SCOPE

The conference “Moral Machines? Ethics and Politics of the Digital World”, in Helsinki, 6th-8th March 2019, is an international and interdisciplinary conference organized by two research fellows, Susanna Lindberg and Hanna-Riikka Roine at the Helsinki Collegium for Advanced Studies. The conference brings together researchers from all fields addressing the ethical and political issues of the digitalizing world. The conference approaches these issues from three perspectives: thinking in the digital world, the morality of machines, and the ways of controlling the digital world.

#moralmachines2019
blogs.helsinki.fi/moralmachines

Organising Committee
Susanna Lindberg, Helsinki Collegium for Advanced Studies
Hanna-Riikka Roine, Helsinki Collegium for Advanced Studies

Organisers and Partners in Cooperation
Helsinki Collegium for Advanced Studies
The Federation of Finnish Learned Societies
Institut Français de Finlande
The project Creative Adaptation to Wicked Socio-environmental Disruptions (WISE; Strategic Research Council at the Academy of Finland in Tampere University)

Design by
Minerva Juolahti
ABSTRACTS

MORAL MACHINES?

ETHICS AND POLITICS
OF THE DIGITAL WORLD

6–8 MARCH 2019
HELSINKI
Session 1: Moral Machines in Literature and Media

The Logic of Selection — Values, Opinions, Ethics and Politics on Contemporary Digital Interfaces

Matti Kangaskoski
University of Helsinki, Finland

In this talk I wish to examine the logic of cultural digital interfaces and how this logic influences its users’ values and opinions. This logic is – as I call it – the logic of selection. In short, this interface logic allows the user to select their desired content (e.g. news, literature, music, movies & television) from a vast array of possibilities at their own convenience. The offered repertoire is seemingly neutral and transparent, and the selection made by the user poses as free. But, of course, all content on digital interfaces is filtered and our access to it is structured. Instead of free choice, we encounter carefully curated content based on popularity, profiling, and assumed user affect.

To identify the logic I examine commercial (e.g. Netflix, Amazon) and non-commercial (Yle, Helmet) interfaces, social media (Facebook, Twitter), and mobile news applications (The Guardian, Helsingin Sanomat), and I close read their privacy policies, terms and conditions.

To examine how this logic influences values and opinions, I discuss three interlinked key effects: 1. Amplification of the “Superstar Effect” (sensu Rosen 1981). Due to the logic of selection, audience attention orients towards an increasingly small amount of ideas and products. 2. Quick affective response, which is propelled by shortened attention spans, hyper reading habits (Hayles 2007; 2012), and the encouragement of relating to media content by selecting emotional reactions (e.g. likes/dislikes, stars, emojis). 3. Profiling Feedback Loop. Products and ideas are created to cater to users’ quick affective selections.

When the interface logic poses as neutral, its effects seem natural. Thereby, high amounts of attention, for instance, can begin to pose in the form of importance (politics, culture), relevance (news), functionality (services), or even truth (science, journalism, public opinion). Through the investigation of interface logic, the talk opens up to a broader conversation on the ethics and politics of the digital world.

Imagining Machine Morals: Predictions on machine morality in science fiction

Teemu Korpijärvi
University of Greifswald, Germany & University of Eastern Finland

Science fiction stories have throughout their history “predicted”, or perhaps more precisely, given ideas to technologies that have later been studied, prototyped and implemented in real life. Examples vary from flight and cloning, to space-travel, to cell-phones, to Skype-video calls, to the development of ever-finer and more independent computers and programs run by them. Thus, dreams of fiction have gradually become reality, from the perspective of historiography; we live in the future of many of those stories of past. The future of many of those stories is now. Some of the changes science fiction has foreshadowed are the rise of computers, information networks, autonomous devices, and the research on Artificial Intelligence. My proposed presentation concentrates on the fictional portrayals of Artificial Intelligence, which in science fiction is often – but not always – described as either being, or becoming, malevolent or otherwise dangerous. After giving a brief overlook on these different kinds of portrayals, I will concentrate on the positive
representations of the A.I. in fiction, and what they may have to say for the future from a moral perspective. I have selected some examples from a set of representations of fictional “machine morals” and use of various degrees of independent machines, which I find representative, but far from exhaustive. As the amount of individual examples is limited only by the human imagination, experience upon which it builds, and the publishing field, this is not intended as a conclusive list or categorization, but a showcase of different approaches and a basis on which to build further discussion.

It can be argued, technology is only as “moral” as it’s designers, and in less independent cases, users. The designers till the “moral ground soil” technology grows from. Even if machines can later develop on their own, the base moral or ethical design gives them a framework within which growth can be achieved. Moral and ethics are subject to debate. Even though many of us share the same basic concepts, there is a lot of variation even within a single, relatively homogenous group, what is moral or ethical. Different moral codes in human society can be observed for instance through the Moral foundations theory proposed by Jonathan Haidt and Jesse Graham.

Thus it may well be more important to ask not, “Moral Machines, or Doom’s Day Devices?”, but rather “Machines with whose morals, and whose ethics, and to what ends?” and “Who imagines them, and through what media?”

**Women Fighting Back: Resisting Dystopia in Battledream Chronicle's Techno-Colonial Universe**

**Maude Riverin**  
University of Quebec in Montreal, Canada

In 1995, Nicholas Negroponte wrote that we were all becoming digital; and we can certainly asked ourselves what are the repercussions of this becoming-digital on the human body. Twenty years after Negroponte’s affirmation, I believe the question should rather be: is there still a human body? Technologies have impacted the definition(s) we at one point attributed to the body; not only has this becoming-digital affected the body in itself, but it also affected the existence of the body. In a context that aims to de-westernize Negroponte’s thoughts, how does this becoming-digital impact contemporary postcolonial understandings? And perhaps more interesting, how are contemporary postcolonial thoughts impacted by technologies? Using as a case study Kenyan filmmaker Alain Bidard’s film Battledream Chronicle (2015), my aim with this paper is to propose an analysis of the ways in which female bodies, as posthuman interfaces, reflects on this relationship between postcoloniality and technologies. Always at the limits between biological flesh and technological body, the film’s female characters represent the links that can be drawn between the claims of contemporary postcolonial theories and the ways in which posthumanism expresses the impacts of technologies.

It is in this context that science fiction becomes an interesting object of study: how does science fiction enable the mise-en-scene of those dystopian postcolonial narratives, and does it, at the same time, could enable utopian openings on African and Afro-diasporic futures? My aim with this paper is therefore to consider the technologized female bodies in Battledream Chronicle as figures of resistance within the spaces of postcolonial affect: in a context of dystopia, does the female main character in Alain Bidard’s film allow utopia?
Just Machines

Susanna Lindberg
Helsinki Collegium for Advanced Studies, University of Helsinki, Finland

Can a machine be just? Would a juridical machine be just, because it calculates its decisions mechanically, impartially, and equitably, and is therefore free of biases? Or would it on the contrary be just a machine that is incapable of understanding the case, therefore of knowing when to show mercy, severity, or pardon, and therefore of really assuming the responsibility that should go with every juridical decision? In Force de loi, Jacques Derrida addressed these questions to the judge. But today we have machines that are gradually assuming the role of dispensing justice, and they oblige us to rethink the relation between justice and machines.

It is tempting to regard a just machine as a developed version of a moral machine. The most famous “real” moral machine is the MIT online survey on the moral reactions caused by different types of accidents that self-driving cars might provoke. It is easy to show that this test does not really prove the moral value of anything but that at most it calculates the emotional cost of different accidents that self-driving cars might have. What the thought experiment really shows is the hollowness of the trolley problem that it is based on.

Today, our societies are gradually adapting much more complicated algorithmic systems that have the function of “just machines”. I use as concrete examples certain recruitment algorithms and the Parcoursup algorithm that operates admission to higher education in France since 2018. Both are expected to increase fairness. However, both have also been severely criticized for 1. Using weak programs 2. Using faulty and irrelevant data 3. Being non-transparent and inflexible. In worst cases, algorithms that were meant to abolish biases actually enforced them.

In my paper, I will show the philosophical grounds for thinking that these flaws do not depend (only) on bad conception, but on the fact that just machines are inevitably also unjust machines – because they are just machines. Firstly, I will remind that these sorting algorithms realize justice in the sense of Plato’s Republic, where justice is essentially finding the right place for everybody in the society, and show why the algorithms necessarily fall short of the task of platonic educators. Secondly, I will analyze the ambiguous existential consequences of being sorted by inexorable machines. Thirdly, I will show why the algorithms themselves, being technological constructs, are inevitably both technological and democratic black boxes with a detrimental effect on justice, if the responsibility of doing the just choice is really discharged on them.

Machine Vision: Selfie Filters and Image Recognition Algorithms as Cognitive Technical Systems

Jill Walker Rettberg
University of Bergen, Norway

Machine vision has become commonplace. Smartphones have advanced image manipulation capabilities, social media use image recognition algorithms to sort and filter visual content, and games, narratives and art increasingly represent and use machine vision techniques such as facial recognition algorithms, eye-tracking and virtual reality.
This paper asks how our intimate and almost continuous use of machine vision in our everyday lives affects our sense of agency and interaction with these technologies. For instance, it’s fun playing with selfie filters, but they’re also a technology that normalises biometrics, a form of machine vision that is otherwise used to identify and control individuals (Rettberg 2017). What does it mean that we are making such intimate use of visual technologies developed for surveillance, war and industry? Contemporary machine vision can certainly be seen as a cognitive technical system, and humans enter into human-technical cognitive assemblages (following Hayles 2017, 3) with selfie filters, image recognition algorithms and other machine vision technologies. How can these contemporary theories of the relationships between humans and machine vision technologies help us understand the specificities of the everyday use of a selfie filter or the automated “memories” generated by facial and image recognition algorithms in the iPhone’s Photo app or Google Photos? In addition to using posthumanist theories I will build upon earlier theories of photography emphasise the technological apparatus in relation to today’s machine vision, such as Vilhelm Flusser’s argument that photographers’ agency remains “subordinate to the camera’s program” (Flusser 2000, 36) hold up today, and Dziga Vertov’s early portrayals of the camera as a subject in its own right (“I am kino-eye, I am mechanical eye, I, a machine, show you the world as only I can see it…” (Vertov 1984 orig. 1923).

Tagshot

Mika Elo
University of Arts Helsinki, Finland

In his essay ‘Little History of Photography’, Walter Benjamin takes up Lazlo Moholy-Nagy’s claim that the illiterate of the future will be those who cannot read photographs. For Benjamin the key question is whether inscription, or more literally “scripting” (Beschriftung) will become the most important part of the photograph. Today, in the era of digital connectivity, we need to ask a similar question concerning the role of algorithmic processes in photography. How do photographs we take, share and look at relate to invisible metadata and computational processes? How is the “script” of photographic phenomena now constituted? Benjamin helps us to explicate how metadata structures incorporated in the photographic apparatus introduce new discursive structures in society. This paper aims at outlining a structural shift in photography culture fostered by the variegated forms of tagging prevalent in digital networks: a shift from snapshot culture to tagshot culture. The photographic apparatus is hereby characterized as an assemblage facilitating intersemiotic entanglements in processes that convey sense. The allover aim of the paper is to highlight, from a media aesthetic point of view, some key aspects of the material-discursive conditions of thinking in the tagshot culture.
**Session 3: Machines as Artists**

**Creativity in Computer-Generated Narratives**

*Mariana Chinellato Ferreira*

University of Coimbra, Portugal

One of the major questions when discussing Computational Creativity is whether human creativity can be simulated (Boden 2009; Chamberlain et al. 2017; Hertzmann 2018). This question can be divided into two aspects: the philosophical and the technical implication of such works. The field of Artificial Intelligence has been experimenting with creativity in a wide context, such as stories (Gervás 2006; Kibartas & Bidarra 2016), poems (Manurung 2000; Gonçalo Oliveira 2012) music (De Mantaras 2006), painting (Zhang et al. 2012) and any other kind of artwork that can be replicated. Over the last few decades, the partnership between Narratology and Artificial Intelligence has resulted in a noticeable amount of systems that generate literary narratives automatically. On one hand, most of the systems present well-developed and complex techniques in Story Generation, on the other hand, compared to the complexity of their architecture, their output might be considered too simple.

This way, this paper intends to observe the creative process of the systems Minstrel (Turner 1992) and MEXICA (Pérez y Pérez, 1999; 2001) and perceive the complexity of their architectures and the output they generate. The focus of this analysis will be the comparison of the human creative process and the computational creative process and the probable influences of these processes in the final output. This way, we can perceive how creativity is emulated in those systems. Some studies (Power et al. 2003; Mckeown & Jordanous 2018) point out that systems that present a higher level of constraints and human input generate more aesthetically pleasing output. These studies might show us that even though the development of AI systems has reached higher levels, the simulation of creativity is still a challenge for the computer scientist, especially when related to the automatic generation of literary texts.

**A listening machine: Sensory agency and the digitalization of the sonic environment**

*Meri Kytö*

Tampere University, Finland

The relationship to the sensory environment is becoming more and more technologically mediated. The everyday heard sonic environment (soundscape) is often mediated for us or by us, by infrastructures sustained by pervasive sensor technology, and for many the sonic environment is enhanced by a speech processor located in their hearing device.

This paper examines post-phenomenological epistemologies of embodied human–technology-relationships (see Hayles 1999) by asking what is meant by listening, learning to listen, listening subjectivities and level of attention in situations where the environment is technologically mediated or the listener digitalized (a machine)?

These questions are approached with two empirical cases: 1) an ethnography with one adult informant learning to listen with a cochlear implant, from deaf to hearing, and 2) studying the developing of machine learning and sensor technology software monitoring sonic environments, focusing especially on the praxis of annotation. Understanding the interfaces between the audio signal and self, the environment and language brings forward the challenges of regularizing (see MacKenzie 2017) sensory phenomena.
The concept of ubiquitous listening (Kassabian 2013) takes as a model for its subjectivity the pervasive structures of ubiquitous computing, information technology that merges into the everyday environment. These kinds of coded spaces shape our understanding of acoustemology, knowing place by listening. People living with cochlear implants hear their environment through microphone and code. Machine learning is developed for the needs of sensor technology by designing artificial intelligence software capable of recognizing sound sources and events from a stream of audio. The notion of ability designed for technology will also be discussed in the frame of the logic of accumulation in surveillance capitalism (Zuboff 2015).

**Automatic and creative: playing with algorithms and fragments in the Book of Disquiet Archive**

Cecilia Magalhaes  
University of Coimbra, Portugal

The Book of Disquiet Archive, or LdoD Archive, (2017, Portela & Silva, ldod.uc.pt) is a digital literary simulator, based on the unfinished and fragmentary Book of Disquiet, written by one of the most important Portuguese modernist authors, Fernando Pessoa. This digital archive has been designed as a dynamic, multilayered system, presenting in its informational structure the Pessoa’s digitised facsimiles and the transcriptions of the four main critical editions of the Book of Disquiet (Prado Coelho, 1982; Sobral Cunha, 2008; Zenith, 2012; Pizarro, 2010). The traces of these authorial and editorial practices were codified, indexed and interconnected in a combined metadata schema, which makes LdoD Archive an environment for play. In this sense, the platform is a literary machine which instantiates practices of reading, editing and writing by managing the fragmentary textual units of the Book of Disquiet in its different variants. Further, the platform permits to design new virtual editions, creatively composed according to the users’ understandings and intentions in interacting with the digital fragments. On the one hand, LdoD Archive is an experiential environment for exploratory literary practices. On the other hand, this same environment is programmed according to automatic procedures and constraints, which conduct the users’ paths in their own creative process. The virtual editions can be planned and designed by using the platform algorithms which recommend, for instance, new reading and editing sequence of fragments. Further, these same algorithms, according to the users’ editing intentions, can track, suggest and define new textual patterns based on semantic values, authorial writing dates, type of text, editions interpretations, heteronomy and so on. In this context, between users’ creative practice and programmed system, this paper aims to present how the algorithms usage works in the LdoD Archive and which kind of results can emerge from these literary activities on screen, here already been observed and analysed in our current focus group tests.

**Moving Failures: A Tentative Approach to Human-Robot Interaction on Stage**

Irene Alcubilla Troughton  
University of Utrecht, Netherlands

In this paper I propose a reading of two artistic practices – a dance technique created by the Latin-American choreographer Lucas Condró and the installations of Bill Vorn in Hysterical machines – as a way of engaging with the ongoing debate about the intersections between new media and human beings. Following the understandings of scholars such as Katherine Hayles, Mark Hansen or Karen Barad, I shall address the reformulation of the limits of cognition, empathy and agency. Furthermore, I shall contribute with political and social considerations about how to dismantle normalizing and ableist conditions in this field.
Bill Vorn started in 2006 the creation of robots in environmental installations in a project named Hysterical machines. These creations are equipped with sensorimotor systems, which allows them to perceive their surroundings and react to them. Their movement constitutes a type of cognition that does not include consciousness, hence operating outside the parameters of rational thinking. Lucas Condró developed his technique of “asymmetrical motion” by basing human movements in the principles of independency – singularity of each part of the body to move in a certain manner – and correlation – limitations of those parts, always in a position of relationality. From these contradictory principles emerge the asymmetries of the body as a base for movement which I propose to be a performance of the non-conscious.

I shall explain their diverse ways of creating a system of punctuated and distributed agency. However, instead of merely addressing their differences, I shall point at a creative connection that might open new paths in considering relations between human and non-human actors on stage, specifically within the field of Human-Robot Interaction. This connection resides in a type of movement that is connected to non-conscious cognition as conceptualised by Katherine Hayles and that, in an ethical consideration, I analyse through the lens of vulnerability (Judith Butler) and failure (Jack Halberstam). Such analysis not only rejects anthropomorphic creations but also delves into a different conceptualisation of the human and the machinic selves that move to resist, on the one hand, the reinforcement of the ideas of the Modernist autonomous subject and, on the other hand, a capitalised and utilitarian use of technology.

Session 4: Manipulation in Web Search and Social Media

Thinking of Google Search as Digital Colonialism

Joshua Adams
Salem State University, United States

I want to make a presentation on how digital tools like search engines can reinforce current and historical inequalities by providing new modes of “ownership” over the digital representation of marginalized groups. To illustrate this point, I will use the Google Search results for the term “Ubuntu.”

Ubuntu is a Zulu word referring to the South African spiritual philosophy of human connectedness and communal responsibility. It is often roughly translated as “I am because you are.” However, the top search result for “Ubuntu” leads you to an open source software operating system. Because Google is primarily a multi-national ad broker, the company Ubuntu has an economic incentive to direct the digital representation of the term “Ubuntu” through search engine optimization.

I want to show how thinking critically about the search results for “Ubuntu” reveals how free market beliefs and values are encoded into search engines, and how users are, often unknowingly, accepting the value systems of neoliberalism, and corporate and colonial individualism. Through search engine optimization, digital tools like Google’s search engine incentivize a kind of “colonial gaze” where prevailing ideas about the democratic potential of the internet blind users from the ways these tools privilege the values, beliefs, ideologies, and ontologies of the Western world. The paper will cite professor Safiya Noble’s “Algorithms of Oppression: How Search Reinforce Racism”, Ramesh Srinivasan’s “Whose Global Village?”, the work of Tarleton Gillespie, Sarah T. Roberts, and others.
Expectations on Automated Future in Social Media Analysis

Juho Pääkkönen
University of Helsinki & Aalto University, Finland

Salla-Maaria Laaksonen & Mikko Jauho
Centre for Consumer Society Research, University of Helsinki, Finland

Social media analytics is a burgeoning new field associated with high promises of societal relevance, business value, and accurate prediction and understanding of human behavior. To unpack the larger discourses of predictive technology and related forms of knowing which underlie these processes (Couldry & Yu 2018; Thrift 2005), studying the epistemic conceptions and expectations that drive data analytics is of paramount importance (Beer 2017). In this paper we build on the sociology of expectations (Beckert 2016; Brown & Michael 2010), and research on expertise in the interaction between humans and machines (Collins 1990; Collins & Kusch 1998) to analyze the role of automation in negotiating the future of social media analytics.

Empirically we draw on participatory observation conducted in four analytics companies and 11 thematic interviews with analysts and their clients to examine the expectations that guide how social media analytics is envisioned as a form of future-oriented knowledge production. Social media analytics appears as a field facing both hype and distrust—either towards the data, analysis methods, or the users of analytics—on part of both clients and the analysts (cf. Kennedy 2016). In this context, the idea of automated analysis through algorithmic methods emerges as a central notion of how these hopes could be realized. Variously by condensing messy and disparate information, or by alleviating the role of subjective expert judgment, automation is hoped to lend analytics with credibility, and make sense of otherwise incomprehensible masses of social media data.

We argue that conceptions of automated knowledge production have an important role in guiding how the future of analytics is envisioned. Thus, our analysis provides an important step towards understanding how knowledge production is negotiated in social media analytics—information that is crucial for assessing how analytics will shape our lives, as the societal role of algorithmic systems becomes increasingly pervasive.

Authorship vs. Assemblage in Computational Media

Hanna-Riikka Roine
Helsinki Collegium for Advanced Studies, Finland

Laura Piippo
University of Jyväskylä, Finland

Our paper outlines an approach to complement previous analyses on social media as storytelling as they have, in our view, neglected the platform interfaces and their effect on users. As digital platforms such as social media services are increasingly embedded in our everyday life and meaning-making, we need to find ways to critically analyse these phenomena. Our particular point of contention is the concept of authorship. It is suited for the analysis of tweets, updates and other content created by a singular user (e.g., small and shared stories in Georgakopoulou 2007, Page 2018, Mäkelä 2019). However, in the analysis of larger phenomena which supposedly emerge from the sharing of those smaller stories in social media, the concept falls short. Furthermore, we argue that the understanding of singular updates, tweets, comments and other shared content as “stories of personal experience” is not always accurate or beneficial. Instead, our aim is to take into the account not only the medium-specificity of computational media, but also the fact that in addition to collaborating and sharing the platforms with other people, we co-exist
and co-create with nonhuman actors such as bots and algorithms (Emerson 2014), whose key feature is to exceed human capacity.

Although digital platforms do exist as sophisticated, designed environments, what appears to an individual user as a single platform is in fact comprised of a multitude of different variations of the content created and shared on the platform. In the case of social media services such as Twitter, the interface (despite undergoing e.g., changes in design from time to time) remains the same, but the feed is assembled in many respects by algorithms and, therefore, appears different to every user (e.g. Bozdag 2013). “Twitter” is, consequently, an “imagined environment” (cf. Anderson 2006) upon which users reflect, react, write and share. Moreover, trolls, doxing, hate campaigns and other mass-produced affective content without anything automatically “personal” or “experienced” behind them pose a problem to the supposed experientiality in the content created and shared, along with the hypothetical event of “the Inversion” which highlights the fact that bots already outnumber humans online (see e.g., Read 2018).

As a result, we suggest that the concept of assemblage (cf. Hayles 2017) could be more beneficial for gaining an understanding of digital environments as platforms for storytelling than the concept of authorship. In our view, assemblage enables the analysis of the platforms as affective environments which are generated by algorithms. Environments such as these are based on a feedback loop of a kind: they are affected by our actions and, in turn, have an effect on how we act and think.

Schizoid Nondroids: Fictions of the Everted Cyberspace

Esko Suoranta
University of Helsinki, Finland

Since the advent of Web 2.0, cyberspace has begun to evert. It increasingly interpenetrates and affects meatspace. On the one hand, this means a more democratic dynamic in the development of digital content and providing outlets for diverse voices through ease of access. On the other, the increase in use and algorithmic surveillance of that use open up avenues for commodifying everything users leave behind online.

In How We Became Posthuman, N. Katherine Hayles evokes the figure of the schizoid android at the center of the complexities of cybernetic systems. She illustrates this figure through Philip K. Dick’s novels, where the schizoid android is characterized as unable “to feel empathy, incapable of understanding others as people like herself” (161). Despite and even because of this, the question of the androids’ possible humanity is central in Dick’s fiction.

Now schizoid androids are among us, even if they most often lack anthropomorphic bodies. We tend to call them machine learning algorithms, artificial intelligences, and social media platforms. For now, the question of their humanity is not pressing, but they hold power over many avenues of human life through their prevalence and efficiency in the everted cyberspace. Like Dick’s androids, they do not understand people as people. Further, they operate in the paradigm of surveillance capitalism where the data exhaust of digital behavior is the avenue to increasing profits.

In my paper, I attempt to update the figure of the schizoid android to correspond to the eversion of cyberspace and the logic of surveillance capitalism. My analysis will draw on works of contemporary speculative fiction by authors like Annalee Newitz, Malka Older, and Dave Eggers as well as theoretical work like Pramod K. Nayar’s concept of traumatic materialism and Shoshana Zuboff’s analysis of surveillance capitalism.
Session 5: Ethical and Political Problems with Machines

Automation, between factuality and normativity

Marc-Antoine Pencolé
Paris Nanterre University, France

Moral or ethical decisions, involving values and norms of a given community, necessarily have to be made, as it seems, by subjects, characterized in particular by their intentionality and autonomy. Thus, delegating them to an automated apparatus would consist either in eliminating their moral dimension, or at least limiting it to the morality of its original designer.

Such classical conception hinges on strictly non-dialectical notions of factuality and normativity. Our claim is that the delegation of a moral or ethical decision to an automaton – complex and powerful enough to be assimilated to a « decision-making » machine – may indeed be considered a dispossession yet also, under certain circumstances, be the very condition of its effectivity. Indeed, as exposed through the hegelian tradition and particularly thoroughly by Habermas, certain normative demands could require, to be fulfilled, the collective power that is being cast into an institutional or a technical system – an algorithm for instance. To be normatively satisfactory, concrete interactions imply a motivational, cognitive and organisational load that the subjects involved have to bear: without being mediated by factuality, such as that of informational machines, these normative demands would remain ineffective.

We will use the case of diverse and ambivalent peer-to-peer files exchange communities to back our claim and help shift the debate about the morality of « decision-making » machines toward the question of the different forms of delegation to automata: what are the intrinsic of contextual elements that make such delegation a successful effectuation of collective norms or a sheer dispossession of our autonomy?

Can machines think? Disambiguating the double talk

Arto Laitinen
Tampere University, Finland

This paper argues that the debate on whether machines think can be clarified by reconstructing it as a debate between two ideal-typical approaches, which use the key terms in different meanings. I call them the “common sense phenomenology” and “cybernetic approach” (cf. Jonas). They seem to disagree on whether machines really think, but arguably they mean a different thing by “really thinking”. As Alan Turing (1950) predicted, the key concepts have undergone a shift in meaning: “The original question, “Can machines think?” I believe to be too meaningless to deserve discussion. Nevertheless I believe that at the end of the century the use of words and general educated opinion will have altered so much that one will be able to speak of machines thinking without expecting to be contradicted.”

There are inchoate reactions to this conceptual expansion, which can however be reconstructed as two idealtypical approaches, common sense phenomenology and cybernetic approach, that correspond to what W. Sellars called the “manifest image” and the “scientific image”. For the former approach, there is a great qualitative difference between sentient, intentional responsible agents such as human persons and “mere” machines that can simulate humans, but for the latter, exemplified by cognitive science, both humans and robots can be seen as cognitive, cybernetic systems.
The confusion consists in the predicament, that seemingly any central concept that the former appeals to, in order to show the difference between humans and robots, such as “thinking” or “feeling” or “being responsible”, or being “self-conscious”, the latter can say that in a sense robots, too, think, feel and are responsible and self-conscious. The aim of this paper is to articulate these two families of meanings, and show that they agree that machines do not Think, or Feel, in some inflated sense, and that it may make sense to say they do “think” or “feel” in some deflated sense. The debate really turns on whether it is true and fruitful to apply the inflated sense to humans. The debate is not merely academic. How we grasp these concepts is constitutive of human social practices: the practice of holding each other responsible depends on how we understand conditions of responsibility. It has been argued that changes in the senses of the central concepts may lead to distorted self-understandings and social pathologies. An argument defending the common sense phenomenological viewpoint to machine thinking can appeal to this constitutive connection to social practices of holding agents responsible.

Applying Principalism to Military Ethics for Human Soldiers and Combat Robots

Marko Ahteensuu
Tampere University, Finland

Mireille Musangamfura, Laura Puumala & Helena Siipi
University of Turku, Finland

In last few decades, principlism has become a major ethical approach in medicine and it has also been applied to ethics of nursing, business and education. The basic idea of principlism is that practical ethical problems can be answered with the help of the following four prima facie principles: respecting autonomy, beneficence, non-maleficence and justice. Is principlism applicable to military ethics? For example, what could respecting autonomy imply in top-down managed military operations where the desirable outcomes are often sought by restricting individuals’ action and freedom? Similar questions arise with respect to the other three principles. It has been predicted that in the next few years, combat robots possessing degrees of autonomy will be employed in military operations. Some of these warbots are then used for killing people. To what extent principlist military ethics for human soldiers can and should be applied to combat robots? Are there different criteria for ethical “choices” for combat robots than for human beings? Several central precepts of human ethics, such as rights, weakness of will and forgiveness, do not seem to concern artificial intelligence (AI). Can we require higher moral standards for the robots than for humans as their choices are not “disturbed” by fear, remorse, cruelty, and other human feelings and emotions?

A Conversation between objects: The Ethical Underpinnings of the IoT

Michael Smith
Purdue University, United States

If the internet is seen as a platform to connect people, the Internet of Things (IoT) refers to the increasing interconnectivity between objects. Once an impossibility, the IoT is now possible with the presence of IPv6, which can accommodate exponentially more online addresses thus enabling the massive interactions and big data needed for the IoT to function. This presentation will provide several examples of the IoT, including its use on production lines and in pharmaceuticals. It will frame the IoT through systems theory, using Luhmann’s writing in particular. The presentation will then connect this systems definition with Latour’s Actor-Network theory (ANT), which states that everything in the social world exists in constantly shifting networks of relation. The ANT argues that objects are designed to shape human actions and decisions. As a result, they mediate human relations and impact our ethics. As a result, this presentation will question what an ethics of objects
might look like, particularly as objects are increasingly becoming actors in the formation and practice of ethics. Object-theory will be explored in relation to the IoT in addressing this question. Finally, this presentation will explore crime and unethical acts that could emerge from the IoT, which may be the new frontier for cyber-crime. Every time that a connection is established, a window is established for a hacker to invade the connected device and potentially use it for illegal activity. More so, given the economic importance of a smooth production process, the theft of information regarding a company’s process of production can be just as damaging as the theft of information regarding the product the process makes. Only in this case, the cyber-criminal may be part object, which only appears to “think” via connections. The presentation will conclude by identifying an ethics of objects in relation to the IoT.

Session 6: Digital Control of Life and Strategies of Resistance
1/2

Dis/assembling Self-knowledge and Temporalizing the Self in Self-tracking Practice

Harley Bergroth
University of Turku, Finland

A plethora of near-body digital self-tracking technologies and related mobile applications are now widely available especially in health- and wellness-related retail contexts (Schüll 2016; Lupton, 2016). These popular devices, such as activity tracking wristbands and sleep tracking devices, enable the real-time gathering of measurements on selected functions of the body, and the assembling of measurements into datasets. As such, self-tracking technologies are typically understood to assist individuals in the production of (health-related) self-knowledge and enable life management and self-control.

This presentation draws from ethnographic data and interview narratives on self-tracking and shows that while self-tracking is often made sense of through the notion of attaining self-knowledge and finding correlations between actions and their consequences, in practice such correlations – and a sense of stability and control in relation to the self – remain persistently elusive in people’s everyday experience. It then argues that self-tracking technologies actively produce their purpose and functionality as knowledge-enhancing technologies in the course of the ongoing technical co-operation, as lived self-knowledge is actively disassembled into temporally extensive trajectories through the affordances of self-tracking technologies. This complexity is made sense of in relation to Stiegler’s philosophy of cinema and technology (2011), and through the idea of the production of the self as a ‘temporal object’; as a flux that is perpetually unfolding.

The presentation contributes to understanding self-tracking as a sociotechnical domain of the production of (ontological) uncertainty (cf. Beer, 2016). The production of uncertainty and related tracking-oriented mindsets is further considered in relation to the alignment of everyday lifeworlds with the requirements of increasingly data-driven capitalism.
What we hide, what we reveal: Expression games with algorithmic systems

Jesse Haapoja
Aalto University, Finland

Airi Lampinen
Stockholm University, Sweden

We build on Erving Goffman’s work on strategic interaction and expression games (1970) to analyze how individuals may manipulate algorithmic systems. Goffman’s work on expression games discusses “the individual’s capacity to acquire, reveal, and conceal information”. He divided these games to five basic moves: unwitting, naive, control, uncovering and counter-uncovering. The first two refer to actions where the observed and observer act naturally. The latter three moves refer to situations where the observed and the observer are locked into a game like situation, attempting to outwit each other. Studying how users try to resist algorithmic systems is important as these systems are an increasingly important part of everyday practices (Willson, 2017), and more and more of our daily behavior is used as a data source. The collection of behavioral data has raised questions on how individuals are placed under surveillance (van Dijck, 2014) and manipulated (Yeung, 2017) by institutions – private companies and public organisations alike – who control the data. Even though human-algorithmic systems interaction differs from interactions individuals have with each other, commonalities can be found since like humans, these systems require information from their interaction partner. Information can be provided willingly and knowingly, or given off without intending to do so or even realizing that data collection is taking place. In addition, what data is gathered is defined by people building these systems. When an individual gains knowledge about the data systems gather from them, they can alter their behavior to play the system to their advantage. System designers often try to make gaming their systems more difficult. We will offer examples to show how the gaming of algorithmic systems can be approached with the help of Goffman’s formulation of moves in expression games.

Delegated moderation work: humans and machines in content moderation

Linda Turunen
University of Helsinki, Finland

Emergence of social media and increase in user generated content has shaped digital work; platforms invite users to post and freely share their thoughts, and people are increasingly responding to this call. People with all sorts of motivations and interests have found social media as a channel to share and post their personal experiences and opinions. Due to these developments, platforms face increasing pressures to govern communication (Hughey & Daniels, 2013), while, simultaneously, the conversation needs to be instantly public to maintain the ‘realtimeness’ of social media (Kaplan & Haenlein, 2010). In the intersection of these diverse pursuits and conditions lies an area of work -- content moderation -- that is becoming central for understanding the social and political tensions around current online cultures and work division. Content moderators are the hidden custodians of platforms (Gillespie, 2018), the unseen and silent guardians who bring safety to online spaces by overseeing the visual and textual content users have generated (Roberts, 2014). They also have an influence on the discussion dynamics and what is archived for readers (e.g., Sherrick & Hoewe, 2018).

As moderation work has turned more labor intensive, specific temporalities and strategies for moderation practices have been applied to enable the flow of discussion in the abundance of messages (Gillespie, 2018). Companies are exploring alternative options to meet the increasing need; for example, involving volunteers or implementing machine-learning-enabled moderation tools. This paper will shed light on the development of content moderation profession and the dynamics of work and labor mediated by digital technologies. More precisely, ethnographically-
oriented interview data is collected to explore the human – machine work division in text-based conversational spaces, as well as the advantages and challenges machines are perceived to bring to content moderation work.

'The Future of the Internet Hangs in the Balance': The Perception and Framing of Political Opportunity and Threat among Activists in Digital Space

Jared M. Wright
Purdue University, United States

How does digital space shape the dynamics of collective action? Specifically, how might the perception and framing of political opportunities and threats differ in the digital environment than in other, more traditional sites of inquiry? Opportunities are typically conceptualized as openings which reduce the cost of collective action, while threat is a force which increases the cost of inaction (Einwohner and Maher 2011; Gamson and Meyer 1996; Goldstone and Tilly 2001; McAdam 1982; Meyer and Minkoff 2004; Tarrow 1998; Tilly 1978; Van Dyke and Soule 2002). But if the affordances of Internet technology can provide new ways of reducing the costs of collective action (Earl and Kimport 2011), might then opportunity become less salient for mobilization, and threat more so, particularly for movements which are most adept at leveraging these affordances in pursuit of their goals? This study utilizes automated text mining and semantic network analysis combined with qualitative thematic coding to examine four years of archival texts from two highly-adept online activist groups, the Electronic Frontier Foundation and AnonNews (a subgroup of the Anonymous hacktivist collective). Analyses confirm that threat appears with much greater frequency than opportunity in the documents, particularly for the more radical Anons in which volatile opportunities are nonexistant. While earlier scholars have shown that the processes of contention are innately and necessarily spacial (Lefebvre 1974; Martin and Miller 2003; Tilly 2000), no one has specifically examined how digital space might reshape the dynamics of political opportunity and threat. This study shows the importance of spacial considerations of the digital realm on collective action by arguing that when technological affordances are most highly leveraged in online social movements, the cost of collective action can become so low that political opportunities lose salience as a motivating factor.

Session 7: Digital Control of Life and Strategies of Resistance 2/2

Towards an Ethics of Personal Data: A Question of Trust

Viivi Lähteenoja
University of Helsinki, Finland

Personal data is everywhere. We leave digital breadcrumbs behind us everywhere we go and whatever we do. Personal data about us as individuals, and especially about us collectively, has enormous potential for social change, as is evident from the Cambridge Analytica / Facebook scandal and its implications for democracy. Personal data also fuels artificial intelligence, which is becoming increasingly omnipresent in our lives and societies.

As society is becoming more and more aware of the potential of (and potential harm from) individual and collective personal data, calls for more ethical handling of personal data by businesses and organisations are becoming more prominent and legislation like the EU General Data Protection Regulation (GDPR) are establishing the legal boundaries for what can and cannot be legally done with personal data.

This paper will address in broad terms what we mean when we say “more ethical handling of personal data” focusing on a case study addressing a question of trust. The material for this discussion is provided by the proceedings of the MyData 2018 conference, a personal data professionals’ meeting held in Helsinki in August 2018, which included discussions focused on the business, legal, technological, and social science aspects of personal data.

The theoretical framework for this discussion is provided by virtue ethics. Virtue ethics has been successfully applied to analogous fields such as environmental ethics, business ethics, medical ethics, and law, and its application to data ethics, and especially the ethics of personal data, promises to yield interesting perspectives as well as to contribute to a shared framework and language to discuss and position emerging issues.


3 The proceedings was commissioned by the Finnish Ministry for Transport and Communications.


Controlling and contextualising digital health data: public perceptions on morality of health data use

Karoliina Snell
Helsinki Collegium for Advanced Studies, University of Helsinki, Finland

The Finnish government aims to create a health data economy that is the most competitive in the world. Great emphasis is put on large data collections and intensified data sourcing to generate economic growth, to enhance medical research, and to boost people’s health in new ways. New data driven practices create vast amount of data about individuals and enable data use across organisational and geographic boundaries. I examine public perception towards digital health data use through focus groups conducted in 2011-2017 in Finland. The analysis illuminates questions of ethics, morality, controllability and especially contextualisation related to digital health data.

Finns are positive towards medical research on health data, but they share a wish for contextualisation of data use – that data would be used only for the original (medical) purposes and that it should only be distributed to a restricted set of users. In contrast, Finnish policies aim at wide, cross-sector use of health data by public and private institutions, nationally and internationally. The data is more valuable if it can be made to “travel” far and wide and can be used extensively. Thus data becomes removed from its original context and translated to serve other purposes and it is analysed with the help of often non-transparent algorithms. I examine discrepancy between the moral contextualisations made by citizens and those promoted in policy. The theoretical reflections build on Nissenbaum’s concept of contextual integrity, which ties protection of privacy to norms of a specific context. I expand the use of the concept from the realm of privacy to control, acceptability, benefits and moral justifications.
Distributed autonomy: smart insurance as a technological imaginary

Maiju Tanninen & Turo-Kimmo Lehtonen
Tampere University, Finland

Minna Ruckenstein
HELDIG, Consumer Society Research Centre, University of Helsinki, Finland

In this paper, we study the so-called “smart insurance” that combines self-tracking technologies with life insurance. Our main question is how smart insurance enacts futures in relation to insurance practices and relationships. Furthermore, we study how the actual and potential consumers of smart insurance products relate to these new technologies and how they perform and negotiate their degrees of autonomy in self-tracking practices. Through the examination of these practices and both the insurance companies’ and the consumers’ imaginaries related to this technology, we develop the concept of distributed autonomy. The paper is based on interviews with active and potential smart insurance customers and product designers of a Finnish insurance company, conducted in 2017–2018.

Smart insurance products aim to track and manipulate the customers’ behaviour by utilizing activity wristbands and other sensory devices. The policyholders are encouraged to share their acts of self-determination in health decisions with the devices while the insurance company gathers data on their activities. The policyholders, for their part, receive free data services; also, the self-tracking devices are meant to aid them in their efforts to have a healthy lifestyle. Our findings demonstrate tensions and possibilities around smart insurance, discussing relationalities that it opens. While asymmetries as regards information flows and control are evident, the case study does not give grounds for a linear story of a disciplinary mechanism. Instead, imaginaries of surveillance, control, self-determination, and freedom are distributed in multiple ways, and in various degrees. The empirical materials make evident that it is not as if either the control is externalized and given up to institutions and gadgets, or it is retained by the subject. Rather, we emphasize the wide variety of the degrees of autonomy evident in the practices and in the fantasies related to the scope of possibilities that the new technologies provide.

‘Going Smart’: The Politics and Ethics of Resilient Smart Machines

Yael Eylat Van-Essen
Holon Institute of Technology, Israel

The dawn of the 21st century has confronted humanity with natural disasters caused by environmental changes, and with a new kinds of global terrorist threats. In order to enhance the resilience of systems which might be confronted in the future with natural or ‘man-made’ disasters, practices involving sophisticated digital technologies based on big data, have been developed. They operate in real time, integrate data from diverse sources and incorporate applications of artificial intelligence, image processing and advanced communication and sensing technologies.

These ‘smart machines’ adopt cybernetic rationality that avoids causality (as well as the need to acknowledge motivations, beliefs and ideologies that shape actual behaviour) and thus, seemingly, deprive from them their ethical dimensions. They give primacy to remote management and elude direct confrontation with problems evolving in real space. They are based on a systemic approach in which the system’s survival is favoured over that of the individual. In many cases, personal information which is being collected using biometric technologies, is utilized for further physical and cognitive control measures, in military, civil and commercial contexts. Since they are addressed to prevent future threats, they function as a general strategy for managing situations of uncertainty, applying predication practices based on algorithmic and quantitative methods. Such
tools produce feedback-loops that can reinforce and deepen states of social inequality and discrimination between populations, while maintaining neo-liberal interests.

In this paper, with extensive reference to the concepts of resilience and the Anthropocene, and with regard to Gilles Deleuze’s concept of the ‘society of control’, and Brian Massumi’s ‘Ontopower’. I will examine the political and moral implications of this new type of ‘machine’. Relating to design and art projects in these contexts, I will demonstrate how strategies for disaster reduction shape the fabric of our daily life, creating new norms, while redefining basic concepts regarding citizenship and governmentality and the relations between man and his environment.

**Session 8: Thinking and Computing**

What should we call ‘thinking’ at the time of ‘artificial intelligence’ and ‘moral machines’?

Anne Alombert
Université Paris Nanterre, France

This talk aims at questioning the notions of Artificial Intelligence and Technological Singularity in light of the work of Gilbert Simondon and Bernard Stiegler. In their philosophical work concerning technics, Gilbert Simondon and Bernard Stiegler both seem to refuse the abstract analogy between humans and machines, in order to focus their attention on the concrete relation between human “thought” or “spirit” and technical artefacts. This perspective lead them to dispel the myths of Artificial Intelligence and Technological Singularity in order to explore the ethical and political consequences of automation and digitalization.

For both philosophers, there is no sense in comparing technical, mechanical or computational operations to human thought, because it is only by relating to each other through their artificial milieu that the technical living beings that we are can become “spiritual”, “noetic” or “thinking” beings – that is to say, that they can individuate psychically and collectively, through a double process of technical externalisation and psychical internalisation.

According to such a vision of the relation between technical objects and human thought, there is no sense in believing that the acceleration of technical evolution can lead machines to think in the place of human beings, but it seems necessary to answer to the problems of the gap between technics and culture and of the disruption of intellectual faculties by algorithmic and digital technologies.

The question is not about guessing when human intelligence will be exceeded by technological performances, it is about asking how human culture could take care of artificial, automated and digital milieus, so that these technologies can become the support of processes of psychic and social individuation, at the root of a new collective intelligence.
Supplementing humanity: technological desire at work in deconstruction

Igors Gubenko
University of Latvia, Latvia

The paper will explore the link between technology and desire in Derrida’s work. Taking off from the Theory and Practice course notes published earlier this year, I draw on two references the word technique (German Technik) acquires in Derrida’s text, namely, the modern technology as interrogated by later Heidegger, and the psychoanalytic technique as developed by Freud and reinvented by Lacan. “Does psychoanalytic technique belong to the age of what Heidegger defines as modern technology (technique)?” Derrida asks, and I read this question as an invitation to once again rethink the genealogy of Derrida’s deconstructive endeavor. As complex and undecidable as the relation of deconstruction to the Western intellectual tradition is, the defining role that both Heidegger’s thought and psychoanalytic theory had to play in the deconstructive inheritance of that tradition is well assessed. The parallel between Heideggerian technology and psychoanalytic technique, whose possibility is laid open by Derrida’s question, lets me to pose the question of the desire of/for technology (désir de technique), or technological desire. In order to frame this desire in deconstructive terms, I provide a discussion of the logic of supplementarity that Derrida has continuously (re)articulated since the mid 1960s. One of the reasons why Derrida introduced the undecidable concept of supplement was to challenge a set of prominent philosophical discourses on negativity, including Heidegger’s existential analytic with its stress on being-towards-death and Lacan’s psychoanalytic theory with its stress on the constitutive role of the lack of being in subject formation. In my exposition, I take supplementation to be the paradigm for every technological intervention, hence also of operation of deconstruction itself. I argue that if anything like the “desire of the deconstructionist” exists, it must be based on the movement of supplementation. The chief implication of this account of deconstructive technological desire bears on the place of technology in the current posthumanist debate and provides a new perspective on the vestiges of anthropocentrism still perceptible in contemporary critical theory.

Algorithmic computation can be successfully carried on in order to predict the possible trajectories of a system within a pre-given space phase, i.e. according to a certain given number of parameters. However, computation cannot account for the historical change of the space phase that we observe, for example, in historical processes, such as, for example, the evolution of the species, where new parameters and observables emerge. The same holds for the historical process which is science: we can build a computational device to describe and predict the behavior of a system according to a particular theory, but this same device cannot account for a change in the scientific theory according to which it was designed. In other words, while describing physical dynamics and processes, computational devices do not actually produce any real knowledge of their objects. We claim that knowledge cannot be reduced to the description of the possible trajectories of a system that a machine can compute, but that knowledge is the activity of asking why the real is such that it can be described within a real interaction. This question, whose answer cannot be provided by any computing machine, is what determine the historical evolution of science and its particular creativity: machines depend on this creativity but they cannot account for it. So, our claim here is that we should avoid to reduce knowledge to computation in order to be able to think of the future as the creative and ever-changing product of evolutionary dynamics and historical interaction. Knowledge does not consist of what can be computed according to a particular scientific theory, on the contrary, knowledge is the incomputable activity of creating new theories and, in this regard, computational devices are useful tools which allows for empirically verify the theory and its power of prediction.

Session 9: Geopolitics and Technology of Space

'Can Moomin Valley challenge Silicon Valley?' – Exploring Alternative Social Imaginaries of Data Activism

Tuukka Lehtiniemi
Aalto University, Finland

In the market economy, new economic forms remain to be invented (Piketty, 2014), based on alternative future imaginaries (Beckert, 2016). The normativity of imagination comes together with technology in sociotechnical imaginaries (Jasanoff, 2015), or collectively performed visions of desirable future, animated by shared understandings of social order attainable through advances in technology. This presentation explores alternative social imaginaries underlying “MyData,” a data activism initiative originating in Finland that aims to shape a more sustainable citizencentric data economy by means of infrastructure level technological intervention.

MyData activism developed as an extension of open data, and has expanded to an international movement, and onto Finnish technopolitical discourse. The cultural trope “Moomin valley,” employed by a Finnish government minister, underlines the openness, egalitarian nature, trustfulness, and inclusiveness of the imagined economic order, contrasting the dominant economic logic embodied by the US data giants (Zuboff, 2015). MyData promises to shape a more just data economy where individuals equally participate by controlling their personal data. By creating a technology environment where individuals control the primary and secondary use of their data, MyData promises to combine “industry need to data with digital human rights” (Poikola et al., 2015), in order to reach “a fair, sustainable, and prosperous digital society” (MyData, 2019).
Based on material collected through participant observation in MyData, I argue that data activism aiming to enhance citizen agency and advance the ethical principle of autonomy can support multivalent and potentially contrasting future imaginaries. It allows for considerable interpretive flexibility that becomes evident in different ways of framing participation as the outcome of agency: either market choice with data, or the exercise of fundamental rights as data citizens. The contrasting nature of the underlying future imaginaries concretizes in how people are imagined as participants in the information society, and what value is realized through participation.

**Responding to Ethical Challenges in a Digital World**

**James Mittelman**  
American University in Washington, United States

A digital world begets distinct challenges to prevailing ethical principles. This paper identifies the challenges, examines responses to them, and suggests how to understand these issues.

In that smart machines are autodidactic and increasingly self-directed, their decisions could be perilous if artificial entities trigger moral competition between nonhuman and human agents or among extrahuman machines themselves. But in a very different scenario, superintelligent machines' moral judgments may be superior to choices by humans and better our species’ condition (Anderson and Anderson 2011, passim).

Insights into the nexus of humankind and algorithmic actors can be gleaned from contemplating the Anthropocene: arguably, the current geological epoch in which human agency is dominant. Picking up on Bruno Latour’s robust conceptualization (2016; 2017), Edwin Sayes (2014) and other authors contend that humans and nonhumans are elements in structured networks with the capacity to be moral and political actors. A major strength of this approach lies in pluralizing agency. It offers a lens for deep thinking about machine ethics and the possibility of training artificial entities to operate in an ethical manner. Extending Latour’s insights also requires emphasis on hierarchies of domination and subordination: specifically, intersectionality, materiality, and postcoloniality (for critiques of Latourian agency, see Harding 2008, 36; Moore 2016; Thatcher et al., 2016).

Building this analytical framework draws attention to strategic responses to ethical challenges of planetary importance. In the digital realm, they are hitherto threefold: codes of conduct, critical research on lethal autonomous weaponry, and innovative training, evident at some universities and within certain firms. Although grassroots movements resisting digital technologies are growing, their impact to date is limited. These responses are the beginning of a journey for investigating ethical conduct in a digital world. For adaptation and survival of the human species, the gravity of this issue must give pause.

**Affective Framing and the Legitimation of Internet Control in Russia**

**Mariëlle Wijermars**  
Aleksanteri Institute, University of Helsinki, Finland

In today’s hyperconnected world, states are confronted with the global challenge of responding to potentially disruptive online communications, such as terrorist propaganda. In Russia, these threats have been instrumentalised to legitimate a dramatic decline in internet freedom (Human Rights Watch 2017). Scholarship thus far has neglected to examine how the Russian government legitimates and cultivates popular support for these policies. Controlling public opinion may well be the decisive factor in Russia's “success” in expanding internet censorship without arousing significant popular resistance. The paper studies how the internet and its regulation are framed
in Russian political and media discourses. It asks what role the mobilisation of affect plays in legitimating internet control, in particular concerning surveillance and restricting online anonymity.

The paper concentrates on the case of Telegram – a messaging app that offers encrypted communications and, consequently, is popular among extremist, terrorist, but also oppositional and activist groups worldwide. In Russia, Telegram’s ‘channels’ have also established themselves as an influential independent news source within the country’s restricted media landscape. Telegram was banned in Russia in April 2018 because of the company’s refusal to supply Russian security services with decryption keys in compliance with a 2016 anti-terrorism legislation package. The paper draws attention to how, in the case of Russia, the restriction of civil rights is legitimated through mobilising sentiments of anxiety, distrust and fear. How is internet control, including surveillance and the banning of online anonymity, reconceptualised into a virtue of effective state governance, rather than a violation of democratic rights? The paper departs from the hypothesis that affective framing may be particularly persuasive among audiences with limited understanding of information technologies. How do framing strategies, then, play into feelings of vulnerability and sense of incapacity to control perceived dangers inherent to online communications?

Session 10: Ethical Machines and Machine Ethics

Social Values and Crowdsourcing Ethics in MIT’s ‘Moral Machine’

Paul Firenze
Wentworth Institute of Technology, United States

MIT Media Lab’s “Moral Machine” (MM) is a massive online experiment to gather human input into projected decisions made by machine intelligence, in particular, autonomous vehicles (AVs) in situations of unavoidable human fatalities (Awad 2017). The designers of MM argue that the kind of descriptive ethics provided by these experiments (revealing people’s actual, considered choices) is a necessary first step toward creating a “machine ethics” (a type of applied ethics). To this extent, MM is built on good impulses toward democratizing design and away from technocracy (Verbeek 2011). But even in such seemingly narrow applications of machine intelligence, the complexities of moral evaluation become evident when the designers seek to make MM’s decision matrix more subtle, flexible, and comprehensive (hoping to go beyond simple trolley-problem-like calculations) by assigning different, more personal dimensions to the parties involved, including age, gender, size, and, more troubling, “social value” (e.g., doctor, executive, robber, homeless person), among others. While accounting for social distinctions supports many of our moral intuitions (e.g., saving children over adults), this paper will argue that, rather than creating a more comprehensive system of valuation, crowdsourcing morality in this way flattens the inherently pluralistic, qualitative nature of human values, reducing diverse individuals and contexts to an all-commensurating, quantitative “social value score,” which, once codified in machine intelligence and applied to circumstances, raises the specter of dystopian outcomes: think China’s proposed Social Credit System combined with the crowdsourced social rating system of the episode “Nosedive” (Wright 2016) from the BBC series Black Mirror, then applied to the behavior of AVs. While potentially a useful and well-meaning beginning to questions of moral machine intelligence, MM ultimately illustrates the pitfalls of applying moral valuations derived from crowdsourced, descriptive valuations filtered through quantifying mechanisms to reveal estimates of social value.
Nursing Robots and Human Moral Psychology

Michael Laakasuo
University of Helsinki, Finland

We present results from five studies (one qualitative, four experimental) mapping people’s feelings regarding robotic nurses. In four laboratory experiments we evaluated our participants’ judgments and attitudes in situations where either a human or a robot nurse forcefully medicates an unwilling patient. We also evaluated how our participants’ judgments are influenced by the possible death of the patient, the character traits of the human or robot nurse; and whether it matters if a Watson-like AI, as opposed to a human senior physician, makes the treatment decision. Our dependent variables ranged from simple approval ratings to comprehensive mapping of assigning responsibility to different parties involved. It seems that human intuition and preferences for medical robotics are complex and not easily summarized. In some situations people show more leniency towards robots than in others.

In our qualitative study we collected in-depth interviews from about 50 inpatients living in elderly care facilities. They were asked to reflect upon their emotions and judgments with respect to a situation where a robot is forcing medication on people. We collected over 200 pages of materials, and observed several common themes; from fear of losing control to fear of losing beloved nursing staff of one’s residential home.

Landscape of Machine ethics Implementation

Vivek Nallur
University College Dublin, Ireland

Isaac Asimov’s Three Laws of Robotics are the first elucidation of rules that bind machines to a sense of morality. Although, these have been rejected as the basis of building ethical machines, they continue to exert an imaginative force on experimental implementations. This paper describes the various current attempts, from constraint satisfaction to reinforcement-learning to rule-based systems, at adding ethics to machines and/or robots from a computer science perspective. Most attempts at implementation, and mathematical verification, have restricted themselves to consequentialist ethics or even simply the Three Laws. This is partly due to engineering limitations, and partly due to lack of engagement from the non-CS community in building artificial moral agents. This paper surveys the landscape of implementations of artificial moral agents, both from the perspective of which sort of ethics are being implemented, their domains of usage, and the techniques being used to implement them. All of these axes create fundamental constraints on the possible experimentation, and therefore require scrutiny from the non-computer science community. We report on the creation of a game-theoretic framework, that is being used as an experimental setting to detect the emergence of game-specific ‘ethics’ after repeated interactions amongst intelligent, social agents. Specifically, the agents use concepts such as mirror neurons to perform learning as rapid model building. The paper also argues for the creation of intermediate ethical goalposts by non computer scientists, so that there is a wider appreciation for, and validation of the kinds of morality future ethical machines will likely possess.

Imagine the year is 1528 and you have been designated Royal Teacher to a young princess, the heir to a flourishing kingdom. Your task is to educate the girl into a good, wise and just ruler who will not turn into a cruel tyrant destined to enslave her people. Her father, the King, asks you to write down an educational plan. You eagerly grab the quill-pen, but immediately you have to pause. How should you define “good”? What are the goals of a good society? Though some goals appear relatively fixed, many others seem variable and dependent on context. Which virtues will work towards these semi-stable goals, and how should you go about instilling them into the heir?  

These questions were all raised in the popular genres of courtier literature and Prince Mirrors (educational treatises) during the Renaissance, but many fundamental questions of culture, ethics and government remain relevant in the current AI debate, albeit in a new historical context. However, whereas AI-systems in fiction easily turn into world dominating tyrants, we argue here that we need to shift our minds to the role of the “teacher.” The modern Prince Mirror is called a Policy and is intended to foster a development towards competitive, beneficial and safe applications of AI. But what are the underlying assumptions shaping these policies?  

Some 500 years after your Renaissance teaching gig, you realize (if you could travel in time) that it’s not even clear who’s teaching who anymore. Having taught themselves, new AI-systems are now providing solutions and recommendations in a range of fields, but you struggle to understand the underlying assumptions behind these suggestions. Fortunately, you manage to find not only a framework for AI policy development but also ethics guidelines. In this presentation we reveal what the time traveler found.

10 Questions raised in Baldesar Castiglione’s Book of the Courtier published in 1528.
11 AI policies have recently been formulated by EU, China, UK, France, Sweden and Finland among others.

Session 11: Personhood in the Digital World

AIs as Legal Persons: Three Discussions

Visa Kurki
Helsinki Collegium for Advanced Studies, University of Helsinki, Finland

In my paper, I structure the debate surrounding the legal personhood of artificial intelligences (AIs). I argue that this debate actually consists of three distinct discussions. The first one concerns the rights of AIs. This discussion is focused on the moral status of AIs and resembles in many ways the debate over animal rights. The second discussion pertains to AIs – and, in particular, robots – as dangerous entities. Here legal scholars and theorists focus, for instance, on liability questions arising from self-driving cars and autonomous killer drones. Finally, the third discussion pertains to AIs as commercial actors that could for instance buy and sell shares independently and in their own name.
In the talk, I apply my general theory of legal personhood to show how all of these discussions pertain to different facets of legal personhood. I then offer certain theoretical tools for analyzing them.

**Planning for Netborgs: Urban Complexity remediated**

**Jenni Partanen & Seija Ridell**  
Tampere University, Finland

Technologies have always boosted humanity in a co-evolutionary manner and this is nowhere more palpable than in urban settlements, where more than a half of the planet’s human population presently lives. The tools we make and the cities we build with them have pushed us, in a circular loop, to extend our limits as a species. What is new in algorithmic technologies is that they intertwine humans and her machines fundamentally at the ontological level. In the contemporary cities, we use and are used by our smart devices and ambient technical systems with no option to opt out. Technology pervades “being human” from the most intimate level to the level of nested complex adaptive networks. The latter form, echoing N. Katherine Hayles, a cognisphere – a meshed assemblage of data, algorithms organizing it, and the humans.

In our interrelational world, rather than maintaining the dichotomies of human/technology, human/animal, or physical/non-physical research should direct attention to the fluid borders of these realms, remaining alert to new vibrant bundles. As individuals, we are incorporated in the cognisphere as its avatars. In the complex urban techno-meshworks, our networked cyborg avatars and our technologically extended bodies inhabit space as reversible figures of a netborg.

In our presentation, we ask what the new algorithmic condition means for urban spatial planning, including its moral and ethical values. Have the fundamentals of planning become irrelevant in information-intensive environments? How should planners and architects take into account that wearable and built-in technologies extend our sensorium transforming our connections with the physical world? How to include into planning thinking that urbanites as networked cyborgs are present in multiple places simultaneously? We tackle these questions in the framework of complex adaptive systems, arguing that the netborgian regime shift warrants a post-human turn in the paradigms of planning theory.

**Machine Learning, Digital Technologies and Discourses on Citizenship for Individuals with Autism Spectrum Disorder and Others with Exceptionalities**

**Natalia Starostina**  
Young Harris College/The University of West Georgia, United States

The paper will address ethical, cultural, and practical debates on the use of digital devices for individuals with disabilities and, especially, Autism Spectrum Disorder (ASD). Considering the alarming growth of autism which has truly reached the status of epidemics in the U.S., the presentation will address the urgent need to expand and to promote the use digital devices for individuals with autism and other disabilities. Many technological devices, software, and alternative communication systems (AAC, PECS) make possible for individuals with autism to communicate more effectively, for autism, by a legal definition, is “the severe disability significantly affecting verbal and nonverbal communication and social interaction.” Can such devices be adapted to the individual needs of individuals with (dis)abilities, and can such devices “learn” from individuals’ characteristics? How is to make the state and educators become actively promoting the widespread use of such devices? Will new technologies help individuals with autism to become active citizens and workers and to obtain satisfying employment opportunities? In the recent decades, innovative pedagogical approaches such as SEE-KS and the SCERTS Model with its emphasis
on engagement have emerged. It is important to improve the quality of life of individuals with ASD and other exceptionalities and (dis)abilities and to expand meaningful citizenship to them. An interdisciplinary paper, it interweaves the approaches of digital and oral history, disability studies, and special education.

Session 12: Thinking Challenges in the Digital World

Panel Description: Thinking challenges in the digital world

We are immersed in the cognisphere. Our thinking becomes dependent on our digital technologies and they are dependent on us as developers and users. This is the co-constitution and co-shaping of humans and technologies, what Bernard Stiegler has termed as epiphilogenetics and what postphenomenology formalizes as I-technology-world scheme. It does not necessarily mean that there is an imitation or a hidden competition between the two parties. Rather there is an increased entanglement between humans and digital technologies. We need our digital technologies to communicate with each other, we need them to read and write, we need them to see the world. Do we need them in order to think?

Our challenge as philosophers is double. On one hand, we should identify the limits of the human and the technological based on past and present, as well as reveal through a future-looking program the potential changes and assess them for the good and for the bad. On the other hand, we need to map the co-shaping processes and re-assess what does it mean to be human in the cognisphere. The questions that interest us are: Is the cognisphere limited to a certain mode of thinking? Is it limited to the way we think? Does it limit how we think so that thinking becomes “mechanical” or “algorithmized”?

From Thinking to Imagining

Galit Wellner
NB School of Design Haifa, Israel

Imagination has been praised as a basic human capability. For Kant it is that which unites perception and understanding. For Marx it is unique for humans and the distinctive characteristic of human labor over animals like bees and spiders. In the age of the machine, when technologies replaced human muscles, imagination could have been considered exclusively human. But in the digital age, when algorithms start replacing human thinking, the faculty of imagination should be reexamined. Hayles’ notion of distributed cognition (Hayles, 2006) is central to the analysis of the proposed new mode of imagination. This imagination does not seek a new point-of-view but rather works in layers. The content of the layers can be produced algorithmically resulting in Pokemons superimposed through an augmented reality app or hotel prices that change according to hidden parameters. All these technological layers can be grouped as what Ihde terms as micro-perceptions comprised of the “immediate and focused bodily [impression] in actual seeing, hearing etc.” (Ihde 1990, 29) The human imagination is required for macro-perceptions, the cultural and political context in which digital technologies operate. In addition to making sense, human imagination is also required for adding new layers thereby enriching the end result.
Thinking Things and Thinging Thoughts

Lars Botin
Aalborg University, Denmark

Thinking things and thinging thoughts is a playful semantic chiasm that tries to show the relationship in between action and reflection in mainly a phenomenological and postphenomenological sense. At the same time, there is a link to Hayles’ Unthought (2017) where she questions the boundaries between things and humans when it comes to thinking. How can things think and how can thoughts become things through the concept of thinging? Another question that is raised through the construction of the chiasm is what comes first, thinking or thinging, or are the two processes of cognition and the autopoietic work of things in orchestral settings of simultaneous character? In this paper, I will discuss how things are basic for any kind of thinking and how any sort of things is propulsar for thinking and reflection. The paper takes an outset in Peter-Paul Verbeek’s Mediation Theory wherein is present a certain verve of pragmatism that deals with how thinking and action should have a purpose of moral and political character. Thinking and thinging in private and public/social spaces has an intention that transcends the mere process of reflection and doing, where intentionality is directed in streams that takes interrelation and inter-contextuality seriously.

Talking versus Doing: Thinking ‘Making’ in the Digital World

Yoni Van Den Eede
Free University of Brussels, Belgium

“You talk the talk, but do you walk the walk?” The distinction between talking and doing is a trope familiar enough from everyday life, but in contemporary philosophy of technology it is something of a central controversy. Pragmatically and pragmatically oriented approaches such as postphenomenology, e.g., in the work of Verbeek, are keen on “doing” – understood in an entrepreneurial way of designing and developing concrete solutions to problems – and reproach more critically oriented perspectives such as critical constructivism (Feenberg) for only “talking.” Vice versa, critical approaches denounce pragmatic perspectives for their (potential) relapse into instrumentalism, suggesting to account also and actually in the first instance for overarching political and/or transcendental conditions. The question is whether in this debate we are not still working with outworn categories, with implicit sedimented meanings. This paper wants to cast a new light on the issue by way of an approach that comes from combining two unlikely partners: object-oriented ontology (OOO) and the systems thinking of Gregory Bateson. This gives us a fresh and provocative understanding of what “making” (e.g., a (digital) object) is, but also how we should – morally and politically – think and re-envision the interrelation, instead of tension, between “talk” and “walk.”

Thinking and Criticizing

Lyat Friedman
Bezalel Academy of Arts and Design, Israel

Gregory Bateson is well-known for coining the term double-bind to denote a discrepancy. Bateson writes: “the individual is caught in a situation in which the other person in the relationship is expressing two orders of message and one of these denies the other” (1972, 206). Problems begin when a person, who is in an intense relationship, does not have the means to correct the doubled message or express her or his frustration. In this paper I would like to borrow the notion of the double-bind and use it in order to map the orders thinking partakes in. On the one hand, forms of thinking, such as judging, solving puzzles, following a set of given rules and insight, are
one type of order of thinking. Katherine Hayles’ recent book, Unthought (2017) provides a different type of order. It distinguishes between conscious thought and non-conscious cognitive processes. I will argue that by comparing thinking machines to thinking bio-organisms while assuming that the distinction is of the order of the distinction on between conscious and non-conscious thoughts, a double-bind is formed which reduces thinking into a single order that invalidates the other. Criticism will provide the means to correct such doubled message.

PROGRAMME

Wednesday, 6 March 2019

9.00 Opening words (Think Corner, Stage)

9.15–10.15 Keynote I: N. Katherine Hayles (Duke University, USA): “Cognitive Assemblages and Their Implications for Ethics” (Think Corner, Stage)

11.00–12.30 Morning sessions
Session 1: Moral Machines in Literature and Media (HCAS common room)
Session 2: What Do Algorithms Make of Us? (Think Corner, Think Lounge)

13.30–14.30 Keynote II: Maria Mäkelä (Tampere University): “Social Media as Moralistic Storytelling Machine: Emergent Authority and the Viral Exemplum” (HCAS common room)

15.00–17.00 Afternoon sessions
Session 3: Machines as Artists (Think Corner, Think Lounge)
Session 4: Manipulation in Web Search and Social Media (HCAS common room)

18.15–20.00 Artistic programme (Think Corner, Stage)
Thursday, 7 March 2019


11.00–13.00 Morning sessions
Session 5: Ethical and Political Problems with Machines (HCAS common room)
Session 6: Digital Control of Life and Strategies of Resistance 1/2 (Think Corner)

14.00–15.00 Keynote IV: Frédéric Neyrat (University of Wisconsin-Madison, USA): “The Zombies of the Digital: What Justice Should We Wait For?” (HCAS common room)

15.30–17.30 Afternoon sessions
Session 7: Digital Control of Life and Strategies of Resistance 2/2 (Think Corner, Think Lounge)
Session 8: Thinking and Computing (HCAS common room)

Friday, 8 March 2019

9.30–10.30 Keynote V: Erich Hörl (Leuphana University of Lüneburg, Germany): “Environmentalitarian Time: Temporality and Responsibility under the Technoecological Condition” (HCAS common room)

11.00–13.00 Morning sessions
Session 9: Geopolitics and Technology of Space (HCAS common room)
Session 10: Ethical Machines and Machine Ethics (Think Corner, Think Lounge)

14.00–15.00 Keynote VI: François-David Sebbah (Paris Nanterre University, France): “The Two Lights of Morality and Machines” (HCAS common room)

15.30–17.30 Afternoon sessions
Session 11: Personhood in the Digital World (Think Corner, Think Lounge)
Session 12: Thinking Challenges in the Digital World (HCAS common room)

17.45 Closing words (HCAS common room)
MORAL MACHINES? ETHICS AND POLITICS OF THE DIGITAL WORLD
6-8 MARCH 2019
HELSINKI COLLEGIUM FOR ADVANCED STUDIES

MAP OF HELSINKI

1. Helsinki Collegium for Advanced Studies (HCAS)
   Fabianinkatu 24, 3rd floor
2. Think Corner
   Yliopistonkatu 4
3. Restaurant Sipuli
   Kanavaranta 7
   www.ravintolasipuli.fi
4. UniCafe - Päärakennus
   Fabianinkatu 33
   www.unicafe.fi