial. To facilitate this, we recommend that the HBP nominate one of its members to act as coordinator on these questions.

- Horizon 2020, which will fund the operational phase of the HBP, aims to position Europe as a world-class competitor in ICT research and digital innovation through 'open science' and the development of innovative public-private partnerships. Those responsible for the scientific direction of the HBP should set out policies that maintain a commitment to openness throughout, and seek to ensure that the benefits of research and development are not 'privatized' but are used in a manner that is beneficial to European citizens.

- With regard to the projected offering of commercial services to industry in particular by the Neurorobotics Platform, an evaluation of proposed applications in terms of social and ethical impact should be an integral part of the terms of service.

- The scientific directors of the HBP should re-iterate their commitment to responsible research and innovation by supporting the development of a coherent programme of researcher awareness, engaging the whole HBP research community, during the operational phase of the project.

1. Introduction

As computing is increasingly integrated into all aspects of societies, economies and everyday life, and as robotic machines play an ever larger role in industrial production and elsewhere, carrying out many tasks that were previously only possible by skilled human labour, there is a growing public debate about the ways that new applications of information and communication technologies (ICT) and robotics are influencing our lives. The impacts on all the organizations and institutions of our modern world are now widely recognised. Power stations, industry, urban food distribution, economic markets, our schools, hospitals, transportation networks, communication systems, and police, civil and military defence infrastructure are all in some way organized by or influenced by computing, and by robotic technologies controlled by computers. As for everyday life, the most obvious revolution relates to the role of mobile telephones, which have been transformed into 'devices' that not only embody previously unimaginable miniaturised computing powers, but link wirelessly into the internet and can connect to multiple previously dumb devices, such as home heating systems and domestic machinery, guide us through unfamiliar territory via connections with geographic information systems, and now are able to respond to some voice instructions and sometimes to answer our questions in spoken language. The power of these innovations has been increased by the development of data mining and machine learning techniques, that give computers the capacity to learn from their 'experience' without being specifically programmed, constructing algorithms, making predictions, and then improving those predictions by learning from their results, either in supervised or unsupervised regimes. In these and other ways, developments in ICT and robotics are reshaping human interactions, in economic activities, in consumption and in our most intimate relations.

Lately, such developments have led to a growing amount of speculation and hype about so called 'intelligent machines'. Some have speculated that we will soon reach the 'singularity', the point at which machine intelligence becomes capable of autonomous learning and self-improvement, including the capacity to build even more intelligent machines, generating a runaway process that will rapidly overtake human intelligence, and indeed operate in ways quite beyond human understanding.[1] The view that we take in this report is that such speculations are, at the very least, premature, and divert our attention from more pressing social and ethical issues that are arising in connection to the growing spread and efficiency of not-so-intelligent machines. It is these less dramatic, and less speculative, implications that we focus on in this report.

1.1 Future Computing and Robotics in the HBP

Information and Communication technologies play multiple roles in the vision of the HBP. On the one hand, ICT is a tool to enable the better integration of neuroscientific knowledge and the building of models and simulations: as the HBP Website puts it, the HBP "aims to achieve a multi-level, integrated understanding of brain structure and function through the development and use of information and communication technologies (ICT). These technologies will enable large-scale collaboration and data sharing, reconstruction of the brain at different biological scales, federated analysis of clinical data to map diseases of the brain, and the development of brain-inspired computing systems. Through the HBP's ICT platforms, scientists, clinicians, and engineers will be able to perform diverse experiments and share knowledge with a common goal of unlocking the most complex structure in the universe. The development and use of ICT over the HBP's 10-year lifetime will pave the way for the project's ultimate goal, simulation of the whole human brain."[2] On the other, the HBP's High Performance Computing Subproject (SP 7) seeks, not only to provide resources for the work of other parts of the HBP, but also to become a major driver for the development of ICT in Europe, and to undertake "path-breaking ICT technology research. Key topics for research will include novel accelerator technologies

addressing highly scalable computational challenges, the use of hierarchical storage-class memory to increase available memory by more than an order of magnitude per core, and the realisation of interactive supercomputing at the exascale level, in particular interactive computational steering, visualisation and Big Data integration."[3]

This work is linked to that of another Subproject on Neuromorphic Computing (SP9) which aims to develop two neuromorphic computing systems which take inspiration from the neurobiology of the brain and will, ideally, radically reduce power consumption (the human brain has the remarkably low energy consumption of around 20 watts, many hundreds of thousands of times less than a conventional supercomputer which has a fraction of its computing power). One approach, led by the UK's SpiNNaker group, has pioneered a system "using scalable digital many-core systems operating at real time with low power… that allows real-time simulation of networks implementing complex, non- linear neuron models. A single chip can simulate 16,000 neurons with eight million plastic synapses running in real time with an energy budget of 1W."[4] The other approach using the hardware-based Neuromorphic Physical Model (NM-PM) developed from the European FACETS program[5] "incorporates $50*10^6$ plastic synapses and 200,000 biologically realistic neuron models on a single 8-inch silicon wafer in 180nm process technology" allowing it to simulate complex, non-linear neuron models faster than real time.

Ultimately, the outputs of these two Subprojects will not only feed into the central aim of simulation of the human brain, but also link to the work of a third Subproject, Neurorobotics (SP10), "in which a virtual robot is connected to a brain model, running on the HPC platform or on neuromorphic hardware… the HBP platform will be the first to couple robots to detailed models of the brain. This will make it possible to perform experiments exploring the link between low level brain circuitry and high-level function."[6]

It is these ambitions that form the focus of the present Foresight Report. We have adopted a deliberately holistic approach to capture some of the deep interconnections existing across the HBP. In particular, we have singled out two cross-cutting topics that help identify broad social and ethical challenges related to the potential contributions of the HBP for future ICT in the light of wider developments and trends: intelligent machines and human-machine integration, including what has come to be known as 'affective computing'. We also comment briefly on the implications of the increasing use of robots in the economy – an issue that has been much analysed by other reports[7] – on the need for Life Cycle Assessment to address the issue of the growing amount of electronic waste produced globally each year, and on the question of 'dual use', that is to say the potential of technologies developed for civilian purposes being utilised for military ones.

1.2 Foresight Lab Reports – Aims and Methods

1.2.1 Aims

This is the third report of the Foresight Lab of the HBP. As with our first two reports – our Foresight Report on Future Medicine[8] and our Foresight Report on Future Neuroscience[9] – the aim of this report is threefold. *Firstly,* the report aims to focus on the key social and ethical issues that are likely to arise in the relatively short term – that is to say in a five to ten year timescale – and to bring these to the attention of the Directors and research leaders of the HBP. Together with the activities of our colleagues working on Researcher Awareness, the aim is to inform and advise those responsible for key decisions and activities in the Project, so that they may take these potential future issues into account in the strategies that they are developing. For example, in our first report, we focussed on questions of data protection and privacy, exploring the potential challenges that legal and ethical considerations posed for the federation of clinical data and its mining for 'disease signatures'. Our recommendations fed into the decision making of those responsible for the Medical Informatics Platform, and to the data governance architecture that they have developed, to feed back into the work of the HBP itself, and to encourage reflection

among researchers and their leaders. This approach, which seeks to develop and enhance reflexivity, and build capacity among researchers and research managers themselves, is a key component of Responsible Research and Innovation (Owen *et al.*, 2012; Stilgoe *et al.*, 2013). Responsible Research and Innovation (RRI) thus distinguishes its approach from that which became widespread under the acronyms of ELSI (Ethical, Legal and Social Implications), which rarely engaged the researchers themselves and instead sought to identify downstream implications, seldom having any effects on the actual content, direct or take up of the research programmes in question.

Our second report, on Future Neuroscience, focussed on the challenge of community building among those within the HBP working on brain modelling and simulation, and between them and the many other groups across the world which were working on similar issues, but in different ways. Here our report was intended, not only to make recommendations to the Directors and researchers of the HBP, proposing strategies to build a neuroscience community, and emphasising the importance of open source modelling in order to build trust, cooperation, to minimise duplication of effort and maximise synergies, but it also was directed to a second audience – that of the broader neuroscience community. Thus, our *second* aim is for our reports also to inform and engage the wider neuroscience community in the work of the HBP, to encourage a wide debate on the promises and challenges posed by the HBP, to help alleviate some of the concerns that have arisen from the perception that the HBP is enclosed and inward looking, and to engage neuroscientists – and indeed other stakeholders, such as regulators, clinicians, funding agencies, non-governmental organizations and commercial enterprises, more closely with the work of the HBP. To this end, our reports are public documents, openly available on our website and that of the HBP itself, and intended to be the subject of wide discussion.

This wider discussion, we hope, will also involve our *third* audience – the general public, that is to say, the citizenry of Europe who have an interest in the work being undertaken by the HBP, and in its potential to fulfil the ambitious promises that it has made in its three key areas of future medicine, future neuroscience and future computing and robotics. This wider work with the European public is also taken forward through our links with colleagues in the HBP's Ethics and Society Subproject, who are undertaking widespread citizen consultations, using a variety of methods which are described elsewhere.

1.2.2 Methods

We have discussed our methods at length in our previous reports, and will be brief here. We use two main strategies in our work. The first is to collect and analyse the views, attitudes and strategies of key stakeholders with methods from the empirical social sciences. To that end, we first carry out an extensive period of 'horizon scanning', examining the literature, both academic and popular, and identifying key themes and questions. In the case of this report on future information and communication technology and future robotics, this task was made more complicated by the fact, as we discuss later, that these issues are currently the topic of intense speculation. This holds both for topics that are – or claim to be - evidence based, as in the many reports on the implications of ICT and robotics for industry, health care, domestic life and much more, and for those that are highly speculative, such as predictions about the supplanting of humans by artificial intelligence, 'the singularity' and so forth. Nonetheless, in attempting to distinguish fact from fantasy, and to focus on the near term, we distilled a number of key themes, which we then used as the basis of webinars with key stakeholders.

We held two webinars co-organized with the Danish Board of Technology which focussed on key themes of dual use (military/civilian), intelligent machines, human-robot interaction, machine-learning, and brain computer interfaces.[10] In these webinars, which are open to an invited audience of 25-35 persons, we gather together key questions in advance, from

literature and discussions with researchers inside and outside the HBP. During the webinar, together with other participants, we pose key questions about future directions, potential alternative pathways, risks and benefits, and we record and analyse the debate that takes place. We also invite participants to continue the debate after the Webinar. In this case, we were delighted to see an extensive email debate between key researchers in the HBP and external experts on many issues, notably the theme of machine intelligence and machine consciousness. In previous reports, we have also used some more systematic foresight techniques, including scenario construction based on narrative and fictional short scenarios (vignettes), and we used questions concerning those scenarios to explore key questions. However, as already suggested, we are currently surrounded by such imaginary scenarios, on television and film, in discussions on the radio and the internet, and we decided not to add our own scenarios to these, but to work on the basis of those imagined futures that were already in wide circulation.

In our discussions with HBP researchers, experts, and stakeholders during our workshops, webinars and on-going participation in the HBP, from all of which we collected the data for this report, we found an active concern by many HBP researchers about many of these issues. This is an important point to make for two reasons. Firstly, because of the common perception that most computing and neuroscience researchers are not considering the societal and ethical issues raised by their research.[11] Secondly, as a methodological point concerning how foresight work is most effectively done. The scientist in the lab working with materials, models, and concepts may actually be the first person to consider a future technology or application, and hence often the first person to consider its societal impacts or potential ethical issues. Therefore, the first data for any type of foresight scoping work comes from the researchers, computer scientists and engineers themselves and they are also potentially the best people to propose solutions (Fisher, Majahan and Mitcham, 2006; Wynne, 2011). Linking our foresight work to researcher awareness is not only important for foresight it also helps develop capacity within the researcher community. When HBP researchers come together with stakeholders and external experts, this helps facilitate refection by bringing people with different types of expertise, interest, and position within the innovation system into conversation with one another. In the long term, this type of integrative work is intended to develop the capacity of researchers to understand their own role in the innovation system and how their work can be most useful. Likewise, it may inform civil society, entrepreneurs, policy professionals, and other stakeholders, and experts from other fields about future applications and their potential impacts.

At this point in our report, it is useful to remark briefly on some of our findings, because it helps to explain how we have structured this report. Because of the diversity of potential applications enabled by HBP research, many of our research participants thought it unreasonable to judge the HBP on all possible applications, for example, of intelligent machines. Often, there seems to be a belief that the research level and the application level of science are two separate strata, without any meaningful connection by which scientists might have an ethical impact on the application of their research. However, in grappling with this separate strata problem, individual researchers were suggesting ways to engage with it. One researcher suggested: "The problem is really about these different phases, research phase, product development phase, deployment phase. And for each of these stages, we need a different approach." Another, a roboticist, clarified this idea of different approaches to the governance of innovation at different phases: "We need one set of ethics underpinned by standards at the product development stage and legislation at the product deployment stage." Another expert participant reminded us that 'intelligent machines' are always closely interlinked with humans as part of sociotechnical systems: "There is too much focus on (an) individual machine. A military drone is an issue of policy. Robots taking jobs is a capitalist economic problem. The underlying problem is with the system, but the robot is easier to demonize."

By bringing together researchers with persons who have specific interests, knowledge or locations within the innovation ecosystem (stakeholders, entrepreneurs, civil society

advocates, policy workers, etc.) there is hope to interconnect the seemingly separate strata of research and applications. Even at the basic research phase, productive exchange that accomplishes this can help build the capacity of later communities to make judgements on applications or potential applications. Early engagement in this way may enable later development of appropriate standards for new technologies and appropriate legislation or regulatory guidelines for their deployment. This process also gives researchers greater insight into which applications will be most useful.

As we have emphasised in our earlier reports, the HBP Foresight Lab seeks to develop a multi-institutional process of capacity building, both within the HBP and with relevant constituencies outside. It uses an iterative process to bring the views and priorities of diverse communities into interaction and dialogue, working closely with the other dimensions of the Ethical and Social Division of the HBP to develop and enhance an expanding dialogue that can encourage wide debate among key stakeholders and the general public, but can also feed back into the direction, management, and priorities of HBP researchers.

1.3 Background and framing of the report

As we have said, at present there is a swirl of speculation about artificial intelligence and robotics, with much talk of 'the singularity'. Widely publicised scientists such as Stephen Hawking have been warning that these developments are leading to the development of 'thinking machines' that will outcompete mankind and could spell the end of the human race.[12] In this report, we try to take a different view, one located at the level of the emergence of interconnected, part-human and part-machine systems. From this perspective, we will first think about various strands of research on ICT and robotics in the HBP and their possible implication. We will then look at broader issues and challenges that go beyond those that are specifically related to the HBP.

Compared to our previous foresight exercises, the present one, on Future Computing and Robotics, has had to confront two new kinds of challenges in relation to the HBP.

Firstly, while many of the researchers involved in HBP activities that could make direct contributions to the broad domain of Information and Communications Technology (ICT) and to robotics do think that these developments will have social and ethical implications in the medium or long term, most do not think they have immediate consequences for their own research. Researchers in these fields seldom need to apply to ethics committees and are not used to thinking through ethical issues as they develop a particular strand of research. Most hold the view that new technological developments can have huge potential ramifications, positive and negative, which are often unpredictable, extremely difficult to envision at the outset, and impossible for the researchers themselves to control. Moreover, many researchers tend to adopt an instrumental rationality approach to their work. That is to say, they build tools to fix specific problems, and publicly profess a distaste for speculative thinking involving many unknowns and uncertainties. In the HBP, there seems to be a prevailing view that ICT- and robotics-related work is geared towards research and not towards commercialisation. In fact, some worry that the BRAIN initiative, the US counterpart to the HBP, is much more focused on developing delivery platforms that could apply the knowledge generated by the HBP, so that the Europeans would not reap the economic benefits of their own research. The consequence is that, in this foresight exercise, we have had to anticipate a translation from research to commercialized and industrialized application that goes beyond the current work undertaken by the HBP. We have had to anticipate what the research in various parts of the HBP could bring to future ICT and robotics, beyond a list of detailed technological features and breakthroughs, in order to highlight the ethical and social issues it could raise.

Secondly, the Subprojects of the HBP have the potential to make contributions to many areas of ICT and robotics, leading to a vast range of future applications and cutting across many domains. To address such a large range of possible products and domains of application, we have decided against taking an inventory-like approach to each kind of product and domain which would generate a long and disparate catalogue, hindering our ability to identify shared features. This would be counter-productive in the context of a foresight exercise, where a goal is to anticipate on future trends and drivers. Instead, we have adopted a more holistic approach, looking at hardware and software, machines and humans, as parts in larger systems.

The dependencies existing between many of the developments made in the HBP fully justify adopting a systemic approach to better grasp the potential of the Project for future ICT and robotics. On the surface, the ICT and robotics work in the HBP may appear to cover an assortment of independent fields, with loosely related outcomes spread across robotics and computing, hardware and software (processors, robots, high-performance clusters, machine learning algorithms, etc.). In reality, these developments are strongly interconnected, and these interconnections would be lost from view if we were to take a piecemeal approach. Besides the variety of machine learning tools (such as artificial neural networks and evolutionary algorithms) and other tools used for data mining and data analytics, which are developed and used broadly across the entire Project, some specific Subprojects present obvious potential for future ICT and robotics. These are the High-Performance Computing Platform (SP7), the Neuromorphic Computing Platform (SP9), and the Neurorobotics Platform (SP10) Subprojects.

SP7, which becomes the High-Performance Analytics and Computing Platform Subproject in the operational phase of the Project, aims to provide the Project with "the high-performance analytics and computing capabilities, systems and middleware necessary for the analysis of massive neuroscience data sets and the simulation of large multi-scale brain models." To this end, in the Ramp-Up Phase, SP7 has chosen two contractors for the pre-commercial procurement of research and development services: Cray, and a consortium of IBM and NVidia. In mid-2016, these will each deliver pilot interactive supercomputers, for deployment at the Jülich Supercomputing Centre. Besides its scientific impact for computational neuroscience, SP7 expects an indirect social and economic impact through the services it provides to other Subprojects, in particular Neuromorphic Computing (for instance, for mapping and routing circuits to neuromorphic substrates, or for testing hardware specifications through simulation) and Neurorobotics. Its innovation potential lies in two main directions. Firstly, SP7 is developing new 'big data' technologies for remote interactive simulation, visualisation and analytics, which fit squarely within the European Commission's strategy for a data-driven economy[13] and have a wide scope of application across both scientific and industrial research. Secondly, SP7 is doing collaborative research on novel High Performance Computing (HPC) hardware based on low-power neuromorphic technologies; energy-efficiency being another strong global driver for future ICT.[14]

The overall goal of SP9 is "to establish Neuromorphic Computing as a new paradigm of computing, complementary to current designs, and to explore potential applications in neuroscience and machine learning… The Neuromorphic Computing Systems developed by SP9 are hardware devices incorporating the developing state-of-the-art electronic component and circuit technologies as well as knowledge arising from other areas of HBP research (experimental and cognitive neuroscience, theory, brain modelling)." Two distinct categories of neuromorphic architecture are being developed in SP9 and will be accessible to researchers through the Platform. The University of Heidelberg team, led by Professor Karlheinz Meier, is developing an architecture based on an analogue neural network core, for a *physical* simulation of brain models running in *time-accelerated* mode. The University of Manchester team, led by Professor Steve Furber, is developing an architecture based on digital multicore systems, for a *virtual* simulation of brain models running in *real-time* mode. Novel applications of the technology will be explored through

Partnering Projects. SP9 envisions that the neuromorphic technologies it develops could generate commercially valuable applications for manufacturing, transport, healthcare, and consumer electronics. Moreover, SP9 expects the European electronics design and manufacturing industry to play a key role in the development and construction of neuromorphic systems. In order to help boost the market for such systems and establish European leadership, SP9 "plans to develop national technology nodes for neuromorphic applications, with a focus on robotics, automotive, manufacturing and telecommunication systems." This could be developed as an offer of commercial services to researchers and technology developers.[15]

The main scientific objective of SP10 is to deliver tools for researchers to test the cognitive and behavioural capabilities of the brain models built by the Simulation Subproject (SP6), as well as the neuromorphic implementations of derived simplified models on SP9 Platform. In the first phase of the HBP, the Neurorobotics Platform relies on simulated robots immersed in simulated environments, offering the possibility "to design simulated robot bodies, connect these bodies to brain models, embed the bodies in rich simulated environments, and calibrate the brain models to match the specific characteristics of the robot's sensors and 'muscles'."[16] During the operational phase of the Project, the Platform will also give access to physical robots controlled by brain models that can be executed in real time on neuromorphic hardware. The HBP Neurorobotics Subproject aims to be the first to prototype applications that will exploit the novel cognitive and behavioural capabilities of physical robots with neuromorphic controllers. It is expected that such robots will have functional capabilities (e.g., learning, effective handling of multimodal real-time input) well beyond what current robotic technologies can offer, with major potential impact over a broad range of domains. A first major applicative strand lies in specialized robotics for industrial applications (from manufacturing to transport) and also household and healthcare applications. A second major strand is that of neuroprosthetic devices, for medical purposes, and labs dedicated to this field of research are joining the Project in the operational phase. Like SP9, SP10 plans to offer commercial services to industry, to experiment with state-of-the-art neurorobotics setups.[17]

In a more indirect manner than SP7, SP9 and SP10, the Theoretical Neuroscience Subproject (SP4) has no less potential for future ICT and robotics. For instance, the work done in SP4 during the Ramp-Up Phase of the Project, on principles of brain computation, on novel computing systems inspired by biology, on learning rules and on spiking learning algorithms holds promises for the fields of machine learning, neuromorphic computing and brain-inspired robotics. Indeed, a major goal of SP4 over the course of the Project is to identify the data and computing principles required to model specific brain functions in neuromorphic computing systems. A key target is the development of models suitable for implementation in neuromorphic and neurorobotic systems as well as in large-scale, top-down simulations of the brain.[18] Large-scale, top-down simulations of the brain are used to explore brain mechanisms involved in cognition and behaviour, and this particular class of simulation is also of interest to the field of machine learning.

2. HBP-related challenges and issues

Our overview of the potential contributions of the HBP to future ICT reveals some of the many interdependencies existing between different parts of the project and, through them, the close and often reciprocal relations existing between the various fields of research involved, and their applicative domains. Our work for this Foresight Report has led us to conclude that there is a need for a systematic, Project-wide review and analysis of this synergetic potential, that would lead to the development of a Responsible Research and Innovation roadmap for building on them. In order to contribute to such a roadmap, in this foresight exercise we have adopted a deliberately holistic approach to capture some of the deep interconnections existing across the HBP, and we have singled out two cross-

cutting topics that help identify broad social and ethical challenges related to the potential contributions of the HBP for future ICT in the light of wider developments and trends: intelligent machines and human-machine integration.

2.1 Intelligent machines

The first cross-cutting topic that we consider is the single broad category of 'intelligent machines'. We have framed the topic in this way because we are not convinced that it is possible to sustain the distinctions between 'robots' and 'non-robots' on the one hand, and virtuality and materiality on the other hand.

2.1.1 What counts as a robot?

To put the issue simply, a 'humanoid-looking ICT-enabled metal thing' is usually considered as a robot, as are non-humanoid ICT-enabled machines that carry out tasks once undertaken by humans. However, we do not normally think of a smartphone, also packed with sensors and software and ever more artificial intelligence, as a robot. Where do we draw the boundary? Isn't the Internet of Things already a giant, networked assembly of robots? While this distinction is highly problematic, that does not mean that it is not meaningful; the public perception of the 'robot' comes with a number of fears and expectations attached, which are very different from those attached to the 'non-robot'. During our activities leading to the preparation of this report, this is an issue that the Neurorobotics Subproject has raised repeatedly. It was also raised at a public event on robots and elderly care that we attended, organised at the House of Lords by BioCentre, a UK think-tank concerned with the ethical, social and political implications presented by new emerging technologies.[19]

2.1.2 What is 'virtual'? What is 'material?

Taken separately, the categories of machine learning – more generally, artificial intelligence – and robotics seem to embody a separation between virtuality and materiality – analogous to the distinction of software and hardware. However, this is another boundary that should be questioned. Some already propose that the definition of a robot should extend to automated computer programmes, and recognize that the definition will need to keep changing to reflect not just new technological developments but also associated social and ethical challenges.[20] To start with, the virtual side of robots and the material side of computing both require some attention. Much of the work that goes into making a physical robot is virtual, using simulated versions of the robot immersed in simulated environments. As we have seen, building such a virtual simulation platform is a primary objective of the Neurorobotics Subproject (SP10). On the other hand, the virtuality of cyberspace has very material impacts on our environment, such as hi-tech waste dumps or the carbon footprint of internet use.[21] Meanwhile, the material and the virtual could merge in novel ways into a new breed of complex systems in the emerging and overlapping fields of cyber-physical systems (composed of networked, collaborating computational entities controlling physical components)[22] and of cloud robotics, which general idea is to build robots which 'brains' are based in the cloud (see illustration below): "cloud robotics happens when we connect robots to the Internet and then, by doing so, robots become augmented with more capacity and intelligence. The cloud lets them communicate with other machines and serve their human operators better. Connected robots equal augmented robots. By collaborating with other machines and humans, robots transcend their physical limitations and become more useful and capable, since they can delegate parts of their tasks to more suitable parties."[23]
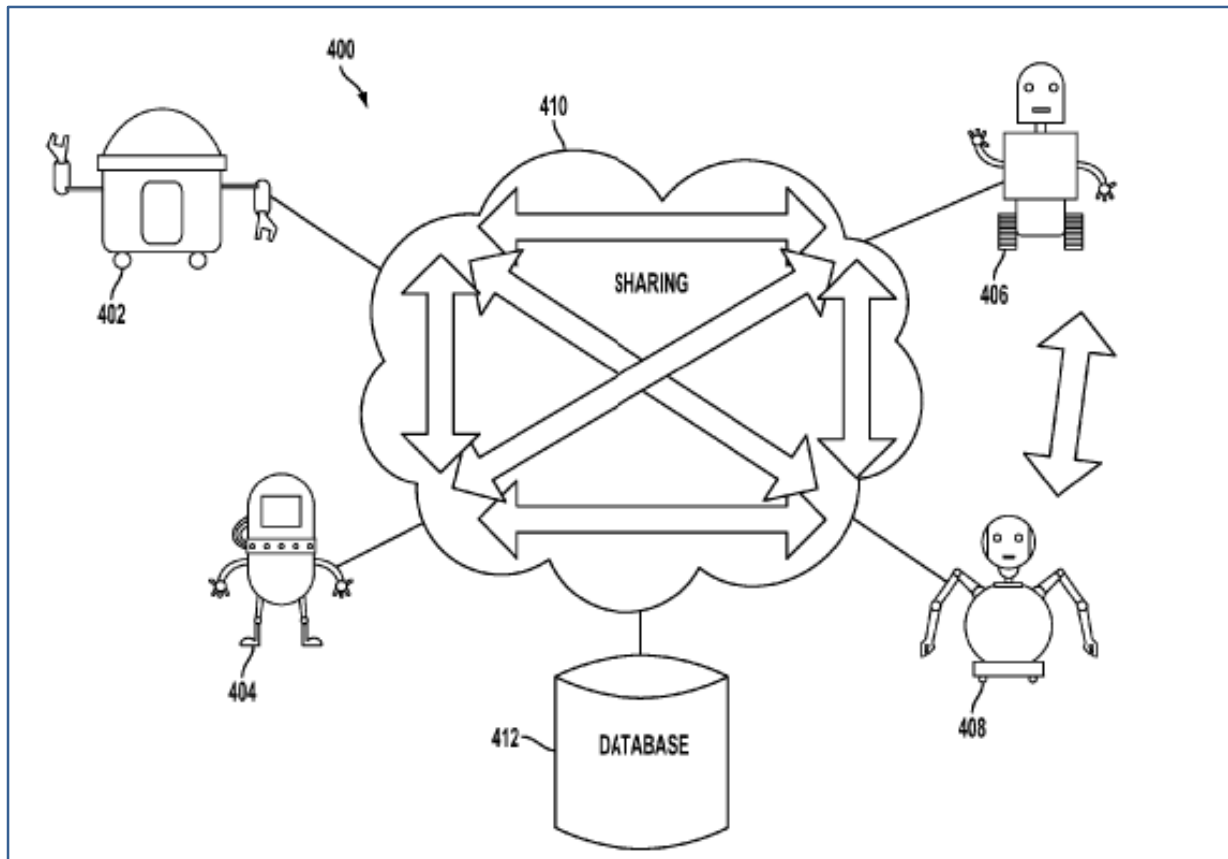
Figure 1: 'Cloud' robotics

Source: _United States Patent US 8,996,429 B1, Francis, Jr et al., Google Inc., Methods and systems for robot personality development, Mar. 31, 2015._

Questioning the clear distinction between hardware and software, between virtual and material, can also bring to the fore differences in philosophical positions and highlight how the adoption of a specific view on this topic can have very practical implications. This is best illustrated by an example, taken from a debate that was initiated during the webinar on Future Computing and Robotics in October 2015.[24] Mark Bishop, Professor of Cognitive Computing at Goldsmiths College (UK), gave a presentation entitled "Dancing with Pixies" in which he defended by _reductio ad absurdum_ the rather poetic view that "if Strong AI (which is the belief that an Artificial Intelligence could functionally become at least the equal of a mind) is possible then we are exposed to a vicious form of panpsychism, wherein an infinitude of disembodied conscious experiences – little `pixies' – dance in every grain of sand." It would be too long to detail the whole argument and the lively conversation that ensued, first online and then by email, between Mark Bishop and, mainly, Steve Furber, ICL Professor of Computer Engineering at the University of Manchester and co- leader of SP9, and his collaborator David Lester. But in the course of this exchange, it appeared that although they were all 'singularity sceptics' doubting the possibility of machine consciousness, there was philosophical disagreement about the extent to which it is possible to separate 'machines' (computers, brains) – the material - from the abstract systems they support (models, minds) – the virtual. The Manchester team held the view that the mind is an information processing system supported by the brain. They saw it as a virtual (but nonetheless very real) system running on a substrate which, although ultimately supported by a physical system (the brain), was separated from it by several layers of abstraction which gave it (the mind) an existence that is independent from the details of the underlying physical system. Thus, while pain is a real sensation supported by the physical (and real) brain, it is sensed by a virtual, yet real, system (the mind) which

could equally be supported by a non-biological physical system of a suitable type. In contrast, Bishop suggests a note of caution; until the issues raised by the Dancing with Pixies *reductio* have been fully addressed and the role of embodiment more strongly engaged (such that neurons, brain and body more properly interact with other bodies, world and society), he is wary of labelling any artificial system as 'strongly intelligent' - a computational mind - in its own right. For Bishop, any 'mind' postulated emerging from a sea of mere 'computations' is just a projection of its engineer's intellect, aesthetic judgement and desire.

The philosophical stance of the Manchester team, on the 'replaceability' of the physical system for running a virtual system, is instantiated in practice in the design of their neuromorphic architecture, SpiNNaker. Contrary to the approach adopted by the Heidelberg team, whose part-analogue neuromorphic architecture aims at physically simulating brain models, SpiNNaker's physical architecture is fully digital and runs virtual simulations of brain models. From a neuroscientific perspective, the SpiNNaker physical architecture is less accurate in comparison to a real brain, but if one believes in the relative independence of the mind (as virtual system) from the brain, then it is not an issue for simulating brain functions (which does not rule out, as David Lester points out, that we may still find that we need to take quantum biology into account to explain macroscopic features of the brain, in which case much more computing power will be required to actually make headway). From an applicative perspective on the other hand, by virtualising the model away from its physical implementation and using industry-proven digital ARM cores, the versatility of the SpiNNaker's architecture gives it greater flexibility in supporting a wide range of biological and artificial neural network models, which could lead in turn to an equally wide range of potential applications in robotics, machine learning, and combinations of the two, and may extend into very diverse industrial applications.[25] This philosophically-driven versatility of SpiNNaker is thus an important asset for the potential of the HBP to translate in the short to medium term from academic research to industrial innovation in the broadly understood domain of 'intelligent machines'.

2.1.3 The HBP and the future of intelligent machines

Our overview of the HBP's activities that hold promise for the future of ICT, and the previous discussion of the SpiNNaker chip in particular, have shown that the Project most certainly has potential to help bring about a very diverse range of variously intelligent machines, and not just in the long term. Yet, when asked if the research done in the Project has some relation to Artificial Intelligence, many collaborators answer that it does not. Some believe that the ultimate goal of the Project could be Artificial General Intelligence or 'Strong AI', i.e. producing general-purpose systems which intelligence would be at least equivalent to that of the human mind, but the feasibility of such an outcome is perceived as being far into the future.[26]

What counts as Artificial Intelligence is not straightforward, and the lack of consensus on defining the term allows for much flexibility in the answers given.[27] There is a general agreement on the distinction between classic AI (broadly speaking, this encompasses symbolic AI and experts systems), Artificial Neural Networks (ANNs are based on highly simplified neuron models that display very little biological realism; different kinds of ANNs are used in machine learning – e.g. for deep learning, supervised and unsupervised learning, reinforcement learning) and Biological Neural Networks (aiming for biological accuracy, they are the models used in computational neuroscience).[28] These days, the vast majority of courses and textbooks on Artificial Intelligence encompass Artificial Neural Networks alongside more traditional methods.[29] Those who deny that the HBP is an Artificial Intelligence project draw a firm line between machine learning and computational neuroscience: the former uses the highly simplified Artificial Neural Networks (ANNs), while the latter, such as what is done in the Simulation Subproject of the HBP, is based upon entirely different kinds of biologically realistic models. Still, we have

already seen that although biological and artificial neural networks may involve different kinds of models, the SpiNNaker architecture developed in the Neuromorphic Computing Subproject (SP9) has the flexibility to support models in both categories. This is one example showing that, although the HBP may not be strictly speaking an Artificial Intelligence project, innovation in Artificial Intelligence could still be an important, if indirect, outcome of the Project. But further, the extent to which it is possible to draw such a firm line between machine learning and computational neuroscience is itself questionable.

A major modelling tool of biological neural networks is the category of Spiking Neural Networks (models comprised of 'firing' neurons that reflect experimental findings of neurophysiology). In its goal to produce a scaffold model, first of the mouse brain then of the human brain, the Simulation Platform of the HBP Subproject (SP6) develops models at various scales with different levels of biological details: a molecular simulator for neuron and synapse models at the molecular level of resolution; a cellular simulator to simulate large-scale, biophysically realistic neural tissue models; and a network simulator for network models at the level of resolution of neurons and synapses. The cellular and network simulators are building up and improving on two open simulation software environments that implement models of spiking neurons, respectively NEURON [30] and NEST.[31] NEST in particular is a simulator dedicated to spiking neural networks, including large-scale neural networks. Meanwhile, in Artificial Intelligence, spiking neural networks have been called the 3rd generation of artificial neural networks (Maas, 1997), but although they can in theory be used for information processing just as traditional artificial neural networks, their theoretical processing power has not yet translated into engineering applications.

On existing hardware architectures, large networks of spiking artificial neurons lead to impractical (perhaps impossible) time-consuming simulations. However, research conducted in different areas of the HBP is working towards removing current limitations and thus bringing spiking neural networks to 'intelligent machines' applications: by building up high performance computing capabilities to support the simulation of large multi-scale brain models in the High-Performance Analytics and Computing Platform (SP7); by developing neuromorphic architectures that can run smaller scale whole brain simulations in biological real-time or even on an accelerated timescale in the Neuromorphic Computing Platform (SP9); by improving how simplified models of neural networks can 'abstract' essential features from larger scale models to remain biologically realistic; and also by finding ways to construct spiking neural networks for efficiently solving computational task[32] in the Theoretical Neuroscience Subproject (SP4). Machine learning, neurorobotics, and autonomous robotics in general, are all hoping to benefit from such advances (Walter, Röhrbein and Knoll, 2015a; 2015b). We can thus conclude that, while the HBP may eventually lead in the long term to General Artificial Intelligence, in a nearer future it is likely to contribute to innovation in 'narrow' or 'weak' Artificial Intelligence – i.e. (non-sentient) intelligent machines focused on specific tasks.

When it comes to Artificial General Intelligence, our colleagues in the HBP are largely 'singularity sceptics'. Thus, in the interviews of senior collaborators conducted during the first year of the Project by the Researcher Awareness team (WP12.4), those who envisioned that the HBP could lead to actually 'conscious' rather than just 'intelligent' machines, viewed this possibility as being too far into the future to deserve serious consideration: "not feasible for many, many decades"; "What is consciousness? What is awareness? These are questions that will take a long time to be fully understood"; "there are all these wild ideas that say, well, we may build systems that are more intelligent than we are and they would take over the world and things like that. That is not what I would call responsible innovation. I would almost call this irresponsible because it is speculations that are not based on knowledge."[33] Similar views were expressed by participants in the webinars we organised with the Danish Board of Technology Foundation in October 2015. This comes as a refreshing contrast to the hype around the existential risk of machines

overtaking humans, which, as we have already noted, has involved scientific luminaries like Stephen Hawking entering the fray to warn that sooner rather than later "artificial intelligence could end mankind."[34] The attitude of our collaborators in the HBP is much closer to that adopted for example by Demis Hassabis, Shane Legg and Mustafa Suleyman, the three co-founders of the British Artificial Intelligence company DeepMind Technologies, now Google DeepMind,[35] who address the topic of machine intelligence in a deflationary manner. In a nutshell, their argument is that 'Strong AI' is currently science fiction; the systems that are currently designed and on which their company itself is focusing are narrow and applied. As for the risk of AI agents reaching moral autonomy, their answer is "Of course we can stop it, we are designing these things." Nonetheless, despite being convinced that the technology is decades away from anything nearing human-level general intelligence ("Think of it as we're trying to build insects. Years from now, you might get to a mouse"), their 20-year roadmap is planned with Artificial General Intelligence as its ultimate goal.[36] Correctly in our opinion, Hassabis argues that unnecessary fear mongering and unsubstantiated speculation are not conducive to stimulating a healthy debate (Rowan, 2015). His associate Suleyman agrees that "On existential risk, our perspective is that it's become a real distraction from the core ethics and safety issues, and it's completely overshadowed the debate".[37] Andrew Ng, head of Artificial Intelligence research at Baidu, is of the same view, and he too sees talk of AI 'superintelligence' as a distraction from the real risks of technology.[38] And certainly, 'the singularity' does not have to be near for 'intelligent machines' to present risks and challenges that raise ethical and social issues in need of serious consideration.

Recent years have seen an ever-growing number of debates and initiatives concerned with the existential, ethical and social implications of Artificial Intelligence. Institutions such as the Future of Humanity Institute in Oxford, United Kingdom,[39] or more recently the Future of Life Institute in Florida, United States,[40] dedicate considerable resources to investigating these dimensions of Artificial Intelligence.[41] Private companies in the domain of Artificial Intelligence have also started equipping themselves with independent AI ethics advisory panels: this is the case of Lucid AI[42]. It is also supposedly the case with Google: the creation of an AI ethics advisory committee was part of DeepMind's agreement to their acquisition by Google in 2014. DeepMind has communicated that no technology coming out of it will be used for military or intelligence applications as part of the acquisition terms and Jaan Tallinn, co-founder of Skype and of The Future of Life Institute, has been instrumental in establishing the ethics advisory committee. However, its composition and its workings are kept confidential, to the extent that some denounce it as a PR stunt.[43]
We have two main reservations with many of these initiatives. Firstly, many of them target speculative 'existential risks' rather than focussing on more immediate and realistic issues. This tends to convey the message – which we consider misguided - that until we reach the singularity there is not much to worry about. Secondly, many of these initiatives are led by the private sector or private philanthropy, and can involve a degree of secrecy (absolute in the case of Google), which makes for an ambiguous alliance with ideas of openness and democratic engagement.

Our analysis suggests that it is misleading for participants in the HBP to insist on a restrictive definition of Artificial Intelligence in order to claim that the Project has nothing to do with Artificial Intelligence. As we have seen, the forums and initiatives aimed at thinking through its potential beneficial and detrimental developments for society usually adopt a broader understanding of the term. It is also not convincing to suggest that these wider social and ethical issues are not the concern of the HBP, as it is geared towards research rather than commercialization. Indeed, HBP researchers' worry that the United States will reap the benefits of translation from research to market shows that participants in the project are well aware of the scope for such translation. For these reasons, the HBP should recognize that a number of Subprojects across the Project will collectively contribute to the broader field of machine intelligence, and act accordingly. The Directors, researchers and others in the Project thus need to work with the Ethics and

Society Subproject, and with others outside the HBP, to participate in current debates around how to make Artificial Intelligence ethical and socially beneficial and bring the weight of publicly-funded, open research to bear on them. This aligns with two of the three policy recommendations recently made by Wendell Wallach, a Hastings Center senior advisor, at a press briefing on the emergence of intelligent machines at the American Association for the Advancement of Science (AAAS) annual meeting in Washington:

- "Direct 10 percent of funding for research in AI and robotics toward studying, managing, and adapting to the societal impact of intelligent machines, including the impact on wages and employment."

- "Create an oversight and governance coordinating committee for AI and robotics. The committee should be mandated to favour soft governance solutions (industry standards, professional codes of conduct, etc.) over laws and regulatory agencies in forging adaptive solutions to recognized risks and dangers."[44]

Here are a few questions, some of them near-term concerns, which could be relevant to the potential outcomes of the HBP:[45]

- Should core questions of safety, or wider ethical worries about machine-powered decision-making usurping human judgment, be society's biggest concern as intelligent machines proliferate? Is it even possible to separate safety from ethics at that fuzzy juncture?

- Should Artificial Intelligence research be open source by default? How can we expect to control and regulate the social impact of increasingly intelligent machines when the largest players in the field are commercial companies that do not divulge their proprietary algorithms?[46]

- Will it be the case that the more generalist our machines become, the less capable and/or reliable for a particular task — and arguably, therefore, the less safe overall?

- Is the umbrella term 'artificial intelligence' actually an impediment to public awareness and understanding of myriad developments associated with algorithms that can adapt based on data input?

- How to regulate and control increasingly powerful data mining technologies that can combine, or hope to combine, data collected in different national jurisdictions, which may be subject to distinct national regulatory regimes on data protection and data use.

- How to protect user privacy from predictive algorithms and ensure informed consent of data processing?

- How to avoid advanced ICT and robotic technologies—driven by algorithms that have incorporate human sensorimotor and other skills that are then honed by either guided or autonomous learning—destroying jobs and concentrating more and more wealth in the hands of fewer and fewer organizations or individuals?

- Will society benefit from the increased efficiency of intelligent machines or will wealth be increasingly concentrated?

These are only a sample of the questions related to social and ethical issues raised by intelligent machines, that the HBP could – and we believe should - engage with. They are especially important as the HBP, one of the two Future and Emerging Technology Flagships of the European Union, enters Horizon 2020, the funding programme that will implement the Innovation Union, which outcomes are notably expected to position Europe as a world- class competitor in ICT research and digital innovation, through 'open science' and the development of innovative public-private partnerships.[47]

2.2 Human-Machine Integration

The previous discussion of intelligent machines has largely left the human out of the picture. Yet the human is present at every step of the way: as inspiration and metaphor, as developer and maker, as voluntary and at times involuntary user, as partner of the machine. This section turns to the second cross-cutting topic that we have chosen for highlighting broad social and ethical challenges of relevance to the HBP, Human-Machine Integration. It expands on the previous one by bringing humans firmly into the ICT picture and focusing on some of the relationships (existing or expected) between the human and the machine, keeping in mind that in the context of the HBP, this means not just computer scientists and engineers but also neuroscientists, clinicians and patient, and eventually a wide array of potential users – literally, everyone.

2.2.1 Replacing the human vs. augmenting the human

In his recent book on *Robotics and the Myths of Autonomy*, David Mindell[48] asserts in his concluding chapter that: "The challenges of robotics in the twenty-first century are those of situating machines within human and social systems. They are challenges of relationships" (Mindell, 2015, p.222). Mindell broadly situates himself on one side of a long lasting rift, supposedly going back to the early 1960s, between the early proponents of 'artificial intelligence', a project aimed at simulating human capabilities, and the proponents of 'intelligence amplification' (who went on to define the field of Human-Computer Interaction, HCI), believing "that computers should be used to 'augment' or extend human capabilities, rather than to mimic or replace them." (Markoff, 2015, p. xii). In the last decade, this rift has been analysed from within not so much as a divide between fields, Artificial Intelligence and Human-Computer Interaction, which in terms of focus, tools and methods are showing some convergence, than as a divide between approaches.

For Terry Winograd, the former still shows more affinity with a 'rationalistic' approach, which aims to model essential aspects of mental processes to eventually replace them in a quest for optimization, while the latter is more guided by a 'design' approach, where the central focus is on the interactions between humans and their 'enveloping environments' (Winograd, 2006, pp 1256-57; see also Grudin, 2009). In keeping with Winograd's 'design' perspective, Mindell proposes that once tempered "the naïve promises and the naïve fears currently offered for autonomous machines [then] we can move the conversation (and the creativity) toward questions of human roles, social interaction, and the challenges of reliability and trust. These sit at the core of new conversations on situated autonomy – Where are the people? Who are they? What are they doing? When? Why does it matter?" (Mindell, 2015, p.226). In what follows we focus on the two strands of the applicative strategy of the Neurorobotics Subproject (SP10) for the operational phase of the Project: specialized robotics for industrial, household and healthcare applications; and neuroprosthetics, including neural and brain-computer interfaces. They are especially interesting for the various – and material – ways in which they can integrate with the human, thus providing a good testbed to contrast the two approaches suggested by Winograd, the 'rationalistic' that tends to 'replace' the human and the 'design' which aims to 'enhance' the human. With the questions raised by Mindell to guide us, we sketch how these two approaches can play out in shaping different future scenarios.

Let us consider the case of self-driving cars, which has been repeatedly presented as an example of technology to which the HBP could positively contribute.[49] In self-driving cars as they are currently conceived, for instance, by Google, human drivers have been designed out of the system, in a typical example of a narrowly instrumental 'human replacement' approach (Mindell, 2015, Chapter 6). But more generally, the terms in which the transportation model of the self-driving car is presented shows no imagination towards exploring other alleys than the individual(ist) vehicle for individual(ist) autonomy, when a human-centred design approach that took into account a wide enough range of dimensions could be an opportunity for re-thinking transportation systems and championing innovative public transport policies – for instance self-driving can be (and already is) applied in many more types of vehicles than cars.[50] Instead, the narrower design model of the self-driving

car does not give centre stage to the richness and complexity of human lives; indeed it largely (almost literally) leaves them out. It is a design model that is rooted in a Western vision of a landscape built for the car not the human, and it offers little room at present for thinking how it will pan out in widely different road systems and urban geographies. It is a design model that puts at its centre existing transportation infrastructures, including car manufacturers - indeed, it is no wonder that research and innovation in this area is largely driven by the private sector.

Further, in contrast to what a more all-encompassing approach to transportation systems could achieve, the self-driving car as transportation model is not in the obvious best interest of the environment, or of public health. The primary benefits put forward for its general adoption are that it will supposedly drastically diminish the number of vehicles on the road by encouraging sharing, and make roads safer.[51] Regarding the former, a large reduction in the number of vehicles on the road could probably be achieved just as well through innovative public transport policies, and anyway it is not clear that automakers will happily let their revenue shrink after the initial surge when car owners convert to self- driving cars, considering their investments in R&D.[52] Regarding the latter, at present accidents are not the most detrimental impact of road traffic on public health: far more important is the contribution to outdoor air pollution.[53] This can indeed be improved by diminishing the number of vehicles on the road, but as we have pointed out, there is more than one way to achieve this. The example of self-driving cars shows that taking a different approach to the design of technological systems has the potential to displace many of what are considered core ethical concerns. In this case, the ethical concern currently monopolising public attention is the moral dilemma of the 'greater good' scenario: how should the autonomous system decide who to kill or seriously injure in an unavoidable collision.[54] Our analysis highlights that although this is certainly a relevant ethical question, there are many more important social and ethical issues that do not currently appear on the agenda of debate.

In the Operational Phase of the HBP, the Neurorobotics Platform (SP10) plans to offer commercial services to industry to experiment with state-of-the-art neurorobotics setups. This simulation platform initially allows virtual robotic bodies to be endowed with virtual brains and to interact with virtual environments and later intends to extend this to real robots in real environments. It could potentially offer significant benefits, for example if it was indeed able to replace some animal experimentation and allow experiments *in silica* that would not be permitted in animals or humans. But it is precisely this advantage that can become a source of concern, when it comes to the specialized industrial, household and healthcare applications that it could help develop: it makes it very easy to elide the human from the design process. This should resonate as a warning if we heed a major governance issue of robotic technology that was raised by panellist Dr Heike Schmidt-Felzmann at the Biocentre event 'Robot & Gran: Would you leave your ageing parent or grandparent in the care of a robot?' held at the House of Lords in March 2015.[55] Schmidt-Felzmann pointed out that at present, involving potential users in research and design requires ethical clearance that is sometimes difficult and time consuming to obtain, whereas placing an artefact in their home that has been developed without their involvement does not.

We consider next the research activities of the different labs from the Center for Neuro-Prosthetics (CNP) at the Ecole Polytechnique Fédérale de Lausanne (EPFL) that will be involved in the Operational Phase of the HBP:[56] neuroprosthetics, extended to neural and brain-computer interfaces. In this domain, the distinction between 'replacing' and 'augmenting' takes on a new dimension. The topic of neuroenhancement has long been raising fierce debates in the context of biochemical substances, but more recently these issues have also been raised in bioelectronics. A major issue in bioethical debate is whether we can ethically, and legally, distinguish treatment/rehabilitation from enhancement, because distinguishing between normalcy and disease or disability is not just subjective, but also politically charged. In bioelectronics, a case in point is the debate

over the use of cochlear implants to 'cure' deafness.[57] Another important issue concerns whether the interlinking of developments in cognitive neuroscience, engineering, brain–machine interfacing and medicine could fundamentally alter an individual's sense of self (Blanke and Aspell, 2009; Clausen, 2009). For more on the issues and challenges raised by neuroenhancement in its various forms, we refer our readers to the project 'Neuro-Enhancement Responsible Research and Innovation' (NERRI), funded under the research framework programme FP7 of the European Commission.[58]

Returning to our initial discussion of the contrast between designing the human out of technology and putting the richness of the human in all its dimensions at the centre, let us consider the controversial question of where to draw a line between treatment and enhancement. We can then see how other issues come to the fore when one brings in some of the socioeconomic aspects of human systems. In the case of cochlear implants for instance, we can see this contrast in the debate between those who conceive of cochlear implants as a great technological fix to remedy an individual disability, and those in the deaf communities who see it as destructive of their minority identity. The contrast between the two approaches could be summarized as between those who focus on adapting the humans to their socioeconomic environment and those who focus on the need to adapt the socioeconomic environment to the humans. But bringing in socioeconomic dimensions into the picture can also highlight more layers to the treatment versus enhancement debate. As argued during a recent panel discussion on neurostimulation co-organised by Virtual Futures and NERRI in London,[59] private companies that want to market a device have an incentive to package it as recreational cognitive enhancement to avoid regulatory hurdles, whereas publicly-funded researchers can only investigate neural technology for therapeutic ends. This makes it difficult to gather reliable and rigorous scientific evidence to inform policy on the regulation of cognitive enhancing devices. These layers need to be integrated in the discussion: bringing socioeconomic dimensions into the picture raises different issues to those traditionally raised by bioethics.

The challenges of human-machine integration go well beyond those raised by an understanding of human-centred design and approaches to 'keeping the human in the loop' narrowly focused on the individual. They encompass wider systemic issues, linked to the socioeconomic systems in which research and innovation in ICT is embedded (see O'Gieblyn, 2016; Morozov 2015). We have given a glimpse of such wider systemic issues in the case of the self-driving cars and also of the neuroenhancement debates. There are many more. For instance, paying attention to human-machine integration at the level of the researchers and developers, and this concerns the HBP, one finds that the immateriality and invisibility of software makes it too easy for scientific projects to equally 'not see', and not give their due, to the very real persons who develop it - unless they are professional software developers who derive their status from non-academic reward systems (Chawla, 2016). This raises the question of recognition and reward, a thorny issue that we have already addressed in our Foresight Report on Future Neuroscience in the context of data sharing and curation, and our recommendations in this area can be extended to software development (indeed, as we have shown in our report, in the HBP, data and models are strongly interconnected, and there is a strong overlap between data producers and model builders).[60] We now examine more of these systemic issues in relation to the HBP, through the lens of 'removing the bottlenecks', a major driver in ICT research and innovation.

2.2.2 Removing the bottlenecks

In the HBP Report to the European Commission, April 2012, which led to the HBP being chosen as one of the two Future and Emerging Technologies Flagships of the European Union, the main objectives of the project fell under three categories: future neuroscience, future medicine, and future computing and robotics. For this last category, the potential benefits of the project were summarized as follows: it had the potential to overcome fundamental limits on the performance and capabilities of conventional computing

technology by contributing to the development of a completely new category of low-energy computing systems with 'brain-like intelligence', and to 'embody' these breakthroughs into neurorobots. [61]

Over three years on, as the initial 30-months Ramp-Up Phase of the HBP comes to a close, and despite significant changes in its governance and scientific agenda, the Framework Partnership Agreement defining the trajectory of the Project for the remainder of its lifetime reiterates such expectations regarding its contribution to the future of computing and robotics: "The human brain performs information processing tasks that are inaccessible to the most powerful of today's computers - all while consuming no more power than a light bulb. Understanding how the brain 'computes' reliably with unreliable elements, and how different elements of the brain communicate, can provide the key to a completely new category of hardware (Neuromorphic Computing Systems) and to a paradigm shift for computing as a whole. What is more, neuroscience will be a driver for more powerful and highly interactive computing systems as well as innovative visualization technologies. The economic and industrial impact is potentially enormous." [62]

This aim of taking inspiration from the human brain to improve information processing performances while lowering energy consumption requires some clarification. There are tasks like routinized logical number crunching that traditional Von Neumann computer architectures are extremely efficient at and much better than human brains, and there is much work on this in data analytics and data mining. The main goal of the HBP may be a completely new kind of computing systems, training the neural networks that will run on the neuromorphic chips developed in the Neuromorphic Computing Subproject (SP9) also requires ameliorations in this type of number crunching, which remains crucial. The HBP, via the High Performance Computing and Data Analytics Subproject (SP7), also aims at improving the performance and number-crunching capacity of traditional computers. However, as is well known, there are many tasks that these same computers are extremely inefficient at, and which require large energy-hungry machines, but which are easy for humans, and for much simpler animals too; for example, navigating unfamiliar surroundings and recognising images. This form of natural computation - autonomous, adaptive and energy-efficient - is the main goal of the HBP, to which several Subprojects contribute and in particular the Neuromorphic Computing Subproject.

This dual goal of improving information processing performances while lowering energy consumption reflects a dominant trend in taking inspiration from the brain for improving ICT: removing the bottlenecks that hinder the optimized working of human-machine systems. In terms of this broader goal, it becomes clear that taking inspiration from the brain extends to interfacing with the brain, be it through the intermediary of the behaving human (for example, human-machine integration in industrial production environments) or in a more direct manner (i.e. neural and brain-machine interfaces). The HBP is one among many initiatives that aim to contribute in this domain. Many joint neuroscience-ICT research efforts worldwide are seeking to remove the bottlenecks, in the machines themselves but also at the interface between the human and the machine. [63]

This brings us back to the two applicative strands of the Neurorobotics Subproject (SP10). In the Operational Phase, SP10 will work in collaboration with SP9 towards integrating neuromorphic technology in neurorobots. This is expected to generate two main kinds of benefits. A dramatic improvement of performance at equivalent energy consumption could allow for functioning on 'biological time', and this has direct implications for human-machine integration in industrial automation, by allowing for more sophisticated robots which can waltz synchronously with humans. Meanwhile, the lower energy use of neuromorphic chips at equivalent performance will improve the portability and mobility of devices, which should greatly benefit the broad domain of neuroprosthetics, including neural and brain-machine interfaces.

In a human-machine integration perspective, from the point of view of 'removing the bottlenecks' we can see that beyond the possible applications of neural and brain-machine interfaces for therapeutics, rehabilitation and enhancement at the level of the individual, there are major future applications at the level of networked human-machine systems. Direct neural and brain-machine interfaces will help bypass the traditional sensory channels (sight, sound and touch) which act as a barrier slowing man-machine interfacing, by establishing direct communication channels between the brain and the machine (see Clausen, 2009; Schalk, 2008). However, this approach can entail conceiving of humans as mere components in complex ICT and robotics systems, components which might be optimized and interfaced in the same terms as the silicon parts. Such a narrow understanding of environment in the 'design' approach of Human-Computer Interface runs the risk, in our view, of merging with the 'rationalistic' approach.[64]

Finally, the concern to 'remove the bottlenecks' directly affects how 'the brain' is conceived and mobilized as an inspiration in ICT and robotics. In the case of the HBP, one key area that is sidestepped is the brain's affective, interpersonal and social capacities. Yet some of the work done by the Project in the domain of ICT may contribute to new developments in this area, for example through work done on modelling aspects of the visual system,[65] and more generally by improving data mining capabilities. We now turn to consider this broad area, usually termed 'affective computing' and to consider some wider challenges and issues, beyond the specific work of the HBP, that are likely to be affected by developments in ICT and Robotics to which the research of the Project is contributing, albeit often indirectly.

3. Wider challenges and issues

In this Foresight Report, we need also to look beyond the direct contributions of the HBP research on ICT and robotics that we have discussed above, and explore the ways that this relates to the many possible general changes (societal and economic) that these developments potentially bring.[66] In principle, the same underlying ICT advances that enable an intelligent application in healthcare might also enable an intelligent application in warfare. As was noted by several of our research participants it hardly seems possible (or appropriate) to judge the HBP on all possible applications of more intelligent machines.[67] The same labour saving automation advances that enable robots to take up necessary but dangerous tasks that are currently putting human labourers at risk might also be utilized to replace jobs in such a way that creates widespread unemployment and exacerbates inequalities in income.[68] Clever human computer interaction protocols that meet users' emotional expectations and thus allow computing to provide new services more effectively could also be used manipulatively or to collect data about users' feelings that they would prefer not to provide. The very same computing technologies that create valuable new commercial and social interaction possibilities can also contribute to the growing toxic burden of Waste Electrical and Electronic Equipment (WEEE). These are just a few of the plausible societal level impacts that might disturb the researcher who hopes that their research will contribute to desirable outcomes but not undesirable ones.

In this section, we select and outline just a few representative examples of these issues. In section 4.1 we discuss affective computing (which is both a computing issue and a robotics issue). Section 4.2 we briefly consider the much-discussed concern that advances in robotics will impact jobs and the economy. We then consider how military applications might emerge from civil research in computational neuroscience. Lastly, we look at the materiality of ICT artefacts – the components that go into them, and the waste that is produced, and we consider the need to think of the environmental and social or economic implications for the future ICT and robotics industry.

3.1 Affective computing

In most discussions of artificial intelligence, it is the cognitive capacities of computers that are at the forefront. However, the increasing sophistication and greater uptake of intelligent machines raises another key issue, which concerns the affective dimension of the relation between humans and such intelligent machines. The recognition of emotions is an important part of human-to-human communication, providing the context and sometimes the content of communications. Even when we do not know the precise feelings of another, we often estimate or impute a sentiment and respond accordingly. As computers come to have a greater capacity to provide information to humans in an interactive and situationally responsive way, it is likely that we will judge these interactions by the standard of our human interactions. We already, even if light- heartedly, often impute sentiment to the interactions that we have with rather stupid machines such as malfunctioning household appliances. If a computer or intelligent machine can estimate our emotions and respond in an appropriate manner, this may lead us to overestimate the actual emotional capacities of the machine. If the interaction seems like the sort of interaction we might have with another person, we may expect the machine to be able to do other things or have other qualities that a person might have. These issues are the concern of 'affective computing'.

Affective computing is the ability of computers to recognise human emotions and thus to be able to respond more appropriately.[69] The broader concerns of affective computing involves not only identifying human emotions and responding appropriately but also potentially conveying them accurately during interaction, such as a human would expect in a conversation with another human. This is to simulate human emotions such that the intended emotion might be recognizable to a human. Affective computing might potentially influence the human user, increasing persuasiveness, and in some cases deliberately generating a particular emotional response in the user.

One reviewer when introducing the logical justification for affective computing wrote (Geller, 2014: 24):

> "Nearly a decade after its retirement, the advice-spewing "Clippy" remains one of technology's most hated characters. As part of Microsoft's Office Assistant help system, the paperclip-faced avatar proposed help based on Bayesian probability algorithms: start a word-processing document with "Dear," and it offered to help you write a letter. Express exasperation, and Clippy would gleefully continue pestering you: it could not sense your mood. Perhaps Clippy would still be with us if it had employed affective computing… "

3.1.1 A brief history of affective computing

Charles Darwin (1872) was one of the earliest writers to suggest that we might be able to classify emotions scientifically.[70] But it is only more recently that science has attempted to classify and categorize these distinct bodily expressions, and therefore also make them measurable. In 1978, Paul Ekman published the Facial Action Coding System (FACS) which is used to describe how faces express emotion through muscle movements.[71] The development of computer systems that can visually recognise these movements and then classify them according to the FACS has been a strong driver of further investigation in affective computing. Facial micro-expressions, brief muscle movements that indicate an emotion sometimes too quickly for another human to see, can be discerned by more sophisticated facial recognition systems.[72]

The growing literature in the 1990s around the physio-psychological FACS system encouraged the development of affective computing. As Ayesh and Blewitt explain "[E]motion detection in general and FACS in particular each lends itself naturally to the field of pattern recognition research that has a rich literature and a wide scope of

techniques from the simple template-matching to multilayered neural nets and scores of classifiers, all of which gave researchers in computational emotions the basis to start developing emotion detection systems" (Ayesh and Blewitt, 2015). Researcher interest then extended to many more potential modes of affective computing such as the analysis of speech, video, text and images, and other forms of physiological response such as electrodermal measurements of arousal linked to the sympathetic nervous system.

Affective computing typically relies on collecting some type of physiological data from the user. The user's emotion can then be classified using some theory of how emotions express themselves physiologically. There are various theories of this; and how precisely to classify emotions is still a matter of debate as discussed below. The physiological data could be as varied as the sound of the voice, a visual image of the face, or galvanic response measured from the skin. Even word semantics or how the user physically interacts (smoothly or agitatedly) with keyboard, touch screen, and mouse are being investigated for recognizable patterns of emotion.

While computing power will increase, new algorithms may improve, and wearable devices will become smaller and more efficient at measuring response, the validity of affective computing will in some ways still have to do with the theories that are used to correlate physiological indexes to their emotion explanation.[73] Many measurement systems use a two dimensional arousal and valence scale. Sometimes a third category 'dominance' is also used to express the strength of the emotion, or in some categorizations 'control', how much control the person expressing emotion feels with regard to an event, is a variable paired with valence as in this emotion wheel.

Figure 2: The Geneva Emotion Wheel

Beyond identifying emotion, some types of affective computing hope to respond to the user with an appropriate emotion. When a computer does this, it simulates the outward signs of an emotion rather than has emotion itself. Thus it is simulating the *expression of an emotion* (providing expected emotional aspects in message communication) rather than *feeling* the emotion (Picard (2003), Schwark (2015). Nonetheless, as we discuss presently, this may lead the human interactor to mistakenly impute an internal world to the computer – or robot – and this will have consequences for future interactions.

It is worth noting that, under the label of 'affective computing', a small subset of researchers is examining the role of emotional memory in cognition. By linking an emotional valence to aspects of a cognitive representation (possibly through the hypothalamus) an organism has an advantage in understanding its environment through its past experience of similar situations. Specifically, emotional memory records a situation in reference to the conative goals of the organism. By modelling in robots this emotional

memory representation of conative goals these affective computing researchers might be said to be going beyond mere surface simulation of an emotion, at least to the extent that one understands the 'feeling' of an emotion as the expression of conative goals (see for examples, Gokcay and Yildirim, 2011).

There are various modalities for the analysis of the emotional content of interaction.[74] While some of these developments may seem complicated and distant, in fact, even today, many personal fitness trackers are used in conjunction with mobile phone applications to provide affective computing information to the increasing number of consumers who use these devices. In addition, while some methods require additional or intrusive equipment, other methods are being investigated by which data can be taken from existing user computer arrangements. Physical human-computer interaction from keyboard, mouse and touchscreen data can be examined for strength and speed of movement or other factors to find patterns typical to the user from which variations can be correlated to their probable emotion (agitation, calm). Affective computing methods are normally accomplished by initially correlating physiological measurement to some judgement of actual emotion (perhaps self-reported). When intense emotions are correlated with physiological measurement values these can be used to predict affect in computing scenarios (Stemmler, 2003; Schwark 2015, p.263). Algorithms are developed that the developers can then test against more examples where it is thought that the actual emotion is known. Often training sets of data are used to train an algorithm to recognize particular emotions.

While current measurements in laboratories can be quite accurate it is harder to create real life out-of-the-lab measurements adequate to everyday user needs. However, as computer processing speed increases, analysis techniques improve, and new consumer wearable devices proliferate, affective computing is expected to become widespread. While none of the direct goals of the HBP relate to affective computing, it is easy enough to see how applications developed from the project (such as smaller, more energy efficient and better computing power devices with a more 'neuromorphic fit' to integrate with the human nervous system) could contribute to this field. Affective computing is thus a good example of a trend in future computing to which HBP researchers will have a medium term and indirect relationship, but as we will see in the next section, its research in some areas is likely to have a significant impact on this field.

3.1.2 Markets and applications of affective computing

Affective computing is still in development but there are, in principle, numerous uses for emotion recognition in human computer interaction (HCI). Those who advocate 'positive computing' think computing could be linked to human flourishing and that computers should assist with this task (Calvo and Peters, 2013, 2014; Pawlowski, *et al.* 2015). In addition to basic desktop help systems, relevant applications might be in healthcare, education and marketing. Some of the first markets in which affective computing is being developed are E-learning, the automobile industry, robot design, smart collaboration and environment for business, game design, and disability services, particularly autism support (Bishop, 2003; Kaliouby *et al.*, 2006). Forecasters predict the affective computing market to grow, pointing to factors such as continuous research in physiological measures, increasing adoption of wearable devices, the arrival of large corporations onto the market, and better processing speed of affective computing to analyse human response in real time.[75] Further some suggest that affective computing will also "find its application in areas such as artificial intelligence where it can help in strengthening machine-human relationships."[76] Hence there is a clear link with research in the HBP on machine learning and neural networks.

Affective computing sits comfortably within the trend of social computing sometimes referred to as Web 2.0 in reference to earlier use of the web as a non-interactive repository for information, as in a static website. In platforms that enable social interaction 'on-line' for recreational or business purposes, affective computing is

attractive because of its capacity to create a more natural interaction with a computing program, or to provide users with feedback on their affective state, for example for medical or self-management reasons. It might also provide advantages to those who wish to identify the others emotions, for example in aiding those who wish to persuade another to take a particular course of action. A health management program that persuaded users to quit smoking might be considered laudable. An automated sales agent who provides accurate information and allows the user to choose whether they will purchase a commodity is no less or more bad than a human sales agent. The boundary between persuasion and manipulation is difficult to define in human-to-human interaction, and also in human computer interactions. This also touches on another difficult question – could a computer lie to you, and what therefore are the ethical issues that arise in the use of affective computing for persuasion in different contexts. In fact, in our view, this is an issue that relates less to computer ethics than to the ethics of the agent (for example, in a government office, in health, education, sales, marketing, customer service, employment assessment, insurance risk assessment) who deploys the computer in a strategy of persuasion (Fogg, 2002; Kafmann *et al,* 2010, cf. the nudge technologies promoted by Thaler and Sunstein, 2008).

Some commercial analytics companies are already beginning to incorporate affective data, particularly in support of customer service units. [77] Personal devices are now being developed to measure the affect of others.[78] Biometric identity systems have for a long time collected physiological data to confirm or deny access to systems, but now affective patterns (for example how someone types on a computer) are being developed to confirm profiles or serve as an additional check in an already secure system (Bakhtiyari, Taghavi and Husain, 2015). We therefore agree with others who have argued that the presence of affective data in social computing (from non-intrusive text based methods, or directly measurable from the increasing numbers of wearable devices that share their output into a network) requires an early debate about how to prevent problems as individuals and commercial enterprises begin to navigate and negotiate the emergent new possibilities (see, for example, Pitt 2012), 2015).

All the social and ethical issues that are already present in contemporary debates about social computing will re-emerge as affective computing becomes more and more a part of the social media and computing environment. Concerns about privacy, anonymity in mass data, the integrity of commercial messaging, invasive advertising, computer security, computer crime and fraud, anti-social behaviour facilitated by computer use, who owns our data, to what uses it is put, are likely to be exacerbated by the presence of pervasive computing, affective computing, the internet of things, and the quantified self.[79] But does affective data raise more or different ethical challenges than other data? What if anything distinguishes from other complex data patterns about users? It certainly seems more 'personal' because it refers to the users' emotional state or mood, even though it is currently largely derived from basic physiological measurement data. One can imagine these data being utilised in evaluations of employment assessment or insurance risk, skills, employment suitability, physiological or health risk. However, the same may be said of many other types of 'big data' that are increasingly used for such purposes. Perhaps the major ethical challenges arise in human-machine integration, discussed earlier in this report, or in the peculiar characteristics of human interactions with intelligent devices such as robots.

3.1.3 Affective relations between humans and robots

Affective computing is particularly relevant for robotics. Humans frequently anthropomorphise robots, whether they have affective abilities or not. In a recent example, *Harvard Business Review*, in an article dedicated to whether or not humans will accept robots and algorithms in the workplace, described an event during a robotics conference where MIT researcher Kate Darling encouraged participants to play with Pleos. Pleos are small animatronic toy dinosaurs that, when played with, react in simple ways

indicating through facial expressions or gestures that when played with they like to be petted and they don't like to be picked up by their tail, etc. Then, after an hour, she asked people to dismantle the little creatures with knives and axes, but the group uniformly refused. "We respond to social cues from these lifelike machines, even if we know they are not real" says Darling (quoted in Frick, 2015, p.84). There are many similar examples of robot devices that are designed to encourage affective relations with the humans with whom they interact. Thus Paro is a robotic baby harp seal designed to respond to cuddling, and produce warm feelings in humans who interact with it. [80] Developed by Japanese industrial robotics company AIST, it is intended to provide a therapeutic effect and companionship as a robot pet. [81] Kizmet, developed at MIT by researcher Cynthia Breazeal is a slightly more human robot designed to explore expressive social exchange between humans and humanoid robots. [82] Even more human-like (but completely without intelligence) is Geminoid DK, an android robot designed by its creator, Danish roboticist Henrik Scharfe of Aalborg University, to look (almost) exactly like himself. [83]

Because of the affective response of humans to a human seeming robot, there are dangers of overestimating the 'intelligence' or other capacities of a humanoid robot with potentially problematic consequences. But further, there are also unsettling consequences for affect itself. When a robot appears human but only somewhat human, this can have an unsettling effect on those who must deal with it. This response is often termed the 'uncanny valley'. The uncanny valley is the region of negative emotional response towards robots that seem "almost" human but not quite (McDorman *et al*, 2015). A less human seeming robot (clearly understandable as a robot) and a perfect human (for example, an actual human) do not invoke the same disquiet. Our tendency to anthropomorphise makes affective computing and humanoid robots a powerful combination. We might anticipate some tension in balancing the desire to make robots more human (but not too human) with the need for integrity of presentation so that humans who interact with robots have a basic understanding of what to expect and what possibilities or dangers there can be in social interaction with a robot.

Do these affective relations pose ethical problems? One of the participants in our research workshops spoke of the ethical danger of android robots (robots shaped like a human) that therefore appear to be intelligent in the perception of many human observers but aren't. "I believe that android robots are a deception. They create an expectation of intelligence, of intelligent behaviours that are simply not present." This is frequently referred to as the brain-body mismatch problem. No specialized or intentional affective computational abilities (expression analysis or simulation for instance) need to be given to a robot for issues of human affect to emerge. A human projects affective qualities onto the machine. However robot designers need to be aware of these affective projections, and perhaps they should be the subject of regulation, so that their potential harms or exploitative potentials can be minimized. [84]

3.1.4 Robots for care work

There have been many suggestions that robotics could provide labour to support health and care industries, particularly in relation to care of the elderly or as part of education systems for children. There are many developments in health care, social care and potentially mental health care which could greatly enhance the quality of many people's lives, and enhance the ability of people to do care work. For example, a difficult but necessary task is lifting people out of their beds or moving those who are confined to bed in order to do cleaning and to eliminate bedsores. This type of activity often results in back injuries among care workers. Robot assisted lifting can potentially extend the dignity of elderly people as well as support care workers.

However, analyses of these developments have highlighted some important ethical issues. Thus Sharkey and Sharkey raise six key areas of concern "(i) the potential reduction in the

amount of human contact; (ii) an increase in the feelings of objectification and loss of control; (iii) a loss of privacy; (iv) a loss of personal liberty; (v) deception and infantilisation; (vi) the circumstances in which elderly people should be allowed to control robots" but conclude that "If introduced with foresight and careful guidelines, robots and robotic technology could improve the lives of the elderly, reducing their dependence, and creating more opportunities for social interaction (Sharkey and Sharkey, 2012. p.27, see also Sharkey, 2008). Analogous questions were also raised by some of our research participants, who asked "What happens when we automate human care work or even the labour required for human socialization?" It is a serious challenge to provide an inexpensive, sufficiently talented, dextrous and multitasking robot which would be necessary to accomplish this. Robots able to lift the elderly are expensive and heavy, and therefore more likely to be in hospitals or nursing homes (Ford, 2015, p.162). Hence there was a concern from our research participants that this might lead to the warehousing the elderly and having them attended to by robots, reducing their amount of social interaction. Our participants argued, as do Sharkey and Sharkey, that the best model was to develop these systems such that they enhance the ability of the elderly to stay in their homes longer, and improve their mobility and independence.

More fundamentally, robotic support in care systems needs some careful consideration of the ethics of care (Mol, 2008). Care is a form of labour which is often under acknowledged. Care is not only crucial for the recipient, there is also a reciprocal aspect of the ethic of care and caregiving: people often enjoy care giving and are rewarded emotionally from giving care. By automating and routinizing care work as merely a set of tasks, we might thus be denying people both the opportunity to receive care itself, and the ability to give care. What would be the long-term impact of creating a set of values where human beings feel that care of the elderly can be left to robots, rather than being a crucial part of human identity and social values? And what of children? One of the expert presenters during our research event commented, "All these issues also exist for childcare but additionally, with children the social interactions that occur are what teach children how to be people. It would be difficult to develop ways that automate this. Certain elements of the task need to be performed by humans so using robotics as tools rather than a replacement for a child carer" (see also Sharkey and Sharkey, 2010).

In the field of care for vulnerable people, the issue of anthropomorphizing was also raised by our research participants. "We humans are compulsive, almost pathological anthropomorphisers. I think that makes us vulnerable." While it was thought that there are potentially good reasons for therapeutic robots, for example, surrogate pets like Paro (the robotic baby harp seal), there will be questions about what might be appropriate for more vulnerable people, such as children or elderly people, for example, with dementia, who are not able to judge clearly what the robotic nature of the interaction means.

Humanoid robots are already appearing in the sex industry. Opinions on this are of course diverse, ranging from enthusiasm (for example, to address how access to sexual satisfaction is unevenly distributed in human society) to anti-prurient moral scruple. But the concern that care is a practice that maintains our societal and individual values of concern for other persons has also been raised in this area. Some proponents of love and sex with robots have imagined robots that could be loved and (to the extent possible) love back. Some opponents felt the paradigm of sex with robots was inspired by prostitution and that, like purchasing prostitution, raised the ethical question of whether someone with money and power should be able to buy the means to satisfy their needs irrespective of the consequences for the other. However, given the ready availability and increasing acceptance of a range of other devices to satisfy sexual desires, it is unlikely that sexual robots will be regulated in relation to any aspects other than their safety.

3.2 The Impacts of Robotics and Automation on the Economy

Most analysis of robotic deployment in the economy to date has been focused on the so-called 4Ds of robotics: areas of work which are dangerous, dull, dirty, or difficult. Robots are often required for these types of labour. In most such situations, such as in an automobile factory, robots do not entirely replace human labour, but they undertake simple or repetitive tasks, replacing many workers, while remaining human workers manage the robots. Humans are still needed because what seems like simple tasks to us may be quite difficult for a robot (anticipating or compensating for some interference in the process - for example a spare part falling into an assembly line). There is little possibility in the immediate future that a robot will be sufficiently autonomous to fully replace a human labourer in all his or her capabilities. But there are still many opportunities for improved intelligence or other capacities in automation to improve production. For example, greater optical processing capacity for robots may in the future enable automated processes to be more autonomous, more capable of recognizing and compensating for interferences in the task processes, thus requiring less supervision. But not all jobs can be done by a robot. In the short term humans tend to overestimate robotic capabilities, projecting human psychology and cognitive abilities that robots do not have. What we humans think of as relatively simple tasks, low skilled labour, involves many situational adjustments that robots do not find it easy to manage. Such tasks often also involve relatively basic human interaction skills which are difficult for intelligent machines to learn. However, as we discussed earlier in this report, much will depend on how new technologies are deployed, whether they are seen as labour replacement strategies, or as labour augmentation strategies creating new products and services that would not have been available before the combination of humans and intelligent machines.

Better machine intelligence may also allow for categories of human labour to be performed by intelligent machines that have previously not been associated with automation. As we have indicated, the typical image of automation is in traditional manufacturing (for example, in an automobile manufacturing plant) with many robots carefully maintained by a few blue-collar labourers. However more intelligent systems will be able to perform tasks in what was traditionally white-collar labour. Lawyers, journalists, and other white-collar professionals are not traditionally thought of as being threatened by new automation technologies. However, in the legal profession, the task of legal discovery, a task often assigned to junior or entry level legal professionals, can now be automated. Existing e-discovery firms are able to reduce costs and expand the productivity per lawyer of a law firm. The task of researching relevant case law can now also be performed on intelligent systems (Colvin, 2015, p.17). Even the prediction of U.S. Supreme court decisions has now been done by intelligent systems (cf. Ruger *et al*, 2004). While it is implausible that an autonomous robotic system would or could be designed to entirely eliminate the need for a human lawyer, such automated systems to assist relatively senior lawyers may still result in significant replacement of those junior legal staff required to undertake this more routine work.

Drawing on considerations such as these, an Oxford Business School report found that up to 47% of total US employment is at risk of computerisation over the next two decades. [85] Transportation and logistics occupations, office and administrative support workers, and labour in production occupations, were deemed to be the most likely to be substituted by ICT. Service sector jobs, where much previous growth has happened, may see greater levels of automation. Occupations with low wages and educational attainment were correlated with high risk of substitution.[86] As perception and manipulation challenges in robotics are overcome, more occupations could potentially be substituted. The professions that required creative and social intelligence were the least likely to be computerised in the near term but if bottlenecks in computing were overcome then a second wave of computerisation might be possible.

A number of factors will potentially limit this process. Regulation may create some barriers. For example, health care is a highly regulated field and therefore we would expect more challenges to automating health care work. In industries with low wage costs

where capital investment would be required to substitute for wages, it might prove more cost effective to remain with human workers rather than automate: for example, the fast food service sector has traditionally not received capital investment in automation for this reason, but this may eventually change (Ford, 2015, p.204).

Much of job loss from the application of new technologies comes from what Brynjolfsson and McAfee (2015) argue is the least creative and most direct application of new technologies: find a human labour process and automate it. They advocate two responses: greater focus by the employee on what they can provide that a machine cannot; and greater focus by the employer on entrepreneurship, on how intelligent machines can amplify the power of the existing workforce. The skill areas that robots still find difficult include high-end creativity, interpersonal relations (so-called affective labour, adding value to a situation because you know how to care about, communicate with, and share interpersonal value with other humans in a transaction), and dexterity/mobility tasks. Education systems may need to emphasize more of these types of skills in the future to increase the employability of graduates.

Machine substitution for labour is not the only way. Greater emphasis on entrepreneurship would instead ask the question "how can I have this machine and this human work together to do something never done before and create something that will be more valuable in the marketplace?" (Brynjolfsson and McAfee, 2015, p.74). By starting with what the workforce can now do and looking at complementarities entrepreneurs may be able to increase productivity and augment demand for skilled labour. Brynjolfsson (2015, p. 74) has argued that engineering design philosophy may be the problem and that a "New Grand Challenge for Entrepreneurs, Engineers, and Economists" is to compliment rather than substitute labour. While this is certainly a potentially positive scenario, author Martin Ford is sceptical, suggesting there may be too few market incentives for this situation to substantial enough to offset job losses (Ford, 2015, p. 253). In a competitive market a people oriented design in competition with a more fully automated business, would have to be able to offset its additional wage costs by being significantly less expensive or provide such additional value to customers as to make these additional costs offset by additional revenue.

We have merely given a brief summary of some of the key points in a growing literature about the impact of robotics and automation on the future economy.[87] While there is much more to be said – and we refer the reader to the reports we have referenced here – it is clear that we do not need to speculate about some future 'singularity' to realise that through such short to medium term impacts, the rise of more intelligent machines, will indeed prove to be a disruptive technology.

3.2.1 Disruptive Technologies

A disruptive technology is an innovation that establishes new value and a new market which displaces a previous market (Christenson, 1997).[88] There are usually advantages for the consumer in that the innovation creates something that is more desirable than its predecessor, that allows it to begin displacing (disrupting) the existing market.[89] Such disruptive technologies can create considerable value for those who hold intellectual property rights or develop a business strategy which can exploit this capture of value from disrupting an existing market.

In digital economies, something of value can be reproduced by its owner at almost no expense (that is to say at close to zero marginal cost), allowing owners who capture something of value to accumulate wealth quickly. As Brynjolfsson and McAfee (2015, p.70) point out, this feature of such economies can greatly increase inequalities in wealth: they have a "fractal-like quality, with each subset of superstars watching an even smaller group of uber-superstars pulling away." The traditional view in economics is that new technology increases labour productivity and therefore wages and employment rise as growth and new opportunities are unveiled. Indeed, this was the case for many advanced countries during

the early second half of the 20th Century. Growth in labour productivity and real Gross Domestic Product (GDP) roughly kept pace with growth in private employment and median family income. However, since the 1980s, median family income began to uncouple from this parallel growth and more recently growth in private employment has also slowed (Brynjolfsson and McAfee, 2015, p.70). While many factors are involved, at least some of these changes are due to the introduction of automation and new information technologies. New applications of intelligent machines have the potential to exacerbate this trend, that is to say, to be both economically and socially disruptive.

The discussion of the impact of robotics on the economy is already being reframed in terms of its potentially disruptive consequences. If robotics and automation significantly replace human employment, this may also undermine consumer purchasing power and consumer demand, as well as threatening the incomes of those who are displaced from the labour force. Some have suggested a collective commitment to 'basic income' for all members of society will be necessary to maintain economic and political stability, and also to continue with the values and way of life that EU citizens have come to expect. While a few have suggested that in the future humans will celebrate a life of leisure while machines look after our needs (Srnicek and Williams, 2015), past historical analysis suggests less likelihood of an equal distribution of income from technological increase.[90] Additionally, countries with high levels of income inequality also have increased social challenges (Wilkinson and Pickett, 2009).

During our research workshops, however, one of our informants voiced scepticism about how applicable the advanced brain modelling undertaken by the HBP will be to the development of intelligent machines. A general model of the human brain may not be relevant to such machines, which tend to be very application specific. However, as we have argued earlier, basic research undertaken by the HBP will have relevance to intelligent machine applications, even if much of the translation to applications is undertaken by private commercial enterprises. It would be unrealistic to expect researchers within the project to address these social and economic issues directly. Potential solutions to these types of large-scale societal challenges are more likely to emerge in the realm of politics than in scientific research policy. Nevertheless, given the commitment of the European Union to enhancing the wealth and well-being of its citizens, it is appropriate, and indeed necessary, to pay attention to the medium term economic consequences of the kinds of developments that might result from the translation of the research of the HBP and similar work in advanced ICT and robotics.

3.3 Further Issues

In this final section of our report, we consider how military applications might emerge from civil research in computational neuroscience. We also discuss the importance of addressing the materiality of computing, what resources are required, what waste is produced and how should we think of the environmental and social or economic implications of sourcing materials for an ever expanding industry of ICT and robotics.

3.3.1 Dual Use

The potential uses of intelligent machine advances in remote-controlled weaponry is a well-recognized concern.[91] In recent years there has been much controversy about the military use of semi-autonomous drones to attack enemy combatants.[92] However there are many other potential military applications from new advances in computing such as battlefield enhancement of soldiers and new forms of intelligence gathering.

In the USA, the Defence Advanced Research Projects Agency (DARPA) is one of several agencies providing the overall budget for the BRAIN Initiative. DARPA's goals in brain research are partly related to veterans' after-combat mental health, but there is also an explicit interest in enhancing the combat effectiveness of military personnel, and in other

technologies, such as 'brain reading' technologies, that may have uses in the security apparatus.[93] In Europe, the situation is different, in that the HBP is committed not to accept military research funding or to engage in research with direct military application. However, history shows us that civilian scientific and neuroscientific research has often resulted in military applications, over which the original researchers have little control. For example, shortly after acetylcholine was discovered to be a neurotransmitter, the G-series of nerve agents (including sarin) was discovered during civilian research into pesticides. Other civilian discoveries led to the more deadly V-series, as well as the development of 'incapacitants' sometimes thought to be 'less lethal' but with potentially lethal consequences.[94]

There are parallels here with current research in artificial intelligence. 'Brain-like machines' are likely to have numerous civilian applications ranging from self-driving cars and medical informatics, to brain controlled prosthetic limbs. Yet these developments may also directly or indirectly lead to complex (autonomous) weapons systems and new potentials for intelligence gathering and surveillance. Thus, despite the prohibition on military funding and applications in the HBP, the advances in knowledge and technologies that it produces, and the basic ICT and robotics tools that it generates, are likely to have an indirect, medium-term impact in the field of defence and security. The challenges raised by such potential 'dual use' issues will be the subject of a subsequent report by the Foresight Lab.

3.3.2 Electronic waste and anticipatory life-cycle assessment

There is one important difference between the brain and its simulation *in silica* that is seldom mentioned: the brain is degradable and can be recycled, while the ICT age has not only led to a geopolitically significant search for the rare metals and other materials that are crucial for components of ICT systems[95] - which is already leading to predictions of shortages and crises[96] - but has generated vast quantities of waste. In taking inspiration from the brain, few consider the materiality of computing: the plastics, metal, the rare earth minerals, the silicon impurity alloys, the molecules, the electricity, cables, energy sources and so forth that circuits, keyboards, robot eyes and Boltzmann machines are made of and powered by. Yet these are increasingly critical issues, if the goals and targets of the United Nations' 2030 Agenda for Sustainable Development adopted in September 2015 are to be met (see Leisinger, 2015). That is to say, we need to consider the type and sourcing of materials required to make new computing technologies; the amount of non-renewable resources required to make new computing products; the amount of electronic waste (e-waste) created by the disposal of computers, computer products, and electronic equipment.

Roughly 42 million metric tons of electronic waste or e-waster are produced globally each year.[97] According to the EU: "Waste electrical and electronic equipment (WEEE) is currently considered to be one of the fastest growing waste streams in the EU, growing at 3-5 % per year. WEEE contains diverse substances that pose considerable environmental and health risks if treated inadequately."[98] Roughly 13% of this e-waste ends up disposed illegally in foreign countries.[99] A significant amount of waste electrical and electronic equipment (WEEE) is comprised of computers (including common household appliances that incorporate simple computing programs). As we witness the development and use of more intelligent ICT applications, where households or institutions might have networked sensors and cloud powered ambient intelligence, we need to consider not just the materiality of new computing and intelligent machines, but more profoundly what potential changes in sourcing, production, circulation and waste they will enable or result.

Such questions point to the need for Life Cycle Assessment to consider possible material futures of computing. Life Cycle Assessment is an engineering and design assessment framework that begins with a use case for product and incorporates knowledge of the extended life cycle of the product (including materials sourcing, energy efficiency and

waste or recycling planning) into the design criteria.[100] Anticipatory life cycle assessment (Dwarakanath, 2013) extends this process to consider how we might include future uncertainty and unpredicted outcomes in the results of a life cycle assessment (use, misuse, creative use, sourcing, missourcing, creative sourcing, disposal, misdisposal, creative disposal or recycling, etc.). These life cycle product assessment techniques enable engineers and designers to consider the environmental and social impacts of new technologies. Can we find ways to make our ICT equipment from recyclable materials? Can we free them from their demand for rare metals? How can we tackle what the US research and advocacy group Demos has termed 'high tech trash'?[101]

4. Researcher Awareness

As we found in our data collection, workshops, and webinars, many HBP researchers have awareness of some of the issues surveyed in our report and some are trying to think through the possible, even if medium term and indirect, implications of their laboratory work. It is true, however, that the many possible uses of the new ICT and robotic technologies enabled by HBP research are exceptionally difficult to predict. This is, in part, because of the gulf between laboratory research and the development of applications and products; as several of our participants argued, the research of the HBP will produce basic tools, analogous to hammers or transistors, which can be utilised in many different applications with a wide range of potential consequences. Bridging the gap between the laboratory research and the application or deployment stage was raised by our participants. It is for these reasons that the Foresight Lab has chosen an anticipatory approach for its work (Barben *et al*., 2008; see also Stilgoe, Owen and Macnaghten, 2013).[102]

What is the role of the scientist in the lab in considering these high variance futures? An important first point of encounter with a future technology or application is the researcher in the lab working with materials, models, and concepts. Some design decisions in an experiment may become part of an industrial process. When new objects or processes emerge from laboratories they remake the world in tiny ways by recreating the circumstances of that laboratory outside of itself (Fisher, Majahan and Mitcham, 2006). Hence it is important that practices of reflection on the potential consequences of decisions in the lab are built into the research process itself (Youtie, 2010). At the very least, it is important for the researcher to understand his or her own location in the research system and to have an understanding of the context and concerns of the potential users of their research – other labs, commercial enterprises, creative adapters, policy makers, and many others; the list is a long one. We drew attention to the importance of such an awareness in our Foresight Report on Future Neuroscience, focussing in particular on an attention to the role of the researcher in their wider research and development community, and the implications of different ways in which they interested with that community, for example in sharing data, making results open source, opening up to contributions from outside and in turn making contributions to other endeavours and projects. This skill of 'reflexivity' is not only crucial for the local goals of building future neuroscience and future computing infrastructure; it is also the basis of thinking about the ethics of responsible research. In the long-term, a culture of Responsible Research and Innovation (RRI) seeks to ensure that the outcome of research contributes to the welfare of the wider community of citizens, both within and outside the European Union.

By taking an integrative approach, where public concerns, external expertise and entrepreneurial, civil society, policy and other stakeholder interests are integrated into the conversation about the strategy and objectives of its research, the HBP could contribute to a culture of well-informed reflection and 'scientific citizenship', which is an aspiration of the European Commission for all its funded research.

5. Conclusions and Recommendations

5.1 Concluding remarks

Let us briefly summarise some of the main issues that we have addressed in this report. First, we have argued that, in view of the current hype and confusion surrounding the potential of machine intelligence to overtake human intelligence, there is a need for short term social and ethical issues of narrowly specialised Artificial Intelligence to be clearly demarcated from speculation about long term potential existential risks, and addressed as a priority, at national and trans-national levels. These debates should not be left to the private sector and private philanthropy, but should be addressed openly and democratically.

Further, we have argued that the human should not be 'designed out' of technology, instead the richness of human life in all its dimensions should be put at the centre. A human-centred design approach that does not narrowly focus on the individual but takes into account the wider socioeconomic context, can bring to light a broader, and different, range of social and ethical issues about ICT and robotics. Our strong view is that such an approach should inform strategic choices and decisions driving research and innovation for future computing and robotics.

The dominant focus on 'removing the bottlenecks' affecting information-processing performance and energy efficiency in ICT directly affects how the brain is conceived and mobilized as inspiration for future computing and robotics. However, we have argued that this narrows the ways in which the brain is taken as inspiration, focussing upon its information processing capacity and its low energy consumption. We have drawn attention to some other characteristics of the brain that are often overlooked in designing future computing and robotics. In particular, no technology that uses the brain as its inspiration should neglect the capacity of the brain, in constant interaction with its interpersonal and physical environment, to develop and sustain a social and emotional mind.

In this context, we have also drawn attention to developments in affective computing, and their implications especially for those who advocate a role for humanised robots in care work, notably for older or vulnerable people. We have suggested that in human relations, care should not be reduced to merely the automation of a series of specific tasks. While there are obvious benefits of robots in this domain, care is a human interaction involving genuine reciprocation of feelings and obligations, and these entail exactly the kind of high-level affective and interpersonal skills that are currently challenging for robots. We should not expect more from robots than they are able to give, and it is important that policy makers recognise the limits of robots for such work, as well as their benefits.

Additionally, we have pointed to the fact that the brain is constructed from readily available organic materials, and is fully biodegradable. No strategy for the future of ICT and robotics can neglect the fact that this is not the case with our current generation of devices, whose growth has not only led to very significant national conflicts and geopolitical rivalries, but which has also generated very large quantities of potentially hazardous waste materials. It is therefore crucial that those urging and funding the development of ICT and robotics, do so within a clear analysis of the life cycle of the material artefacts that are proliferating in consequence, and devote both attention and funding to the development of more sustainable and ecologically desirable alternatives.

Like many others, we have also drawn attention to the potentially disruptive consequences of likely developments in ICT and robotics on economic and social life. Given the commitment of the European Union to enhancing the wealth and wellbeing of its citizens, we have suggested that it is appropriate, and indeed necessary, for policy makers at

European level to pay attention to the medium term economic consequences of the kinds of developments that might result from the translation of the research of the HBP and similar work in advanced ICT and robotics.

5.2 Recommendations

- Our overview of the potential contributions of the HBP to future ICT reveals the many interdependencies existing between different parts of the project, and through them, the close and often reciprocal relations existing between the various fields of research involved and their applicative domains. We recommend that those responsible for the coordination of the Project pays close attention to these deep interconnections. The Cross-Design Projects planned for the Operational Phase will use these interconnections to build some synergies, however we consider that more is needed. Hence we recommend that a systematic, project-wide reflection should be conducted to take stock of all these synergetic potentials and devise a Responsible Research and Innovation roadmap for building on them.

- In view of the potential contributions of the HBP to machine intelligence, it is necessary to recognize that a number of Subprojects across the project will collectively contribute to this domain. Thus we recommend that HBP researchers engage with current initiatives and debates, in Europe and elsewhere, that consider ways to make Artificial Intelligence ethical and socially beneficial. To facilitate this, we recommend that the HBP nominate one of its members to act as coordinator on these questions.

- In its Operational Phase, the HBP moves into Horizon 2020, the EC funding programme expected to position Europe as a world-class competitor in ICT research and digital innovation through 'open science' and the development of innovative public-private partnerships. We recommend that those responsible for the scientific direction of the HBP set out policies that seek to ensure that the research results of public-private partnerships are subject to the same requirements of openness, so that they can be ethically and in a manner that is beneficial to European citizens.

- With regard to the projected offering of commercial services to industry in particular by the Neurorobotics Platform, we recommend that an evaluation of proposed applications in terms of social and ethical impact should be an integral part of the terms of service.

- We recommend that the scientific directors of the HBP re-iterate their commitment to responsible research and innovation by supporting the development of a coherent programme of researcher awareness, engaging the whole HBP research community, during the operational phase of the project.

6. Endnotes

[1] Originally popularized by science fiction writer Vernor Vinge, the idea of a singularity has been most prominently promoted by futurist Ray Kurzweil through books and film such as "The Singularity is Near". While "the singularity" is an important cultural reference for some A.I. enthusiasts, and has more broadly become a discussion point within A.I. societal impacts round tables, not all speculative futures or conceptions of A.I. surpassing human intelligence are organized in relation to this. Humans, even from ancient myth, have often held concerns that their creations would somehow surpass or undo them.

[2] https://www.humanbrainproject.eu/en_GB/mission

[3] https://www.humanbrainproject.eu/en_GB/high-performance-computing-platform

[4] https://www.humanbrainproject.eu/en_GB/neuromorphic-computing-platform1

[5] http://www.kip.uni-heidelberg.de/cms/vision/projects/facets/neuromorphic_hardware

[6] https://www.humanbrainproject.eu/en_GB/neurorobotics-platform1

[7] For example in the many reports produced by the 'innovation charity' NESTA: http://www.nesta.org.uk/

[8] Available at https://www.humanbrainproject.eu/documents/10180/1055011/SP12_D12.1.1_FINAL.pdf

[9] Available at https://www.humanbrainproject.eu/documents/10180/1055011/SP12+D12.1.2+FINAL.pdf/.

[10] These webinars occurred on January 7th and October 9th 2015. Some of the presentations can be viewed at http://www.tekno.dk/article/dual-use-and-neuroscience-invitation-to-an-online-debate/?lang=en and http://www.tekno.dk/article/future-computing-and-robotics/?lang=en

[11] For example, the perception that ethics are to be outsourced to sociologists, philosophers and historians of science with research policy insight.

[12] http://www.bbc.co.uk/news/technology-30290540

[13] European Commission Digital Agenda for Europe: Towards a thriving data-driven economy. https://ec.europa.eu/digital-agenda/en/towards-thriving-data-driven-economy.

[14] Human Brain Project Framework Partnership Agreement. https://www.humanbrainproject.eu/documents/10180/538356/FPA++Annex+1+Part+B/.

[15] Ibid.

[16] Ibid.

[17] Ibid.

[18] Ibid.

[19] Robot & Gran: Would you leave your ageing parent or grandparent in the care of a robot? http://www.bioethics.ac.uk/events/Robot-and-Gran-Would-you-leave-your-ageing-parent-or-grandparent-in-the-care-of-a-robot.

[20] Wilson, HJ. (2015). 'What Is a Robot, Anyway?', *Harvard Business Review*. https://hbr.org/2015/04/what-is-a-robot-anyway.

[21] See for instance, http://time.com/46777/your-data-is-dirty-the-carbon-price-of-cloud-computing/; http://www.theguardian.com/environment/2011/sep/08/google-carbon-footprint; http://www.businessgreen.com/bg/news/2288738/googles-footprint-falls-as-users-emit-eight-grams-of-co2-per-day;

[22] See for instance, http://www-cps.hb.dfki.de/about/mission.

[23] Extract from the transcript of interview by Robotics Trends of Mario Tremblay, co-founder and CEO of RobotShop, posted 07/07/2011 (http://www.robotshop.com/blog/en/myrobots-com-pioneers-cloud-robotics-interview-with-mario-tremblay-1348). For a general introduction to cloud

robotics, see for instance https://sites.google.com/site/ruijiaoli/blogs/page; for current efforts in this domain, see for instance the project RoboBrain at Stanford (http://robobrain.me/#/), backed by funding from the National Science Foundation, the Office of Naval Research, Google, Microsoft, and Qualcomm (http://www.wired.com/2014/08/robobrain/).

[24] A summary of the event along with video recordings of the online debate is available at: http://www.tekno.dk/article/future-computing-and-robotics/?lang=en.

[25] The entire presentation "Dancing with pixies" by Professor of Cognitive Computing J. Mark Bishop, Goldsmiths, University of London is available at: https://www.youtube.com/watch?v=esiiTEJHKaw. The Commentary by ICL Professor of Computer Engineering Steve Furber, University of Manchester, co-director of HBP Subproject SP9, is available at: https://www.youtube.com/watch?v=uHJg6w5fIrY. The link to the SpiNNaker Home Page is http://apt.cs.manchester.ac.uk/projects/SpiNNaker/. The virtual versus physical systems argument was part of the email exchange that followed the webinar. We would like to thank Mark Bishop, Steve Furber and David Lester for their reviewing of, and invaluable contribution to, this part of the report.

[26] Personal communication Bernd Stahl / Mark Shaw, De Montfort University, in Researcher Awareness Work Package (WP12.4).

[27] A HBP Subproject director has ironically commented that, throughout its history, whatever has worked in AI has become a field in its own right, ostensibly distancing itself from AI, while AI has been left with all the things that have failed to work (Personal communication Bernd Stahl / Mark Shaw, De Montfort University, in Researcher Awareness Work Package (WP12.4)).

[28] See for instance, Hawkins, J. & Dubinsky, D. (2016). 'What is Machine Intelligence vs. Machine Learning vs. Deep Learning vs. Artificial Intelligence (AI)?', Numenta blog (http://numenta.com/blog/machine-intelligence-machine-learning-deep-learning-artificial-intelligence.html).

[29] See for instance a classic textbook of Artificial Intelligence, by now in its 3rd edition: Russell, S. & Norvig, P. (2010). *Artificial Intelligence: A Modern Approach, 3rd Edition*. New York: Pearson. The link to the associated website is: http://aima.cs.berkeley.edu/.

[30] https://www.neuron.yale.edu/neuron/.

[31] http://www.nest-simulator.org/.

[32] Jonke, Z., Habenschuss, S. & Maass, W. (2014). 'A theoretical basis for efficient computations with noisy spiking neurons'. *ArXiv* (http://arxiv.org/abs/1412.5862v1).

[33] Personal communication Bernd Stahl / Mark Shaw, De Montfort University, in Researcher Awareness Work Package (WP12.4).

[34] Cellan-Jones, R. (2014). 'Stephen Hawking warns artificial intelligence could end mankind'. http://www.bbc.co.uk/news/technology-30290540.

[35] Google DeepMind website can be accessed at http://deepmind.com/.

[36] Rowan, D. (2015). 'DeepMind: inside Google's super-brain'. *WIRED UK*. http://www.wired.co.uk/magazine/archive/2015/07/features/deepmind/viewall. For a recent debunking of the current 'myths' of Artificial Intelligence by an 'anti-hype' proponent, see for instance Goodwins, R. (2015). 'Debunking the biggest myths about artificial intelligence'. *Ars technica*. http://arstechnica.com/information-technology/2015/12/demystifying-artificial-intelligence-no-the-singularity-is-not-just-around-the-corner/.

[37] Mizroch, A. (2015). 'Google on Artificial-Intelligence Panic: Get a Grip'. *WSJ.Digits*. http://blogs.wsj.com/digits/2015/06/08/google-on-artificial-intelligence-panic-get-a-grip/.

[38] Rowan, D. (2015). Ibid.

[39] https://www.fhi.ox.ac.uk/.

[40] http://futureoflife.org/.

---

[41] See for instance the project 'Control and Responsible Innovation in the Development of Autonomous Machines' launched in August 2015 by the Hastings Center, funded by the Future of Life Institute: http://www.thehastingscenter.org/Research/Detail.aspx?id=7754.

[42] See https://www.lucid.ai/ethics-advisory-panel and http://www.crassh.cam.ac.uk/events/26614?utm_source=CRASSH+Newsletter&utm_campaign=de9a58b254-Next+week+at+CRASSH%2C+University+of+Cambridge&utm_medium=email&utm_term=0_3a4085b61b-de9a58b254-214706093.

[43] Levy, S. (2015). 'The Deep Mind of Demis Hassabis'. *Backchannel*. https://backchannel.com/the-deep-mind-of-demis-hassabis-156112890d8a#.eai43qowf. Rowan, D. (2015). Ibid. For a sceptical opinion on the existence of Google's AI ethics advisory committee, see O'Gieblyn, M. (2016). 'As a God might be. Three Vision of Technological progress.' *Boston Review*. https://bostonreview.net/books-ideas/meghan-ogieblyn-god-might-be?platform=hootsuite.

[44] http://www.thehastingscenter.org/News/Detail.aspx?id=7755.

[45] Reported by TechCrunch, they were raised during a roundtable entitled 'Artificial Intelligence: a force for good?' co-hosted by SwiftKey and Index Ventures in London in September 2015, which brought together Artificial Intelligence experts from academia, the private sector and the literary world, to discuss how AI could be transformed into a force for good and its risks managed: see Lomas, N. (2015). 'Not Just Another Discussion About Whether AI Is Going To Destroy Us'. *TechCrunch*. http://techcrunch.com/2015/09/06/ais-real-risks-and-benefits/. The podcast of event is available from: http://indexventures.com/news-room/index-insight/artificial-intelligence-a-force-for-good).

[46] Over the past few months, companies like Google, Facebook, Baidu, Microsoft – Add refs) have somewhat responded to this often raised criticism by releasing open source a number of their algorithms (and even some hardware design in the case of Facebook). This is a first step, but it remains to be seen how critically central to their core R&D are the released sources - and if anything, it should draw attention to the fact that algorithms on their own may not have such intrinsic value, and that instead the real value may lie in a combination of algorithms with (proprietary) datasets and (proprietary) hardware architectures. (On Facebook: http://www.theregister.co.uk/2015/12/10/facebooks_open_ai_hardware_release/; on Microsoft: http://www.i-programmer.info/news/105/9171.html; on Google: https://www.tensorflow.org/; on Baidu: https://www.technologyreview.com/s/545486/chinas-baidu-releases-its-ai-code/).

[47] See https://ec.europa.eu/programmes/horizon2020/en/what-horizon-2020; https://ec.europa.eu/digital-agenda/en/research-and-innovation.

[48] David A. Mindell is a professor of aeronautics and astronautics and the Dibner Professor of the History of Engineering and Manufacturing at MIT, and with twenty-five years of experience as an engineer in undersea robotic exploration.

[49] Personal communication Bernd Stahl / Mark Shaw, De Montfort University, in Researcher Awareness Work Package (WP12.4).

[50] This particular critique has been formulated for instance by Morozov, E. (2015). 'The Taming of Tech Criticism'. *The Baffler*, No. 27. http://thebaffler.com/salvos/taming-tech-criticism.

[51] See for instance, http://www.investopedia.com/articles/investing/052014/how-googles-selfdriving-car-will-change-everything.asp; http://www.businessinsider.com/report-10-million-self-driving-cars-will-be-on-the-road-by-2020-2015-5-6?IR=T; http://www.autobytel.com/car-ownership/advice/10-benefits-of-self-driving-cars-121032/.

[52] Beside the well-known example of Google, see for instance a recent WIRED post on Ford: http://www.wired.co.uk/news/archive/2016-02/24/ford-autonomous-cars-kill-people?utm_source=Adestra&utm_medium=email&utm_campaign=wired%20weekender%2026.02.16.

[53] For traffic-related death-rate, see for instance https://en.wikipedia.org/wiki/List_of_countries_by_traffic-related_death_rate. For reduction in numbers of road traffic deaths expected from deployment of self-driving cars in the UK, see for instance, http://www.businessinsider.com/report-10-million-self-driving-cars-will-be-on-the-road-by-2020-2015-5-6?IR=T. For contribution of road traffic to air pollution and related deaths, see for

instance: https://www.rcplondon.ac.uk/projects/outputs/every-breath-we-take-lifelong-impact-air-pollution; http://www.nature.com/nature/journal/v525/n7569/full/nature15371.html; https://www.foe.co.uk/sites/default/files/downloads/road_air_pollution_health.pdf; http://acm.eionet.europa.eu/reports/docs/ETCACM_TP_2012_14_traffic_contribution_city_aq.pdf.

[54] See for instance, https://www.technologyreview.com/s/542626/why-self-driving-cars-must-be-programmed-to-kill/; http://ed.ted.com/lessons/the-ethical-dilemma-of-self-driving-cars-patrick-lin; http://www.iflscience.com/technology/should-self-driving-car-be-programmed-kill-its-passengers-greater-good-scenario; http://www.theatlantic.com/technology/archive/2013/10/the-ethics-of-autonomous-cars/280360/; http://www.wired.co.uk/news/archive/2016-02/24/ford-autonomous-cars-kill-people?utm_source=Adestra&utm_medium=email&utm_campaign=wired%20weekender%2026.02.16.

[55] Robot & Gran: Would you leave your ageing parent or grandparent in the care of a robot? http://www.bioethics.ac.uk/events/Robot-and-Gran-Would-you-leave-your-ageing-parent-or-grandparent-in-the-care-of-a-robot.

[56] Human Brain Project Framework Partnership Agreement, document available at: https://www.humanbrainproject.eu/documents/10180/538356/FPA++Annex+1+Part+B/.

[57] See for instance: Hu, J. (2014). 'The Silenced Voice: Examining the Evolving Debate on Pediatric Cochlear Implantation'. (http://business-ethics.com/2014/09/30/1852-the-silenced-voice-examining-the-evolving-debate-on-pediatric-cochlear-implantation/); Mager, WL. (2013). 'Billy 2.0 – The Beginning of the End'. (http://wlmager.com/billy-2-0-the-beginning-of-the-end/); Gupta, S. (2014). 'The Silencing of the Deaf - How high-tech implants are being blamed for killing an entire subculture'. *Matter*. (https://medium.com/matter/the-silencing-of-the-deaf-22979c8ec9d6#.jh4okusuu).

[58] NERRI project homepage: http://www.nerri.eu/eng/home.aspx.

[59] http://www.virtualfutures.co.uk/event/vfsalon-neurostimulation/.

[60] Rose, N., Aicardi, C., and Reinsborough, M. (2015). Foresight Report on Future Neuroscience. http://www.kcl.ac.uk/sspp/departments/sshm/research/Research-Groups/Biomedicine-Ethics-and-Social-Justice/BESJ-Projects/HBP/SP12-D12-1-2-Future-neuroscience-online.pdf.

[61] The Human Brain Project-Preparatory Study (HBP-PS) Consortium (2012). The Human Brain Project: A Report to the European Commission, April 2012. https://www.humanbrainproject.eu/documents/10180/17648/TheHBPReport_LR.pdf/.

[62] Human Brain project Framework Partnership Agreement, p. 8. The document is available at: https://www.humanbrainproject.eu/documents/10180/538356/FPA++Annex+1+Part+B/.

[63] For recent projects, see for instance: OUTREACH@DARPA.MIL. (2016). 'Bridging the Bio-Electronic Divide. New effort aims for fully implantable devices able to connect with up to one million neurons'. http://www.darpa.mil/news-events/2015-01-19; Gershgorn, D. (2016). 'IBM research thinks it's solved why the brain uses so much energy'. http://www.popsci.com/ibm-research-thinks-its-solved-brain; http://www.salk.edu/news-release/memory-capacity-of-brain-is-10-times-more-than-previously-thought/; http://www.eurekalert.org/pub_releases/2016-02/cmu-cmj020316.php; http://science.sciencemag.org.libproxy.ucl.ac.uk/content/351/6275/799.full; http://www.theregister.co.uk/2015/02/17/darpas_google_glass_will_plug_straight_into_your_brain/.

[64] The paper by John Licklider (Licklider, 1960) that many in HCI consider as foundational of their field was already pregnant with this risk.

[65] In the Ramp-Up Phase of the Human Brain Project, Work Package 11.1 (WP11.1) in the Applications Subproject (SP11) developed a retina model to be used as a sensor as input for cortical models, and Work Package 11.3 (WP11.3) in the same Subproject prepared for application cases from the Neuromorphic Systems, among which "Exploitation of Feedback in Ultra-fast Spiking Visual Architectures" (Task T11.3.3) and "Asynchronous Computational Retina" (Task T11.3.5) (Deliverable D11.4.3 "Applications: First Science Report", available at https://www.humanbrainproject.eu/documents/10180/1055011/SP11_D11.4.3_FINAL.pdf/).

[66] An EU funded research project for Ethical Issues of Emerging ICT Applications (ETICA) identified "11 emerging ICTS, understood as those socio-technical systems that are currently being seen as having the potential to significantly change the way humans interact with the world in the medium term future of 10-15 years" (B. Stahl, 2013). These included affective computing, ambient intelligence, artificial intelligence, bioelectronics, cloud computing, future internet, human-machine symbiosis, neuroelectronics, quantum computing, robotics, virtual/augmented reality. For more details see http://www.etica-project.eu

[67] The theory that the research level and the application level of science are two separate strata without any meaningful connection by which scientists might have an ethical impact on the use of their research has been discussed earlier.

[68] Indeed, in some extreme scenarios, the displacement of paid human labour by intelligent robotic machines may even pose a threat to the imperative of ever increasing consumption – using purchasing power derived from wages – that is the foundation of the present economic system, and provides the demand necessary to drive the system forward. This is discussed later in this report.

[69] The term 'affective computing' is often associated with Rosalyn Picard of the MIT Medialab, who wrote an MIT report book of this name in 1995, later published as a book (Picard, 1997).

[70] Other writers in the 19th century had also considered this issue, see Sir Charles Bell (1824) *The Anatomy and Philosophy of* Expression. Henry James (1884) suggested that some emotions have "distinct bodily expression" and thus opened up the possibility of physiological measurement (James, 1884, "What is an emotion?" Mind, 9, 188-205.

[71] Available at https://www.paulekman.com/product-category/facs/

[72] In some cases, these cues to what another human is feeling might be more readable by machine observer than a human observer. See Ekman and Friesen (1969); Ekman (2009).

[73] A brief literature search revealed affective computing researchers were building their affective models from psychology theorists, including: Millenson's 3 dimensional scale of emotional intensity (1967), in the lineage of Watson (1930) with a behaviourist model; Scherer (2005) with the Geneva emotion wheel- valence and control- or the Circumplex model- valence and arousal wheel adapting Russell (1980) in the lineage of Wundt (1903); Watson et al (1999) with the PANA model- quite similar to the Circumplex; Ekman (1967); Damasio (1996); Plutchik (1980); and Pörn (1986).

[74] From speech acoustic properties such as pitch, intonation, loudness, or speaking rate and voice quality can be combined to estimate emotional content. Body gesture recognition focuses on finding body alignment patterns of movement that indicate some emotional state, perhaps agitation, enthusiasm, or being tired. By using a Facial Action Encoding System (FACS), developed by Ekman, muscle movements in the face can be categorized and learned by a computer with a camera and a visual recognition system. Slightly more intrusively, facial electromyography measures muscle movement in the face. Electro encephalography (EEG) measures electrical patterns in the cortex which can sometimes be correlated to emotions. Normally EEG is not thought to provide a good indication of the activity of the lower part of the brain which some theorists associate with certain types of affect. Galvanic skin response can be used to measure arousal. The changing electrical properties of the skin are thought to be closely associated with the sympathetic nervous system (which, for example, in situations of extreme arousal, is associated with the so-called 'fight or flight' response). Other physiological measures influenced by emotions include heart rate, breathing, blood pressure, perspiration, or muscle contractions.

[75] Between 2015 and 2020 the affective computing market is predicted to grow USD 9.35 Billion to USD 42.51 Billion, a Compound Annual Growth Rate of 35.4%. (from press release September 2015 http://www.researchandmarkets.com/research/bqk6nm/affective)

[76] Research and Markets http://www.researchandmarkets.com/research/bqk6nm/affective

[77] For example, http://www.fiercebigdata.com/story/speech-analytics-aim-improve-sales-real-time-fraud-detection/2015-01-14; See also Gong, et al. (2015).

[78] https://source.wustl.edu/2004/02/research-casts-doubt-on-voicestress-lie-detection-technology

[79] The quantified self refers to the self of a person whom aspects of their daily are recorded by incorporating technology into data acquisition, often physiological data.

80 According to Martin Ford, the cost of these robots is up to USD 5000 (Ford, 2015, p.162).

81 http://www.parorobots.com

82 http://www.ai.mit.edu/projects/humanoid-robotics-group/kismet/kismet.html

83 https://henrikscharfe.com/project/lima-peru

84 There have, of course, long been discussions about ethics for those who create robots, and the ethical principles that should govern robots themselves. Many begin from those articulated in science fiction novel popular with roboticists, *Runaround* (1948) by Isaac Asimov, which famously invokes ethical laws for robotics. A UK funded research project has developed these into further ethical principles for robotics work. These principles discourage roboticists from designing robots "solely or primarily to kill or harm humans", acknowledge that robots are not responsible agents and should thus be designed & operated "as far as practicable to comply with existing laws, & fundamental rights and freedoms, including privacy", and that as products robots should be designed to "assure safety and security" just like any other product from washing machine to automobile, and that as manufactured artefacts robots "machine nature should be transparent" rather than "designed in a deceptive way to exploit vulnerable users", and finally that "[t]he person with the legal responsibility for the robot should be attributed." While these suggestions are for roboticists by roboticists (with consultation with a broad sample of stakeholders and researchers) they also make potential future policy suggestions such as licensing of robots, or "if not licensing per se, then perhaps some means of saying who is the owner of the robot, how to get in touch with them, perhaps a license plate, be it company or individual person." All of these principles are online with commentary at https://www.epsrc.ac.uk/research/ourportfolio/themes/engineering/activities/principlesofrobotic s. During our Foresight lab work we arranged a presentation from this project to HBP roboticists and programmers. In this presentation it was clear that one important tool that we have with robots, despite their somewhat ambiguous nature, is that as products they are subject to laws like any other products, like health and safety, but also laws of contract and accountability of owners and users of robots to use them in accordance with the laws of their country.

85 Frey and Osborne, 2015, *Technology at Work: The Future of Innovation and Employment*, available online: http://www.oxfordmartin.ox.ac.uk/publications/view/1883

86 It is worth noting this is different from the effect of information technology during the 20th Century (primarily from the 1980s onward) when middle income jobs were lost. Automation lacked the dexterity and manipulation to substitute low wage labour (as it had down in the 19th Century) and it lacked problem-solving and coordination abilities used by high wage workers. See Autor and Dorn, 2013.

87 One absence in this literature is much research on how artificial intelligence will affect agriculture. For example, the substitution of alternative industrial processed proteins for egg proteins to make an alternative mayonnaise - by the food innovator company Hampton Creek – shows the potential for a big data approach to food content to have a disruptive impact in agriculture: http://www.cnbc.com/2015/05/12/hampton-creek-disruptor-50.html This would have the potential to be particularly impactful when combined with new protein generating methods from synthetic biology.

88 For recent examples see Manyika, et. Al. (2013) available at http://www.mckinsey.com/business-functions/business-technology/our-insights/disruptive-technologies

89 Significant societal impact is often considered to be part of the term's definition.

90 Even past economy optimists such as Laurence Summers, former U.S. treasury secretary, have become sceptical that technological progress will directly result in job growth and well-being (Colvin, 2015, pp. 12-13).

91 Much debate has dwelt on whether or not such machines will be allowed to become 'autonomous' or whether or not a human "in the loop" will make the final decision. Perhaps a more subtle concern is whether or not improved remote killing capability is already encouraging greater levels of military adventurism, whether it corrupts the moral logic of warfare. That is to say, whether the possibility of military action without risk of casualties encourages decision makers to solve problems

militarily rather than diplomatically, whether it is the military equivalent of moral hazard in bank lending.

[92] Current military drones capable of armed action are semi-autonomous. Targeting and weapon release are performed by human pilots. The International Committee for Robot Arms Control and other initiatives are presently pushing for international treaties to prevent advances in drone warfare. See www.icrac.net

[93] For DARPA, see http://www.darpa.mil/program/our-research/darpa-and-the-brain-initiative, and for some discussion of these issues, see, for example, https://www.technologyreview.com/s/527561/military-funds-brain-computer-interfaces-to-control-feelings/

[94] These issues were discussed by contributors to our webinars. See also Bartolucci and Dando(2013).

[95] From the mid-1990s, conflicts in the Democratic Republic of Congo were fuelled by money made from trading the rare metals that are crucial for ICT equipment, and local militias fought over who would control these commodities, that that were mined in appalling conditions: http://www.techrepublic.com/article/how-conflict-minerals-funded-a-war-that-killed-millions/ . Today China controls much of the world supply of these rare metals, leading to concerns about its potential to block or limit their export: http://www.globalsecurity.org/military/world/china/rare-earth.htm. For a good summary of the materials used and their sources, see http://energyskeptic.com/2014/high-tech-cannot-last-rare-earth-metals/

[96] See http://www.techradar.com/news/computing-components/the-pc-rare-metals-crisis-921363

[97] http://www.theguardian.com/environment/2015/apr/19/worlds-mountain-of-electrical-waste-reaches-new-peak-of-42m-tonnes

[98] Taken from http://ec.europa.eu/eurostat/web/waste/key-waste-streams/weee. European legislation on EU legislation promoting the collection and recycling of such equipment (Directive 2002/96/EC on WEEE) has been in force since February 2003. The recast Directive (2012/19/EU), which entered into force on 13th of August 2012, introduces stepwise higher collection targets that will apply from 2016 and 2019.

[99] Lucy McAllister "The human and environmental effects of e-waste" available at http://www.prb.org/Publications/Articles/2013/e-waste.aspx   For more on dumping http://www.bbc.co.uk/panorama/hi/front_page/newsid_9483000/9483148.stm

[100] ISO 14040 (2006): Environmental management – Life cycle assessment – Principles and framework, International Organisation for Standardisation (ISO), Geneve

[101] See http://www.demos.org/publication/tackling-high-tech-trash-e-waste-explosion-what-we-can-do

[102] For an overview of this approach see section 1.3 of Rose, Aicardi, and Reinsborough (2015) available at http://www.kcl.ac.uk/sspp/departments/sshm/research/Research-Groups/Biomedicine-Ethics-and-Social-Justice/BESJ-Projects/HBP/Report-on-future-medicineWebsite-version.pdf

7. References

Asimov, I. (1942). 'Runaround'. *Astounding Science Fiction, Vol. 29, 1 (*March).

Autor, D., Dorn, D. (2013). 'The Growth of Low-skill Service Jobs and the Polarization of the U.S. Labor Market.' *American Economic Review*, 103(5): 1553-1597.

Ayesh, A., Blewit, W. (2015). 'Models for Computational Emotions from Psychological Theories Using Type I Fuzzy Logic' *Cognitive Computation*, June, Volume 7, Issue 3, pp. 285-308.

Bakhtiyari, K., Taghavi, M., Husain, H. (2015). 'Hybrid affective computing- keyboard, mouse and touch screen: from review to experiment', *Neural Computing and Applications* 26:1277-1296.

Barben, D., Fisher, E., Selin, C., Guston, D. (2008). 'Anticipatory Governance of Nanotechnology: Foresight, Engagement, and Integration.' pp. 979-1000 in Hackett, Amsterdamska, Lynch, and Wajcman, eds.*, The New Handbook of Science and Technology Studies*. MIT Press.

Bartolucci, V., Dando, M. (2013). 'What does neuroethics have to say about the problem of dual use.' Chapter 3 in Rappert, B., Selgelid, M. *On the Dual Uses of Science and Ethics: Principles, Practices and Prospects*, Australian National University Press.

Bell, C. (1824). *The Anatomy and Philosophy of Expression*.

Bishop, J. (2003). 'The Internet for Educating Individuals with Social Impairments.' *Journal of Computer Assisted Learning*. 19: 546-556.

Blanke, O. & Aspell, J. (2009). 'Brain technologies raise unprecedented ethical challenges'. *Nature*, vol. 458, p. 703.

Brynjolfsson, E., McAfee, A. (2015). 'The Great Decoupling: an interview with the authors of The Second Machine Age.' *Harvard Business Review*, June.

Calvo, R. A. & Peters, D. (2013). 'Promoting Psychological Wellbeing: Loftier Goals for New Technologies.' *IEEE Technology and Society*, December vol 32, 4. pp 19-21.

Calvo, R. A. & Peters, D. (2014). *Positive Computing: Technology for Wellbeing and Human Potential*. The MIT Press.

Chawla, DS. (2016). 'The Unsung Heroes of Scientific Software', *Nature*, Vol. 529, 7 January 2016, pp. 115-116.

Christensen, C. (1997). *The innovator's dilemma: when new technologies cause great firms to fail*, Boston, Massachusetts, USA: Harvard Business School Press.

Clausen, J. (2009). 'Man, machine and in between'. *Nature*, vol. 457, pp. 1080-1081. Colvin,

G. (2015). *Humans are Underrated: What high achievers know that brilliant machines never will*. Finland: Nicolas Brealey Publishing.

Damasio. A. (1996). 'The Somatic Marker Hypothesis and Possible Functions of the Prefrontal Cortex.' *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 351, 1413-1420.

Darwin, C. (1872). *The Expression of the Emotions in Man and Animals*.

Dwarakanath T., Wender, B., Seager, T., Fraser, M. (2013). 'Towards Anticipatory Life Cycle Assessment of Photovoltaics' published in *Photovoltaic Specialists Conference (PVSC)*, IEEE 39th.

Ekman, P. 'Lie Catching in Microexpressions' in Martin, ed. (2009) *The Philosophy of Deception*, Oxford University Press, Oxford. pp. 118-133.

Ekman, P. & Friesen, W. (1967). 'Head and Body Cues in the Judgement of Emotion: A Reformulation.' *Perceptual and Motor Skills*, 24, 711-724.

Ekman, P. and Friesen, W. (1969) 'Nonverbal Leakage and Clues to Deception.' *Psychiatry* 32, 88-97.

Fisher, E., Mahajan, R., Mitcham, C. (2006). 'Midstream Modulation of Technology: Governance from Within,' *Bulletin of Science Technology Society* December 2006   vol. 26, no. 6, pp. 485-496.

Fogg, B. J. (2002). *Persuasive Technology: Using Computers to Change What We Think and Do*. Morgan Kaufmann.

Ford, M. (2015). *The Rise of the Robots: Technology and the Threat of Mass Unemployment*. London: Oneworld Publications.

Frey, C., Osborne, M. (2015). 'Technology at Work: The Future of Innovation and Employment.' Oxford: Oxford Martin School.

Frick, W. (2015). 'When Your Boss Wears Metal Pants: Insights from the frontier of human-robot research'. *Harvard Business Review,* June.

Geller, T. (2015). 'How do you feel? Your computer knows'. *Communications of the ACM* Jan Vol 57 No 1 p24-26.

Gokcay, D., Yildirim, G., ed. (2011). *Affective Computing and Interaction: Psychological, Cognitive and Neuroscientific Perspectives*. IGI Global: Hershey, Pennsylvania.

Gong, S., Dai, Y., Ji, J., Wang, J., Sun, H. (2015). 'Emotion Analysis of Telephone Complaints from Customer Based on Affective Computing'. *Computational Intelligence and Neuroscience*, Volume 2015, Hindawi Publishing Corporation, Article ID 506905, 9 pages. http://dx.doi.org/10.1155/2015/506905

Grudin, J. (2009). 'AI and HCI: Two Fields Divided by a Common Focus', *AI Magazine*, Winter 2009, pp. 48-57.

James, H. (1884), 'What is an emotion?' Mind, 9, 188-205.

Kaliouby, R. Picard, R. Baron-Cohen, S. (2006) 'Affective Computing and Autism.' *Annals New York Academy of Sciences*. 1093: 228-248.

Kaptein, M., Markopoulos, P., de Ruyter, B., Aarts, E. (2010). 'Persuasion in ambient intelligence,' *Journal of Ambient Intelligence and Humanized Computing* March, Volume 1, Issue 1, pp 43-56.

Leisinger, K. (2015). 'Business needs to embrace sustainability targets', *Nature*, *528* (7581), 165.

Licklider, JCR. (1960). 'Man-Computer Symbiosis', *IRE Transactions on Human factors in Electronics*, March 1960, pp. 4-11.

Maass, W. (1997). 'Networks of Spiking Neurons: the Third Generation of Neural Network Models', *Neural Networks*, Vol. 10, No. 9, pp. 1659-1671.

MacDorman, K., Minato, T., Shimada, M., Itakura, S., Cowley, S. J., & Ishiguro, H. (2005). 'Assessing human likeness by eye contact in an android testbed.' *Proceedings of the XXVII Annual Meeting of the Cognitive Science Society* (pp. 1373-1378). July 21–23, 2005. Stresa, Italy. (6 pages).

Manyika, J., Chui, M., Bughin, J., Dobbs, R., Bisson, P., Marrs, A. (2013) 'Disruptive technologies: Advances that will transform life, business, and the global economy.' MCKinsey Global Institute.

Markoff, J. (2015). *Machines of Loving Grace. The quest for common grounds between humans and robots*. New York: HarperCollins.

Millenson J. (1967). *The psychology of emotion: theories of emotion perspective*. 4th ed. Hoboken Wiley.

Mindell, DA. (2015). *Our Robots, Ourselves. Robotics and the Myths of Autonomy*. New York: Viking.

Mol, A. (2008). *The Logic of Care: Health and the Problem of Patient Choice*. Routledge.

Owen, R., McNaughton, P., Stilgoe, J. (2012). 'Responsible research and innovation: From science in society to science for society, with society'. *Science and Public Policy,* 39, pp.751-780.

Pawlowski J., Eimler S., Jansen M., Stoffregen J., Geisler S., Koch O., Muller G., Handmann, U. (2015). 'Positive Computing: A New Trend in Business and Information Systems Engineering?' *Business Information Systems Engineering*. 57(6): 405–408.

Pörn L. (1986). 'On the nature of emotions.' *Changing Positions, Philosophical Studies*, vol. 38, pp. 205-214.

Picard, R. (2003). 'Affective computing: challenges,' *Int. J. Human-Computer Studies*. 59:55-64.

Picard, R. (1997). *Affective Computing*. Cambridge: MIT Press.

Pitt, J. (2012). 'Design Contractualism for Pervasive/Affective Computing'. *IEEE Technology and Society Magazine*, Winter.

Plutchik, R. (1980). *Emotion: A Psychoevolutionary Synthesis*, New York, NY: Harper & Row.

Ruger, T., Kim, P., Martin, A., Quinn, K. (2004). 'The Supreme Court Forecasting Project: Legal and Political Science Approaches to Predicting Supreme Court Decision-making.' *Columbia Law Review*, 104: 1150-1210.

Russell, J. (1980). 'A circumplex model of Affect.' *J. Pers. Soc. Psychol*. 39:1161-78.

Schalk, G. (2008). 'Brain-computer symbiosis', *Journal of Neural Engineering*, vol. 5, pp 1-15.

Scherer K. (2005) 'What are Emotions and how can they be measured' *Soc. Sci. Inf*. 44(4): 469-729.

Schwark, J. (2015) 'Toward a Taxonomy of Affective Computing', *International Journal of Human-Computer Interaction*, 31:11, 761-768.

Sharkey, N. (2008). 'The Ethical Frontiers of Robotics.' *Science*, 322(5909): 1800-1801.

Sharkey, A. Sharkey, N. (2008). 'Granny and the Robots: Ethical Issues in Robot Care for the Elderly.' *Ethics and Information Technology*, 14(1): 27-40.

Sharkey, N. and Sharkey, A. (2010) 'The crying shame of robot nannies: an ethical appraisal', *Journal of Interaction Studies*, 11, 161-190.

Srnicek, N., Williams, A. (2015). *Inventing the Future: Postcapitalism and a World Without Work*, Verso.

Stahl, B. (2013). 'Virtual Suicide and other ethical issues of emerging information technologies.' *Futures,* 50, pp. 35-43.

Stemmler, G. (2003). 'Introduction: Autonomic psychophysiology' in Davidson, Scherer, Goldsmith, eds., *Handbook of Affective Sciences*. New York, NY: Oxford University Press. pp. 131-134.

Stilgoe, J., Owen, R. & Macnaghten, P. (2013). 'Developing a framework of responsible innovation'. *Research Policy*. 42:1568–1580.

Thaler, R., Sunstein, C. (2008). *Nudge: Improving Decisions about Health, Wealth, and Happiness*. Yale University Press.

Walter, F., Röhrbein, F. & Knoll, A. (2015a). 'Neuromorphic implementations of neurobiological learning algorithms for spiking neural networks', *Neural Networks*, Vol. 72, pp. 152-167. doi:10.1016/j.neunet.2015.07.004.

Walter, F., Röhrbein, F. & Knoll, A. (2015b, in press). 'Computation by Time', *Neural Processing Letters*. doi: 10.1007/s11063-015-9478-6.

Watson, D., Wiese, D., Vaidya, J., Tellegen, A. (1999). 'The Two General Activation Systems of Affect: Structural Findings, Evolutionary Considerations, and Psychobiological Evidence.' *Journal of Personality and Social Psychology*, 76(5), 820-838.

Watson, J. (1930). *Behaviourism,* Chicago: Chicago University Press.

Wilkinson, R., Pickett, K. (2009). *The Spirit Level: Why More Equal Societies Almost Always Do Better*. London, Allen Lane.

Winograd, T. (2006). 'Shifting viewpoints: Artificial intelligence and human-computer interaction', *Artificial Intelligence*, vol. 170, pp. 1256-1258.

Wundt. W. (1903). *Naturwissenschaft Und Psychologie*.

Wynne, B. (2011). 'Lab Work Goes Social, and Vice Versa: Strategising Public Engagement Processes', *Science and Engineering Ethics*, 17(4): 791-800.

Youtie, J. (2010). 'Silos or systems in emerging science domains? The use of environmental, health and safety knowledge by nanotechnology researchers.' Darmstadt Conference Society for the Study of Nanoscience and Emerging Technologies, October 2010.