

EFAK XV

XV Eesti Filosoofia Aastakonverents
Digiühiskond: filosoofilised perspektiivid
Tallinna Tehnikaülikool, 30-31. August 2019

XV Annual Estonian Philosophy Conference
The Digital Society: Philosophical Perspectives
Tallinn University of Tallinn, Estonia, 30-31 August 2019

Teeside raamat
Book of Abstracts

Catherine Allamel-Raffin, University of Strasbourg, catherine.allamelraffin@unistra.fr
Jean-Luc Gangloff, University of Strasbourg, gangloff@unistra.fr

What Daston and Galison do not say about scientific objectivity

The complex nature of the concept of scientific objectivity is probably due to the fact that its meaning has been constructed in the course of history as normative positions and new practices emerged, especially in the field of natural sciences. New defining characteristics can always be added to the previous ones. In their book *Objectivity* (2007), Lorraine Daston and Peter Galison consider that "the history of scientific objectivity is surprisingly short. It first emerges in the mid-nineteenth century and in a matter of decades became established not only as a scientific norm, but also as a set of practices, including the making of images for scientific atlases" (2007, 27). For Daston and Galison, scientific objectivity arises at the time when the word, the concept and the types of realities that it subsumes, merge. However, according to them, it is necessary to distinguish what they call "mechanical objectivity" from two other epistemic virtues, "truth-to-nature" and "trained judgment". After having listed some objections addressed to the book since its publication, we will emphasize the fact that Daston and Galison do not mention in their definition of the concept of "mechanical objectivity" some essential modalities of what we will call "objectifying detachment", because of their epistemological and ontological options that deserve at least to be discussed.

Bibliography:

Daston, L., Galison P. (2007). *Objectivity*, New-York, Zone Books.

Nathan Andersen, Eckerd College, andersnt@eckerd.edu

Being Online: Existential Phenomenology and the Digital Lifeworld

The central maxim of the phenomenological method is "to the things themselves." We should let go of our theories about and prior understandings of the subject matter to be investigated, and focus on how we encounter that subject matter experientially, in the world as we live it. I propose that this is a productive approach for investigating philosophically the "digital society," or what Jamie Susskind calls the "digital lifeworld." While an understanding of computers, and of the underlying technologies of the internet and networked devices, and an understanding of the companies that control them and laws that regulate them are valuable for a variety of purposes, coming to grips with the nature of the digital society requires, first and foremost, an appreciation of the ways in which it modifies the experience of its participants. Experience, as Heidegger clarified, is always the experience of someone who experiences herself to be among others and in a world. My experience, for example, is always at once "my" experience — so that I have a sense of a "self" as the one having the experience — and I experience myself as one among others — a "we" with whom I share or in principle could share that experience — and it is always experienced as of the "world," or as proceeding from a shared reality that is independent from this or that experiencing agent. The question, then, of what it means to "be online" is how networked technologies modify our experiential sense of ourselves, of others, and of reality. As a starting point I examine Sartre's account of the kind of agency involved in listening to the radio, which he develops in his Critique of Dialectical Reason, and consider whether his approach might be adapted to account for the experience of interacting online.

Bibliography:

Sartre, J.-P. (2006). *Critique of Dialectical Reason*, Vols. I & II. Trans. Alan Sheridan-Smith. Verso.

Zsuzsanna Balogh, Eötvös Loránd University, baloghzsphil@gmail.com

Can carebots care about us?

In my paper I would like to reflect on the importance intersubjectivity has in communication and base my view of human, person-to-person communication on a phenomenological theory. I argue that communication with existing as well as future social robots has to be laid on different (non-intersubjective) foundations, which includes a sharp awareness of their nature and their cognitive abilities in order to avoid misunderstandings and potential emotional trauma (to those concerned). My starting assumption is that we should take human-to-human communication as the basis of how to think about communication with social assistive robots. In human-to-human or intersubjective communication, the cognitive capacity of social cognition, mindreading or empathy (e.g. Hutto et al. 2011, Zahavi 2014) is an essential ‘ingredient’. According to a somewhat radical but well-supported phenomenological view of intersubjectivity (Zahavi), this means that we perceive the mental states of the other person *directly* and as a result also perceive ourselves (‘the social self’) in certain ways as seen from the other person’s perspective. These subtle but real processes, as I argue, are essential for us to feel human and to perceive each other as partners in communication. Social assistive robots, to this date, however strikingly smart they may be, are not capable of having such cognitive processes, therefore we cannot and should not view them as equal partners in intersubjective interaction.

Empathy however also has another and more common meaning (closer to ‘compassion’), which involves caring and feeling for the other, as most people expect healthcare and social assistive agents to do, especially in the case of elderly people, who may be more vulnerable and susceptible to manipulation. Due to their age, their potential cognitive condition and their need for company, elderly people may be more prone to believing that a robot/carebot’s certain behaviour expresses genuine emotion and empathy as well as real intersubjectivity without any reflection on the matter, which in my view strongly indicates that they should be properly prepared and educated to communicate with and approach social robots to avoid any self- or otherwise deception and emotional trauma. I conclude my paper by making a number of suggestions about potentially (mostly emotionally) harmful scenarios and how previous awareness of the above points could help avoid these.

Bibliography:

Hutto, D. D., Herschbach, M., & Southgate, V. (2011). Mindreading and its alternatives. *Review of Philosophy and Psychology*, 2(3), 375-395

Zahavi, D. (2014). *Self and Other: Exploring Subjectivity, Empathy, and Shame*. Oxford University Press.

Camilla Cannon, camillakcannon@gmail.com

Nonbinary Gender Identity and Algorithmic-Psychometric Marketing Legibility

Recently, discussions of non-binary gender identity have been increasingly featured in mainstream media. These appear primarily in three kinds of publications: mass market magazines such as Vogue, USA Today, and The New York Times; advertising and business editorial magazines such as Adweek and Forbes; and marketing reports from consumer insight strategy groups. I examine how intra-community discussions of non-binary identity are mined in order to create a rhetorical facsimile of non-binary identity that encourages all consumers to attain an “authentically” gendered self through consumptive self-creation. I argue that these forms of ostensibly liberatory consumptive gender expression serve to render all consumers more legible within algorithmic-psychometric marketing systems. The strengthening of these systems helps shore up various forms of surveillance (governmental, financial, healthcare) that have negative effects on the lives of trans and non-binary people. I am hopeful that my analysis of the instrumentalization of non-binary gender identity in these surveillance systems will lead to increased scrutiny of the desirability of pop cultural and marketing recognition, as well as encourage further interdisciplinary investigation of the various intersecting surveillance networks that negatively affect the lives of trans and non-binary individuals.

Eleonora Cugini, University of Padova, eleonora.cugini@gmail.com

For a Critique of Artificial Intelligence: Ethics, Normativity and Freedom.

AI ethics faces core questions that challenge both the idea of progress and the stability of the democratic system. For instance, AI ethics reflects on technologies that, if on the one hand seem to reproduce human forms of intelligence, on the other hand do so by developing artifacts and softwares that often blur the boundary between peaceful/civil use and military or control (e.g.: drones, bots, facial recognition, algorithms, etc.).

Philosophical reflection (from Marx to Heidegger, from Benjamin to Foucault, from Pasolini to Simondon, Derrida, Stiegler) has very often addressed technology and its contradictions with respect to progress, understood as the improvement of human living conditions (and therefore with the same problematic notion of ‘progress’). Yet the AI ethics seems to provide philosophy with an even more radical issue. The AI ethics seems in fact to reshape the question about the relationship between ethics and intelligence, understood as the process of self-understanding of the human being.

The hypothesis I would like to put forward is that the AI ethics ultimately consists of a critical reformulation of the question about human essence.

I will begin by addressing it through the method of critical theory: I will investigate the meaning of ‘ethical’ good by preliminarily highlighting its specific type of immanent and essentially critical normativity (Honneth; Jaeggi; Menke) which differs from the normativity (external and universally abstract) of the ‘moral’ just.

Starting from this approach, I would like to understand:

a) whether the criteria for judging both the patterns of data which informed an AI and the process of elaboration of those data (machine learning) as “good” or “bad”, “just” or “unjust” can or

should be the same criteria for judging and criticize the social norms and the process to learn such norms and which are these criteria;

b) to what extent the ethical problems of programming an AI reflect the process of self-understanding of the human being, of its limits and of its capacity for emancipatory progress;

c) whether the ethical question on AI is at its core a political question: that is, a question that, while critically reformulates the notion of 'essence', also interrogates the emancipatory processes of the realization of human freedom.

Biobibliography:

Benjamin, W. (2003). *On the Concept of History, Selected Writings, Vol 4 (1938-1940)*.

Cambridge: Harvard University Press.

Derrida, J. (1989). *Of Spirit: Heidegger and the Question*, London: The University of Chicago Press.

Derrida, J. (1994). *Specters of Marx*, New York/London: Routledge.

Derrida, J. (2005). Cinema and Its Ghosts: An Interview with Jacques Derrida, *Studies in the Cultural Politics of Education*, 37(1-2), pp.22-39

Foucault, M. (1977). *Discipline and punish: the birth of the prison*, New York: Pantheon Books.

Foucault, M. (1978). *The History of Sexuality. Vol 1: The Will of Knowledge*, New York: Pantheon Books.

Heidegger, M. (1977). *The Question Concerning Technology*, New York/London: Garland Publishing.

Honneth, A. (2007). *Disrespect. The normative Foundation of Critical Theory*, Cambridge: Polity Press.

Honneth, A. (2014). *Freedom's Right: The Social Foundations of Democratic Life*, New York: Columbia University Press.

Jaeggi, R. (2008). Rethinking Ideology, *New Waves in Political Philosophy*, Eds. C. Zurn & B. de Bruijn, Palgrave, Herbst, pp. 63–86.

Jaeggi, R. (2018). *Critique of Forms of Life*, Cambridge: Harvard University Press.

Marx, K. (1992). *Capital*, Vol 1, New York: Penguin.

Menke, C. (2015). *Kritik der Rechte*, Berlin, Suhrkamp.

Menke, C., Khurana, T., (2019). *Paradoxien der Autonomie. Freiheit und Gesetz I*, Berlin: August Verlag.

Pasolini, P. P. (1997-2014) *Scritti corsari*, Milano: Garzanti.

Stiegler, B. (1998-2010). *Technics and Time*, Vol 1-3, Stanford: Stanford University Press.

Stiegler, B. (2016). *Automatic Society, Vol 1: The Future of Work*, Cambridge: Polity Press.

Stiegler, B. (2018). *The Neganthropocene*, London: Open Humanities Press.

Dragana Ćirić, dragana.ci@gmail.com

Digital Spatial Diagrammatics: Space and Information in the Age of Strategic Design Intelligence [datascares]

The core interest of this study will be the relationship between space, data/information, spatial inscription and representation, reasoning and decision-making as part of the current digital phase. We will look at how digitality (as the cultural and philosophical framework), digital sciences and technologies (as its practical counterpart) and digital media and creativity have transformed spatial awareness, action, notation and imagery, and what role spatial/architectural design thinking and performance have taken within these tendencies as both - their initiators and leading developers, and their contributors highly influenced by digital ideologies, tools and the unlimited design possibilities warranted in the digital world.

Starting from the outline of space informatisation and cyberisation as a broader social condition, more focused arguments will be built upon the question of what digital diagrammatics stands for. Operating along the scale from the logical abstractions and complexities in inferencing that guide intellectual plots, processes and relational networking behind the ideas and concepts to their

materiality, objecthood or different physical manifestations, digital diagrammatics has acquired a specific role of spatial agency by data. Datascape – environments defined/designed/mapped by data; scapes of information themselves; superpositions of multiple physical and virtual realities inasmuch as one of them alone - will be examined from the diagrammatic architectural perspective asking: How diagrammatic and spatial strategies employ and articulate data for different purposes and what are the outcomes of digital data operations in a diagrammatic spatial layout? What are the ways they are practiced through across the disciplinary systems while dealing with data in documentary, scientific or creative manner? Supported by comments on historical precedents and forms that made their foundations, the answers on how digital language and literacy (as pivotal features of the underlying system of new digital media and its hybrid aesthetics) have transformed the ways spatial or any other information is used and comprehended, are expected to come as a conclusion.

Hermann Diebel-Fischer, University of Rostock, hermann.diebel-fischer@uni-rostock.de

Re-Discovering the Humanoki: Robotics, Algorithms, and the Quest for the Good Life

Recent developments in the field of information and communication technologies caused a new wave of ethics which seeks to render the »digital society« a society and an environment that on the one hand does not reproduce the injustices, discriminations, and biases of our current world and that on the other hand will employ technology in a way which serves humans and does not endanger them.

This endeavor has its outset in the philosophical question What is Man? which has been addressed by Kant, Cassirer, Scheler, Plessner, etc. In my paper (which is based on my habilitation project), I argue that in this new wave of ethics, which we are currently experiencing, we - once again - enter the field of philosophical (and theological) anthropology. Technology-induced questions in the field of ethics explicitly or implicitly address or create a concept of the human, which puts humanity in opposition to technology, as can be seen e. g. in the debates on algorithms or healthcare robots, and in groups such as AlgorithmWatch, the IEEE Ethically Aligned Design-initiative, or the EU's HLEG AI's Ethics Guidelines for Trustworthy AI.

Technology does not only serve as a cultural means to solve problems in human life, it furthermore serves as a means with the help of which humanity gains an insight into itself. Technological artefacts, be it an algorithm or a robot, do not only reflect mankind's self-image (Schwarke 2015) but also help to answer the question What is Man? by means of ethical assessments. In the construction of right and wrong concerning the design and employment of technology, the anthropological assessment of mankind's being in the world is not merely a by-product but a precondition. Thus, under the condition of »digitality« (Stalder 2016), the digital society does not only demand ethical but also anthropological (re)considerations.

Bibliography:

- Ala-Pietilä, P., et al.(2019). *(EU Commission's independent HLEG): Ethics Guidelines for Trustworthy AI*.
IEEE: Ethically Aligned Design (2016, 2019).
- Schwarke, C. (2015). Technik und Natur. Über Nachhaltigkeit im Umgang mit dem Bild der Natur, in ed. E. Gräb-Schmidt, *Was heißt Natur? Philosophischer Ort und Begründungsfunktion des Naturbegriffs*, Leipzig, pp. 83-95.
- Spielkamp, M., et al. (2016). AlgorithmWatch, algorithmwatch.org (accessed May 15, 2019).
- Stalder, F.: Digitalität, Berlin.

Why philosophers of science should care about the problems of the “pure public”

Several philosophers of science, including Heather Douglas and Alison Wylie, have recommended some involvement of non-scientists in science. The aim of my presentation is to argue that such arguments can benefit from analyses of public participation initiatives in practice.

Two common aims of these initiatives are to involve either “lay experts” who have relevant experience or representatives of the general public. In the latter case, these representatives are often conceived in terms of the “pure public” – persons who can represent the general citizenry and take part in open-ended deliberation because they have no pre-existing opinions or experience concerning the issue. Several analyses attract attention to the problematic character of the notion of “pure public” – persons lacking previous involvement with the issue may also lack motivation to participate or experience enabling them to work with experts and policymakers on more equal terms. Selecting the “pure public” may thus undermine success of public participation.

I argue that philosophers who recommend forms of public participation that rely on the “pure public” – such as Douglas who proposes consensus conferences – should take these criticisms into account. Philosophical arguments that focus on alternative forms of public participation – such as Wylie’s proposal to collaborate with lay communities offering epistemically relevant perspectives – avoid this issue. Nevertheless, the discussion of the latter can also benefit from awareness of different expectations about public participation. While involving “lay experts” has a variety of epistemic and political benefits, it makes science more democratic in just one particular sense and does not address the issues of representation for the general citizenry. Being clear about the aims and type of public participation is important, since a mismatch may cause disappointment in democratisation of science, including proposals by philosophers of science.

Teaduse institutsionaalne mandumine

Alates Popperist ja Lakatosist on teadusfilosoofia üheks põhiküsimuseks peetud eraldusjoont teaduse ja mitte-teaduse vahel. Lakatosi järgi on pseudoteadus selline mitte-teadus, mis esineb teadusena, kasutades teaduse ühiskondlikku autoriteeti. Sellest ajast peale on avaldatud mägede viisi tekste sellest, kuidas mitte-teadused nagu näiteks astroloogia ei ole teadused.

Ent vaenlane ei ole teisel pool piiri. Vaenlane on meie keskel.

Paljud mitte-teaduslikuks tunnistatud uurimused on vägagi teaduslikud, sellal kui tohutu kogus teaduslikuks tunnistatud tekste on mõttetu ebateadus.

Mitte-teadlase mõtestatud tekst mitte-teaduseks tunnistatud valdkonnast võib olla teaduslikum kui teadlase palka saava isendi publikatsioon teaduseks tunnistatud alalt.

Parim viis teha pseudoteadust on jahmerdada teaduseks tunnistatud suure ürituse sees.

Teadus on institutsionaliseerunud, teadlased saavad palka. Nad moodustavad preestritekasti. Nende huvides on mitte lasta piruka juurde teisi.

ETIS on üks kaasaegseid teadusbürokraatia mooduseid, kuidas tekitada müra tootev preestriteklass ning hoida nende võimu. Süsteem toodab kvantiteeti ja selles püsivad normaalteadlased Kuhni tähenduses. Nad kordavad mehaaniliselt ühte ja sama meetodit erinevatel objektidel. Täna loevad nad liivaterasid siin, homme seal.

Niinimetatud koolkonna väljaandes tuleb tsiteerida pühasid lehmasid ja ei tohi tsiteerida kriitikuid. See on ringkäendussüsteem.

Mõne aja pärast on päris teadlased hakanud kaevureiks, tuletõrjujaks ja öövahtideks, sest preestriteklass on nad olushierarhiast välja tõrjunud.

See on pokazuha-teadus.

Tulnukas, kes planeedile satub, on alul segaduses. Lõpuks ta taipab: see on täpselt nagu Poe õudusnovellis hullumajast, mida külastasid inspektorid. Arstid olid keldrisse luku taha pandud ja hullud olid endile valged kitlid selga tõmmanud ning esinesid arstidena.

Kuidas teisiti tõlgendada ühiskonda, mis põletab ja näljutab oma geeniuseid ja nimetab teaduseks keskpäraste interpretaatorite armeed, kes ülikoolides õpetavad geeniuste tekste, aga alati alles pärast nende surma?

Isegi teadusfilosoofia on silmakirjalik teoloogiline õigustus preestrite klassile.

Bibliograafia:

Lakatos, I. (1991). Teadus ja pseudoteadus, *Akadeemia*, nr 2, lk 258 – 268.

Poe, E. A. (1844). *System of Doctor Tarr and Professor Fether*, autograph manuscript.

Kelly Hrupa, Tallinna Ülikool, kellyhrupa@gmail.com

Andres Luure, Tallinna Ülikool, andresluure@gmail.com

Vabadus, vastutus ja armastus

Kui mees on ebatervislike eluviiside tõttu haigeks jäänud, kas siis tema naine vastutab selle eest? Kas ta on liiga vähe armastanud? Naine ütleb: "Ma armastan teda. Tahan, et ta oleks õnnelik ja vaba. Kuidas ma võiksin teda takistada?"

Armastus jätab armastatule vabaduse ja edendab tema vabadust olla see, kes ta on, ja saada selleks, kes ta on, teha seda, mida ta teeb, ja hakata tegema seda, mida teeb tema. Armastaja vastutus on aidata armastatul kanda vastutust oma vabaduse eest.

Selleks et ma saaksin armastada, tuleb mul iseendaga vabaduses koos elada, iseennast armastada. Kuidas ma vastutan iseenda eest? Kas ma pean end distsiplineerima ja piirama, kas ma pean armastuse nimel iseennast sundima? Kui mina armastan, kes siis armastab? Mind on ju kaks, kuidas muidu ma saaksin end armastada? Mulle tundub, et mina ei armasta ning seetõttu ma pean mind distsiplineerima, selleks et mina viiks mu armastust ellu. Kas ma siis armastan mind? Kui ma armastan mind, siis ma jätan minule vabaduse ning edendan minu vabadust, ja mitte sellepärast, nagu mul oleks ükskõik, kas või kuidas mina armastan, vaid sellepärast, et ma ei saa mind armastama õpetada, vaid vastupidi, ma saan õppida armastama ainult minult, kes on objektiivne armastaja ega lase armastust hinnangutel moonutada.

Roomet Jakapi, Tartu Ülikool, roomet.jakapi@ut.ee

Alvin Plantinga, religioossed uskumused ja kujutlused

Alvin Plantinga (snd. 1932) on kuulus analüütiline religioonifilosoof, epistemoloog ja loogik, kelle arvukatele publikatsioonidele lisandus aastal 2000 suurteos Tagatud kristlik uskumus. Selles teoses arendab ta välja epistemoloogilise mudeli, mida ta teistlike uskumuste puhul nimetab A/C–mudeliks ja spetsiifiliselt kristlike uskumuste puhul laiendatud A/C–mudeliks („A“ tähistab Aquino Thomase ja „C“ Jean Calvini vaadet, nagu autor neid tõlgendab). A/C–mudeli järgi kuulub inimese loomuliku kognitiivse varustuse hulka spetsiaalne võime nimega *sensus divinitatis*, mille

abil omandatakse teistlikke uskumusi. Täpsemalt, kui see spetsiaalne kognitiivne mehhanism on töökorras, siis teatud tingimustes tekitab see inimeses tõeseid uskumusi Jumala olemas- ja kohalolu kohta. Tegemist on alususkumustega, mis tähendab, et need tekivad spontaanselt, ilma argumenteerimisega ja järeldamiseta, sarnaselt sellega, kuidas väidetavalt tekivad tajukogemusel põhinevad uskumused. Laiendatud A/C–mudeli järgi tekivad ka tõesed kristlikud uskumused inimeses teatud tingimustes spontaanselt ja selle põhjuseks on Püha Vaimu ergutus. Viimase näol on tegemist üleloomuliku mõjutustegevusega, mis kutsub evangeeliumi kuuljas või lugejas esile usku Kolmainususse, inkarnatsiooni, lunastusse jne. Plantinga mudeli järgi ei koosne usk siiski üksnes uskumustest selle sõna kitsas epistemoloogilises tähenduses. Mudeli kirjelduses mainib ta ka usulisi tundmusi kui faktoreid religioossete uskumuste omandamisel ja omamisel. Leian, et ühe olulise faktori jätab Plantinga siiski usu kirjeldusest välja ja selle poolest on tema mudel ka ebarealistlik abstraktne konstruktsioon. See puuduv faktor on kujutlemine. Püüan ettekandes näidata, et ja kuidas toimivad religioossed kujutlused religioossete uskumuste omandamisel, püsimisel ja ka hülgamisel. Ühtlasi vaatlen ja hindan epistemoloogilisi tagajärgi, mida kujutluste sissetoomine võiks põhjustada kõnealuses mudelis.

Lauri Kallio, University of Turku, lauri.kallio@alumni.helsinki.fi

Late German idealism and the 1860s Kant revival

The talk falls into two parts. First, I briefly introduce the concept of late German idealism. It refers to the idealist philosophy approximately between 1840 and 1880, which emerged in the aftermath of the German idealism (approx. 1780–1830). I begin with a comment on F.C. Beiser's concept of the late German idealism.

I also outline the history of the Philosophical Society of Berlin (Philosophische Gesellschaft zu Berlin), which I consider as an organization of the late German idealists.

Second, I discuss the reaction of the late German idealists towards neo-Kantianism, which occupied a leading position in the German universities from the 1870s until the First World War. The idea that "getting back to Kant" resolves the crisis of German philosophy became popular in the early 1860s. The early neo-Kantians argued that the failure of the various philosophies of the German idealism was due to the misinterpretation of I. Kant's critical philosophy. On the other hand, early neo-Kantians renounced "dogmatism" concerning Kant. They argued that there are flaws in Kant's theory and it has to be revised according to the latest results of science.

As an example of the discussion between the neo-Kantians and the late idealists, I address C.L. Michelet's (1801–93) – secretary of the Berlin Society – critique of Eduard Zeller (1814–1908).

Zeller focuses on Kant's epistemology and adheres to the dualism between the thing-in-itself and its appearance. Yet Zeller argues for stronger realism than Kant and that the thing-in-itself is eventually not unknowable. Michelet proceeds from the possibility to overcome the Kantian dualisms. He disagrees with Zeller's attempt to make epistemology the formal foundation of philosophy and also criticizes Zeller's concept of "the given", which is tied up with representations rather than concepts.

Bibliography:

- Beiser, F. C. (1862). *Late German Idealism. Trendelenburg and Lotze*. Oxford: Oxford University Press.
- Michelet, C.-L. (1862). Zeller, als Kritiker und Empiriker, *Der Gedanke Philosophische Zeitschrift* 2, no. 4, pp. 288–92.
- Michelet, C.-L. (1873). *Hegel und der Empirismus: Zur Beurtheilung einer Rede Eduard Zellers*, Berlin: Fr.-Nicolaische Verlagsbuchhandlung.
- Zeller, E. (1862) *Ueber Bedeutung und Aufgabe der Erkenntniss-Theorie: Ein akademischer Vortrag*, Heidelberg: Buchhandlung von Karl Groos.

On the Domestication of Chatbots

Developments in natural-language processing algorithms (i.e. chatbots) have been driven by one main imperative: to become as anthropomorphic as possible. While the early success has been lauded as technological breakthroughs, we now have reached levels that almost skip over the “uncanny valley”.

The reaction from the general public as well as professional philosophers signaled a lack of preparation on how to deal with a future of intermingled AI and human speakers. And while this technology has yet to reach mass application, the method of its success is replicable and will likely lead to different forms of machine-learned chatbots, often indistinguishable from human speakers. I am arguing that for a norm- and goal-guided development of chatbots, a framework ought to be developed that recognizes both a chatbot’s non-humanness and its superiority over other forms of socially incorporated entities, such as domesticated pets. This leads to a unique social position within human society, for which we currently lack both terminology and intuitions, leaving AI researchers without clear guidelines and the public without opinion.

The proposal in this talk is to understand AI that is capable of establishing relationships with humans as a new social category. With a genuinely new approach on how to incorporate AI into human society as quasi-agents, old and misleading fault lines will be less problematic. The imperative of developing chatbots to become more and more human can be replaced to adjust for social expectations of how AI is to behave both in a private manner as well as in public.

Understanding chatbots as a new social category would allow for a framework to be developed that recognizes the benefit of socially integrated AI as well as the future the development of chatbots can take.

“Truth” and democracy: adversaries or allies?

A lot of political thinkers have seen claims to truthfulness and democratic politics as something inherently incompatible. Some have seen facts as tyrannic and unsuitable for politics, others are fearful of removing the “political” from politics and emphasize the necessity of agonism in politics. The word “truth” in political philosophy brings to mind authoritarianism and an unpleasant history. Should the justification of democracy exclude any pretension to epistemic success due to its authoritarian character?

Invernizzi-Accetti claims that relying to epistemic success in the justification of democracy:

- 1) takes away autonomy from the people. It treats politics as technical problem-solving (but one of the questions of politics is *what is a problem*, i.e. what is to be solved).
- 2) justifies epistocracy, rather than democracy. If epistemic standards are the only ones, then if any other procedure (rather than democracy) can be proven to be epistemically the most successful, the justification of democracy fails.
- 3) undermines pluralism since *truth* is necessarily one.

He concludes that since epistemic success conflicts with other democratic values, attempts to use it as part of justification should be abandoned.

I will defend the use of arguments relying on epistemic success as a part of the justification for democracy. I will answer to Invernizzi-Accetti and argue that epistemic success as a part of the justification of democracy:

- 1) is necessary for preserving autonomy

- 2) does not lead to the conclusion of preferring epistocracy to democracy and
- 3) is compatible with reasonable pluralism.

Bibliography:

Invernizzi-Accetti, (2017). Does Democratic Theory Need Epistemic Standards? Grounds for a Purely Procedural Defense of Majority Rule, *Democratic Theory*, 2, pp. 3–26.

Eve Kitsik, Tartu Ülikool, eve.kitsik@ut.ee

Eetika ja metafüüsika suhtest tehnoloogia arengu kontekstis

Tehnoloogia arenedes võib juhtuda, et metafüüsilised küsimused, mida seni on arutatud vaid teoreetilisest huvist lähtuvalt, muutuvad praktiliselt tähtsaks. Näiteks muutub teleportatsiooni võimalikuks saades praktiliselt tähtsaks metafüüsiline küsimus, kas teleportatsiooni käigus algne isik hävib – sellest näib sõltuvat, kas teleporteerimisjärgne isik peab maksma ära algse isiku võlad või kandma tema kriminaalkaristust. Samuti muutub vaadeldava funktsioneerimise poolest inimesega sarnaste robotite loomisel praktiliselt tähtsaks metafüüsiline küsimus, kas sellistel robotitel on teadvusseisundid samas mõttes, nagu inimestel. See küsimus näib olevat praktiliselt tähtis nii robotitest tulenevate võimalike ohtude kui ka nende moraalse staatuse hindamiseks. Kuidas sellistele praktiliselt tähtsatele metafüüsilistele küsimustele läheneda?

Oma ettekandes eristan kaht lähenemist: metafüüsikast ja eetikast tõukuvat. Metafüüsikast tõukuval lähenemisel tuleks kõigepealt lahendada metafüüsiline probleem (näiteks teleporteerimisele isiku A samasus teleporteerimisjärgse isikuga B) ning alles seejärel leida selle lahenduse praktilised järelmid (näiteks kas B peaks maksma A võlad). Eetikast tõukuva lähenemise kohaselt on aga metafüüsilise probleemi lahendamiseks vaja neidsamu praktilisi järelmeid arvesse võtta. Äärmuslikul juhul metafüüsiline küsimus taandatakse eetilisele küsimusele: kuidas oleks eetiliselt õige kasutada mõisteid nagu 'sama isik' või 'teadvus', arvestades, mida samasuse või teadvuse (mitte)omistamine praktikas kaasa toob? Konkreetselt robotite teadvuse küsimusega seoses on sellist eetikast tõukuvat lähenemist toetanud H. Putnam, laiemalt (sh ka isikusamasusega seoses) näiteks A. Thomasson.

Lisaks esimese tasandi metafüüsikast ja eetikast tõukuvate lähenemiste eristamisele teen oma ettekandes sarnase eristuse ka järgmisel tasandil, kus me eelkirjeldatud lähenemiste vahel valime. Teise astme metafüüsikast tõukuva lähenemise kohaselt sõltub valik esimese astme lähenemiste vahel metafüüsilise probleemi olemusest ja lahendatavusest. Teise astme eetikast tõukuva lähenemise kohaselt peaksime aga valima esimese astme lähenemiste vahel eetilistest kaalutlustest lähtuvalt. Oma ettekandes kaitsen esimese astme eetikast tõukuvat lähenemist, tõukudes seejuures ka teisel astmel eetikast. See tähendab, et esitan eetilistele kaalutlustele tugineva argumendi (esimese astme) eetikast tõukuva lähenemise poolt.

Lilian Langsepp, Tartu Ülikool, lilian.langsepp@gmail.com

Juhtumianalüüs: monokord teadusfilosoofia kontekstis

Ettekande eesmärk on tuua esile antiik-kreeka päritolu monokordi-juhtumi see aspekt, mille sihiks oli helide omavaheliste konsonantsete ja dissonantsete suhete erinevust demonstreeriva ja usaldusväärsele teadmisele toetuva tõestuse väljatöötamine. Taju ja teadmist linkivana pakub see juhtum siinkohal huvi teadusliku instrumendi ja teadusliku eksperimendi raamistikus. Instrumendina kujutab monokord endast helitugevust võimendavale kõlakastile pinguletõmmatud keelt, mida saab liigutatava roobiga kaheks osaks jaotada. Keele osi puudutades tulevad kuuldavale

helid, mis vastavad keele osade ja terviku arvulistele suhetele, seega muutuvad arvulised suhted kuuldeliselt tajutavaks. Selle ülimalt minimalistliku seadeldise abil osutus võimalikuks erinevate intervallide võrdlemine ning kõige olulisema tulemusena lihtsate arvsuhete ja konsonantsete kooskõlade omavahelise vastavuse teoreetiline põhjendamine. Sajandeid hiljem sai see teoreetiline teadmine aluseks mandri-euroopa keskaegse õpetatud mitmehääle muusika kõlapildile, mõjutades muusika esteetilist hindamist ja reeglistades muusika komponeerimist. Monokordi ning selle abil helisuhete uurimise kirjeldamine ja kujutamine traktaadist traktaati ning muusikaõpikutes kuni meie kaasajani teeb sellest ühe enim kujutatud seadeldise muusikateooria ja muusikaajaloo valdkonnas. Nii on ta leidnud ka koha Whipple'i teadusajaloo muuseumi kollektiooni akustika osas, kus esitletakse monokordi kui antiikset muusika- ja teaduslikku instrumenti. Kõik selle instrumendi leiutamise seonduv on omistatud Pythagorasele, kuigi otsesed kirjalikud allikad selle kohta teadaolevalt puuduvad. On võimalik ka see, et monokord oli talle omistatud eksistentsi esimestel sajanditel idealiseeritud instrument, mitte-eksisteeriv, ainult diagrammi kujul traktaatides esinev. Kui aga võtta arvesse Pythagorasega seotud legendi, illustreerib see teaduslikule eksperimendile iseloomulikku käiku probleem – juhus – vaatlus – eksperimendid – mudeldamine – teoreetiline tõestus. Käesoleva ettekande teadusajaloolise konteksti aluseks on David Creese'i vana-kreeka autorite allikatele tuginev ulatuslik ja põhjalik monokordi käsitlev monograafia.

Bibliograafia:

- Creese, D. (2010). *The Monochord in Ancient Greece Harmonic Science*, Cambridge University Press.
- Anicius Manlius Severinus Boethius (1989). *Fundamentals of music*, with Introduction and Notes by Calvin M. Bower, ed. by Claude V. Palisca, New Haven & London: Yale University Press.
- Rees, T. (2009). *Monochord: an ancient musical and scientific instrument*. Explore Whipple Collections, Whipple Museum of the History of Science, University of Cambridge.
<http://www.sites.hps.cam.ac.uk/whipple/explore/acoustics/monochord/>

Juho Lindholm, University of Tartu, juho.lindholm@gmail.com

Rein Vihalemm as a Deweyan Pragmatist

John Dewey anticipated much of the post-Kuhnian philosophy of science. He conceived of science as concrete and practical problem-solving rather than as a set of true propositions or statements: science was a mode of dynamic action done by people rather than a static and abstract system which is considered as independent of scientists. He argued that knowledge is a kind of action; that this identification has already been made in physical science but not in philosophy; that theories and ideas are hypothetical and are not independent of practice because they direct experiment; that all discoveries of science are tentative and instrumental (which does not preclude them from being true); and that the end of science is not the contemplation of eternal and immutable truths but the intelligent and technical regulation of the human environment and society. These ideas were central also for Rein Vihalemm. He does not cite Dewey – he only mentions him once – but his program of practical realism in philosophy of science is in many respects identical to Dewey's later philosophy.

I will argue that (1) Vihalemm's project is a special case of Dewey's later philosophy and that (2) in order to convince the logocentric opposition of the validity of Vihalemm's program, Dewey's ideas provide arguments which ground the latter. I will focus on Dewey's epistemology as he presents it in *The Quest for Certainty* (1929), but the motivation behind it is wider: moral and social philosophy and a reform in education.

Nietzsche kaleidoskoop ehk suure perspektiivi võimalikkusest filosoofias

Ettekandes tahaksin esitada ühe spekulatiivse tõlgenduse Nietzsche meetodi kohta. Väidan, et Nietzsche panus lääne filosoofiasse seisneb perspektiivide pööratavuse avastamises ja selle rakendamises kontseptuaalsele mõtlemisele. Sellelt pööratavuse pinnalt on mõistetav ka tema kriitika õhtumaise dualistliku metafüüsika aadressil: “Metafüüsikute põhjapanevaks usuks on usk väärtuste vastandlikkusesse” (Sealpool head ja kurja, lk 11). Seevastu Nietzsche alushoiak dualismide suhtes väljendub vastandite (ümber)pööramises, mida nimetan tema kaleidoskoopiliseks eripäraks: “mul on kätt, et perspektiive ümber pöörata” (Ecce homo, lk 15). Esitan oma mõttekäigu järgmiste punktide varal: (1) Nietzsche kasutab omadussõna “suur” (gross) sageli talle iseloomulikus tehnilise termini tähenduses, osutamaks et sellega markeeritud mõiste hõlmab ühtlasi ka oma (näilist) vastandit. (2) Nietzsche “suurt jaatust” tuleks mõista topeltjaatusena, kus suur JA jaatab väikest ja’d ehk seda konjunktsiooni, mis toob kokku ja hoiab lahus vastandid, millega mõtlemine opereerib (elu ja surm, keha ja vaim, tõde ja vale, tervis ja haigus, mees ja naine jne). (3) Vasturääkivus ei ole mõtlemise vajak, välistus, mis sätestab mõistuse legitiimsuse tarvituse lõpliku piiri (kehtestades ratsionaalsuse kui sellise); vasturääkivuse mitmesugused ilmingud (paradoks, antinoomia, apooria) on hoopis filosoofiline ajend, mis sunnib mõtlema ja hoiab mõtteprotsessi käigus. (4) Kui Hegeli dialektika neutraliseerib vasturääkivusi n-ö eituse eituse kaudu, siis Nietzsche “kaleidoskoopia” kui jaatuse jaatus kätkeb kahe perspektiivi samaaegset omaksvõttu, parallaksinihet, mille põhjustab “maailma seirav topeltpilk, mis on omane igale suurele teadmisele” (Ecce homo, lk 93). (5) Pööratavuse idee seab õigesse valgusesse ka Nietzsche kuulsa “perspektivismi”, mis ei tähenda sugugi postmodernistlikku anything goes relativismi, vaid kahe ühismõeldutu vaatepunkti kaudu olemise ja mõtlemise, substantsi ja subjekti, reaalsuse ja ratsionaalsuse mitteidentsuse tunnistamist. (6) See tunnetusteoreetiline hoiak implitseerib ka teatava kõlbelise ideaali, zarathustraliku “vastuvõtlikkuse vastuoludele”. Nietzsche eetilise minapilt on seega kenasti kokkuvõetav Walt Whitmani tõdemusega: “Räägin ma endale vastu? / Olgu nii – ma räägin endale vastu, / (olen avar ja paljudes on mu sees.)”

Giulia Lorenzi, giulia.lorenzi.philosophy@hotmail.com

Mirror and echo phenomena: between analogies and disanalogies.

For many centuries, features such as the rapidity and the wealth of information supplied by vision have pulled the interest of many thinkers to the study of the functioning of sight. Indeed, vision is regarded as the primary sense, and there is a vast and multidisciplinary (Palmer 1999) scientific literature on its mechanisms and phenomena. For these reasons, in philosophy, vision has shaped the debates on perception in a general sense. However, recently there has been growing interest in studying the other sense modalities and their interaction: smell (Richardson 2013), touch, hearing (O’Callaghan 2007) and taste have become objects worthy of study in the philosophy of perception. Given that, a common strategy to study phenomena of these different sense modalities, is to postulate analogies with better known visual cases.

In the contemporary debate about the metaphysics of sounds, Casey O’Callaghan (2007; and, more briefly, Casati and Dokic 1994), compares the sound of an echo with a mirror image in defence of his position about the event nature of auditory object. Does the comparison between the echo phenomenon and that of mirroring work properly, however?

The aim of my talk is to investigate more extensively the analogies and dis-analogies between the echo phenomenon and mirroring, considering both metaphysical issues, such as temporal and spatial coordinates, as well as perceptual aspects such as human reactions to and intuitions about them. My starting points will be, on one hand, O'Callaghan's considerations about echoes and, on the other, Steenhagen's (2017) work on mirrors.

This more detailed analysis supports the accuracy of the analogy between echo and mirror phenomena, which is intuitively attractive, and provides a more solid argument in favour of O'Callaghan's general theory of sounds as distal events.

Bibliography:

- Casati R., Dokic J., (1994). *La Philosophie Du Son*, Nîmes, Editions Jacqueline Chambon.
O'Callaghan, C., (2007). *Sounds. A Philosophical Theory*, New York, Oxford University Press.
Palmer, S. E., (1999). *Vision Science: Photons to Phenomenology*, MIT Press, Cambridge.
Richardson, L., (2013). Sniffing and Smelling, *Philosophical Studies*, 162, pp. 401-419.
Steenhagen, M., (2017). False Reflections, *Philosophical Studies*, 174, pp. 1227-1242.

Leo Luks, Eesti Maaülikool, leo.luks@emu.ee

Reaalsuse infotehnoloogilised laiendused — kas üksnes viimse inimese meelelahutus või sild üle inimse poole liikumisel?

Kuigi *üle inimene* on Nietzsche filosoofias äärmiselt keeruline ja segane mõiste, pole kahtlust selles, et Nietzsche propageerib inimese kui sellise ületamist, liikumist üle inimse poole. Küsimus pole valikus — Nietzsche jaoks on humanistlikult määratletud inimene nihilismi kui ajastu kultuurilise loogika tingimustes paratamatult hääbumisele määratud, üle inimene aga on ainus potentsiaalne võimalus uuenemiseks ja nihilismi ületamiseks.

Paljud filosoofid (nt Adorno, Benjamin, Habermas, Debord, Baudrillard) on 20. sajandi vältel kirjeldanud, kuidas meedia tuginev ühiskond (mida hoiab töös nähtamatu võrgustikuna kapital) põhjustab allakäiku nendes võimetes, mida humanistlik traditsioon inimese puhul keksetena väärtustab. Interaktiivse sotsiaalmeedia levik pakub sedasorti allakäigunarratiividele veelgi rohkem ainet. Oleks küllaltki lihtne kirjeldada sotsiaalmeediasse ja nutiseadmetesse neelduvat kaasaja inimest Nietzsche vaimus *dekadendina*, oma elu tühiste naudingutega täitva *viimse inimesena*.

Oma ettekandes küsin, kas infotehnoloogilistes virtuaalreaalsustes leiduvad võimalused inimesele omase iseduse fragmenteerimiseks võiksid olla sillaks üle inimse suunas. Kuigi Nietzschest ammutavad inspiratsiooni mitmed post- ja transhumanistlikud mõtlejad, ei võta ma ettekandes vaatluse ega hindamise alla visioone inimjärgsusest kui hüpoteetilisest kehatust eksistentsist küberruumis. Võtan aluseks tänase olukorra, tuginedes põhimiselt diagnoosile, mille annavad sellele Zygmunt Bauman ja Rein Raud teoses „Iseduse praktikad“. Erinevalt mainitust analüüsin kaasajal ilmnenu infotehnoloogilisi võimalusi lahus humanistlikest moraali-intuitsioonidest, amoraalselt. Uurin võimalusi, mida olukord pakub *eneseloomest esteetikaks* ning skisoidseks, pendeldavaks subjektiivsuseks, kasutades põgusalt Foucault' ja Deleuze'i poolt tehtud edasiarendusi Nietzsche filosoofiale. Kitsamalt küsin, lähtudes nii isiklikust kogemusest kui ka Nietzsche hilisteoste implikatsioonidest, kas ja kuidas võiks just filosoferimine, filosoofielamine olla teekond üle inimse suunas. Kuigi üle inimse mõiste tõlgendamine nõuaks omaette mahukat uurimistööd, on selle mõiste piiritlemine ettekande küsimuse püstituse sihis vältimatu.

Bibliograafia:

- Bauman, Z. (2008). *The Art of Life*, Cambridge/Malden: Polity Press.
Bauman, Z., Raud, R. (2016). *Iseduse praktikad*, Tallinn: TLU Kirjastus.

- Boella, S. (2015). The Nihilist as a Not-Man. An Analysis of Psychological Inhumanity, *Philobiblon*, Vol XX, 1, lk 33-44.
- Brogan, W. A. (1991). The decentered Self: Nietzsche's Transgression of Metaphysical Subjectivity, *The Southern Journal of Philosophy*, Vol XXIX, No 4, lk 419-430.
- Constâncio, J., Branco, M. J. M.; Ryan, B. (toim), (2015). *Nietzsche and the Problem of Subjectivity*. Berlin/Boston: de Gruyter.
- Luks, L. (2015). *Nihilism ja kirjandus*, Tallinn: TLU Kirjastus.
- Luks, L. (2016). Nietzsche moraalikriitilise hoiakuga seotud tõlgendusprobleemid, *Akadeemia*, nr 12, lk 2174-2198.
- Luks, L.; Moor, A. (2016). Family Resemblances Between Contemporary Post-Subjectivist Philosophy and Animistic Religions, *Problemos*, no 89, lk 48-61.
- Mellamphy, D., Mellamphy, N.-B. (2016). *The Digital Dionysus: Nietzsche and the Network-Centric Condition*, CTM Documents Initiative.
- Nietzsche, F. (2004). Fragmente nihilismist. - *Akadeemia*, nr. 8, lk. 1793-1832.
- Nietzsche, F. (2006). *Nõnda kõneles Zarathustra*, Tallinn: Olion.
- Nietzsche, F. (2015). *Moraali genealoogiast. Vaidluskiri*, Tallinn: Varrak.
- Nietzsche, F. (2017). *Sealpool head ja kurja. Tulevikufilosoofia eelmäng*, Tartu: Ilmamaa.
- Pearson, K.-A. (1999). Loving the Poison: On the „Meaning“ of the Transhuman Condition, in ed. J. Lippitt, *Nietzsche's Futures*. Hampshire: Macmillan, lk 189-206.
- Ransom, J.-S. (1997). *Foucault's Discipline. The Politics of Subjectivity*, Durham/London: Duke University Press.
- Sedgwick, P. R. (2007). *Nietzsche's Economy. Modernity, Normativity and Futurity*, Hampshire/New York: Palgrave.
- Zimmerman, M. E. (2011). Last Man or Overman? Transhuman Appropriations of a Nietzschean Theme, *The Hedgehog Review: Critical Reflections on Contemporary Culture*, Volume 13, No. 2, lk 31-44.

Endla Lõhkivi, Tartu Ülikool, endla.lohkivi@ut.ee

Infotehnoloogia interdistsiplinaarses teadustöös: arusaamise loomisest eri distsipliinide vahel

Infotehnoloogia on järjest enam kasutusel teadusliku uurimistöo loomuliku osana, alates andmete kogumisest, salvestamisest, töötlemisest kuni analüüsi, modelleerimise ja arvutisimulatsioonideni, rakenduste ja teenusteni, mis teadustöö alusel luuakse. See eeldab uurimistöös erinevate pädevuste ühendamist ehk interdistsiplinaarsust. Üheks valdkonnaks, kus infotehnoloogia pikka aega on rakendust leidnud, on keeletehnoloogia ja arvutilingvistika, näiteks kõnetuvastussüsteemide loomine. Lihtsustades võib sellise uurimistöo osapooli nimetada keeleinimesteks ja arvutiinimesteks. Tulemuslik koostöö eeldab, et distsipliinide esindajad üksteisest aru saaksid. Mida see tähendab? Kas arvutiinimesed peaksid õppima keeleteadust ja vastupidi? Mil määral peaksid ühe distsipliini esindajad omandama teise distsipliini teadmisi ja oskusi?

Oma ettekandes käsitlen arusaamise mõistet interdistsiplinaarse teaduspraktika näitel. Arusaamiseks ei piisa ühisest sõnavarast, vaid enamal või vähemal määral peavad kattuma ülesanded ja nende lahendamise viisid. Vaatlen, milliseks kujunevad teadlaste valikud, kui kattumine on väiksem või suurem.

Ettekanne põhineb kahe ETAGi personaalse uurimistoetuse projekti, PUT 732 ja PRG 462, raames kogutud empiirilisel materjalil.

Elektrooniline informeeritud nõusolek inimuuringutes: otsides tasakaalupunkti isikautonoomia austamise ja paternalismi vahel

Arengud infotehnoloogias mõjutavad seda, kuidas inimuuringuid tehakse. Näiteks on muutunud viis, kuidas võimalikku uuringus osalejat informeeritakse ning temalt nõusolekut küsitakse. Kui varasemalt leidis informeerimine aset vahetult uurija ja võimaliku uuritava vahel ning protsessi lõpus andis uuritav allkirja paberil olevale dokumendile, siis nüüd võib kõik aset leida elektrooniliste kanalite kaudu. Viimane on ka Eestis praktiseerimist leidnud, nt eelmisel aastal kasutas seda TÜ Eesti Geenivaramu, värvates aastaga sadatuhat uut geenidoonorit.

Inimuuringute üks peamisi printsiipe on olnud vabatahtlik osalemine. On leitud, et selleks, et potentsiaalne uuritav saaks osalemise üle otsustada, tuleb teda uuringuga seotud asjaoludest informeerida. Üksnes teavitamisest ei piisa, oluline on veenduda selles, et inimene saab talle esitatud infost aru. See tähendab ka selle mõistmist, et pakutakse osalemist uurimistöös ja nt mitte personaalset ravi või mõnda muud teenust.

Informeeritus on vajalik selleks, et inimese saaks langetada autonoomse enesekohase otsuse selle kohta, mis toimub tema keha või tema isikuandmetega. Vahetu kontakt uurija ja võimaliku uuritava vahel võimaldas suhteliselt lihtsalt kindlaks teha, kuivõrd infost on aru saadud.

Ettekandes võtan vaatluse alla väljakutsed, mida esitab elektroonilise informeeritud nõusoleku kasutusele võtmine. Teen seda tuginedes näitele, kus uuritav annab elektroonilise nõusoleku ilma et ta oleks tutvunud uuringu kohta käiva infoga.

Uurin, kas praegune inimuuringute eetilise raamistik annab rahuldava vastuse kuidas sellises olukorras toimida või tuleks uute tehniliste lahenduste valguses asjad ümber vaadata? Kas tuleks kasutada tehnoloogilisi lahendusi, mis sunniksid inimest uuringu kohta käivat infot läbi lugema? Sõna otseses mõttes lugema, mitte vaid linnukest panema? Näiteks kas võimalik uuritav peaks pärast informeerimist ja enne allkirjastamist täitma mingi testi või vabas vormis mingitele küsimustele vastama või oleks see liiga paternalistlik? Või peaksime kaalutletud otsust, infot mitte lugeda, austama?

Ettekandes näitan, et vastus eeltoodud küsimustele võib olla erinev, sõltuvalt sellest, millisest autonoomia teooriast lähtuda.

Kant on the rule of existential generalization

The subject of my presentation is Kant's claim in the Critique of the Pure Reason that existence is not a real predicate. My aim is twofold. On the one hand, I wish to argue that Kant's claim amount to an implicit endorsement of the rule of existential generalization (EG); while, on the other hand, I wish to critically assess Kant's reasons for supporting EG. In particular, I wish to argue that some of Kant's best known arguments in support of the claim that existence is not a real predicate deserve further scrutiny.

Luis Fernández Moreno, Complutense University of Madrid, luis.fernandez@filos.ucm.es

Considerations on the Nature of Theoretical Identifications

In the third lecture of *Naming and Necessity* Kripke focuses on natural kind terms, and he claims that one of the similarities between them and proper names is that both appear in identity statements that, if true, are necessary. Kripke calls this type of identity statements containing natural kind terms theoretical identifications. In my paper I will assume the interpretation of them that is most plausibly Kripke's view, according to which theoretical identifications are identity statements and natural kind terms are general terms. Nevertheless, Kripke asserts that the said similarity follows from another, namely that proper names and natural kind terms are rigid designators.

The aim of my talk is twofold. Firstly, I contend that it can be held that natural kind terms are rigid designators, whose referents are universals. Secondly, I argue that the conception of natural kind terms as rigid designators of universals hinders the establishment of the truth of theoretical identities and hence of their necessity.

Bibliography:

Kripke, S. (1980). *Naming and Necessity*. Oxford: Blackwell. (Revised and enlarged edition, first published in D. Davidson and G. Harman (eds.), *Semantics of Natural Language*. Dordrecht: Reidel, 1972).

Bruno Mölder, Tartu Ülikool, bruno.moelder@ut.ee

Mõistuse üleslaadimine ja interpretivism

Filosoofid ja futuroloogid on kaalunud võimalust, et tulevikus suudetakse laadida inimese mõistus või vaim üles arvutisse ning et sealt saab vajadusel selle alla laadida uude kehasse. Mõistuse üleslaadimise võimalikkuse hüpotees laseb mõnedel traditsioonilistel filosoofilistel probleemidel paista uues valguses. Iseäranis tõuseb üles küsimus sellest, mis on inimõistuse jaoks tarvilik ning milline vaimukäsitlus sobib kokku üleslaadimise võimalikkusega. Näiteks funktsionalism tundub üleslaetud mõistuse jaoks kõige sobivam teooria ning neid teooriaid, mis rõhutavad vaimu olemuslikku ihulisust, on raske kui mitte võimatu ühitada üleslaadimise võimalikkusega. Antud ettekandes vaatlen lähemalt, kuidas paistab mõistuse üleslaadimine välja interpretivistlikust vaatevinklist, mille järgi on mentaalsus olemuslikult tõlgendusest sõltuv. Kas interpretivism sobib käsitada sellist vaimu või mõistust, mida on võimalik arvutisse üles laadida või tekivad siin põhimõtteliselt raskused?

Peeter Mürsepp, Tallinna Tehnikaülikool, peeter.muursepp@taltech.ee

Haavatava maailma hüpotees

Simulatsiooniargumendi looja, Rootsi päritolu Briti filosoof Nick Bostrom esitas hiljuti uue intriigeriva lähenemise, mida ta nimetab haavatavushüpoteesiks, mis ütleb järgmist. Kui tehnoloogia areng jätkub, siis omandab inimkond kunagi võimekuse, mis väga tõenäoliselt tsivilisatsiooni hävitab, kui just tsivilisatsioon ei eemaldu pool-anarhisest seisundist. Viimase all mõistab Bostrom järgmist:

- 1) korrakaitse võimekuse piiratus ennetada valdava enamuse poolt taunitavaid aktsioone;
- 2) globaalse administratiivse võimekuse piiratus;
- 3) inimeste motivatsioonide paljusus.

Tsivilisatsiooni ähvardavate põhiohtude kirjeldamisel jääb Bostrom tavapäraselt teadaolevate juurde nagu tuumaohut või kliimakatastroof.

Olukorda stabiliseerivateks peab ta järgmisi teoreetilisi võimalusi:

tehnoloogia arengu pidurdamine;
tegutsemismotiivide paljususe vältimine;
eriti efektiivse ennetustegevuse juurutamine;
efektiivse globaalse administreerimise rakendamine.

Esimene abinõu ilmselt garanteeritud tulemust anda ei saa, sest Bostromi enese sõnutsi on hävitusliku potentsiaaliga tehnoloogiad juba olemas. Inimeste tegutsemismotivatsioonide reguleerimine näib samuti üsna lootusetu ülesandena. Kolmanda moodusega seoses pakub Bostrom välja mõtteeksperimenti, mida ta nimetab Hi-Tech Panopticumiks. Terminoloogiline seos George Orwelli loominguga on siinkohal tahtlik. Tegemist oleks üleüldise jälgitavuse süsteemiga. Mõneti üllatavalt ei pruugi ülemaailmse jälgitavuse rakendamine rahaliselt ülesamatult kallis olla. Küll aga kaasnevad muidugi muud probleemid, millest põhilised on juba Orwell omal ajal välja toonud. Edukas globaalne administreerimine näib Bostromi jaoks kõige sobilikuma lahendusena. Selle rakendamise korral pole tingimata vaja tehnoloogia arengut pidurdada, mis niikuinii suhteliselt lootusetu ülesanne. Eriti suuri lootusi paneb Bostrom efektiivse administreerimise seostamisele tulemusliku ennetustegevusega. Viimane poleks ilma esimeseta ju tegelikult võimalikki.

Taalise üleüldise kontrolli ja tegutsemisvalmiduse kujundamine tõstatab loomulikult tõsised poliitilised probleemid. Tegemist on ju totalitarismiks valmisoleku (turnkey totalitarianism) tekitamisega. Ohtlik olukord tekib tahes tahtmata, sõltumata sellest, kas keegi kavatseb võtit keerata või mitte. Bostrom näeb siiski lahendusvõimalust, arvates et olukorra võib päästa nn struktureeritud läbipaistvuse rakendamise abil, mis hoiab ära võimaluse vaikseks ja salajaseks võimu koondamiseks. Avalikkuses püsimine võib tõesti totalitarismi suunas liikumise ära hoida, kuid vaevast saab siin ikkagi täit garantiid olla.

Rita Niineste, Tallinna Ülikool, rita.niineste@gmail.com

Feministlik filosoofia ja ühiskondlik debatt

Feministliku teooria üks käibefraase on olnud, et isiklik on poliitiline, kuid feministlikus uurimistöös ei iseloomusta eetilise ja poliitilise mõõde mitte ainult naiste isikliku elu ja kogemuse uurimist, vaid praktiliselt kõiki aruteluteemasid. Milliseid piiranguid seab või võimalusi avab see olukord konkreetsemalt feministlikule filosoofiale, mis nime poolest peaks olema võimeline sõnastama piisavalt abstraktseid ja võimalikult suure üldistusjõuga mõttearendusi selle kohta, milline paistab maailm läbi naise kogemuse, kuidas analüüsida soolist erinevust kui filosoofilist probleemi jms? Kas feministlikul filosoofial on midagi pakkuda näiteks inimestele, kes tunnevad, et muutunud soorollid ja teisenenud ühiskondlikud konventsioonid võivad pikas perspektiivis küll ühiskondlikku vabadust suurendada, kuid esialgu on selle hinnaks kohmakas ja mõttevabaduse seisukohalt küsitav korrektsusparadigma? Kas kandepinda koguv posthumanistlik vaade võiks anda feministlikule filosoofiale uusi vahendeid nii sisemiseks arenguks kui ka ühiskondlikus debatis kaasa rääkimiseks?

Bibliograafia:

- Ferrando, F. (2014). Is the post-human a post-woman? Cyborgs, robots, artificial intelligence and the futures of gender: a case study. *EUROPEAN JOURNAL OF FUTURES RESEARCH*, 2(1).
- Fielding, Helen A and Olkowski, Dorothea E. (2017). Part 1. The Future Is Now. In Helen A. Fielding and Dorothea E. Olkowski, eds., *Feminist Phenomenology Futures*. Bloomington: Indiana University Press.
- Grosz, E. (2006). The Force of Sexual Difference. In: Mortensen, E. (ed.) (2006). *Sex, breath, and force. Sexual difference in a post-feminist era*. Lexington Books.
- Grosz, E. (2011). *Becoming Undone: Darwinian Reflections on Life, Politics, and Art*. Duke University Press.
- Tuana, N. (2004). Coming to Understand: Orgasm and the Epistemology of Ignorance. *Hypatia*, 19 (1), 194–232.

Takashi Oki, Nagoya University, Yale University, takashi.oki.new@gmail.com

Aristotle on Deliberation and Necessitarianism

The problem of compatibility/incompatibility between the possibility of meaningful deliberation and necessitarianism (the view that everything happens of necessity) has long been a topic of discussion, and it is well known that Aristotle is concerned with the problem in *De Interpretatione* 9. He thinks that if everything happens of necessity (18b30-31), then ‘there would be no need to deliberate or to take trouble, thinking that if we do this, this will happen, but if we do not, it will not’ (18b31-33). In this paper, I argue that Aristotle is a deliberation incompatibilist, and consider why he thinks that it is reasonable to endorse this position.

First, while interpreting his presentation of the necessitarian argument (18a34-18b16) as a *reductio* (*pace* Bobzien 2011; Nielsen 2011), I show that Aristotle thinks that deliberation is inefficacious if the future is necessary (in the sense of being fixed/irrevocable) in the way the past and present are.

Second, I argue that the necessitarian conclusion that ‘everything is and happens of necessity’ (18b30-31), which is considered to be incompatible with deliberation, should be distinguished from the view that everything that happens happens of necessity, independently of antecedent conditions (*pace* Nielsen 2011). Further, I also show that Aristotle’s argument on the inefficaciousness of deliberation is not a sort of ‘Lazy Argument’ (*pace* Sorabji 1980).

Third, I argue that, in Aristotle’s view, the principle that ‘if we do this, this will happen, but if we do not, it will not’ (18b32-33) would hold even if everything happens of necessity. By pointing out that Aristotle accepts that one could still ‘causally affect some future events’ even if everything happens of necessity, I show that Fine’s (1984) contention that ‘since one can causally affect some future events, one can deliberate about them’ misses the point.

Bibliography:

- Bobzien, S. (2011). ‘Afterword’ to *the Philosophy of Aristotle*, Signet/Penguin.
- Fine, G. (1984). Truth and Necessity in *De Interpretatione* 9. *History of Philosophy Quarterly*, 1-1: 23-47.
- Nielsen, K. M. (2011). Deliberation as Inquiry: Aristotle’s Alternative to the Presumption of Open Alternatives. *Philosophical Review*, 120-3, pp. 383-412.
- Sorabji, R. (1980). *Necessity, Cause, and Blame*. Cornell University Press.

Hume and the Guise of the Bad

‘tis observable that an opposition of passions commonly causes a new emotion in the spirits, and produces more disorder, than the concurrence of any two affections of equal force. This new emotion is easily converted into the predominant passion, and increases its violence, beyond the pitch it wou’d have arriv’d at had it met with no opposition. Hence we naturally desire what is forbid, and take a pleasure in performing actions, merely because they are unlawful. The notion of duty, when opposite to the passions, is seldom able to overcome them; and when it fails of that effect, is apt rather to encrease them, by producing an opposition in our motives and principles. (T 2.3.4.5)

Hume scholars so far seem to have paid little attention to this passage. But, as I will try to show, its analysis bears fruit for both Humean scholarship and contemporary moral psychology. First, it needs to be understood why the phenomenon pointed out—desiring and taking pleasure in performing certain actions merely because they are unlawful—poses a challenge of sorts to Hume’s theory of evaluation (sections 3-4). Second, I propose a solution of the challenge which draws on Hume’s treatment of malice, and illuminates the role played by comparison and the self in these apparently paradoxical passions (sec. 5). Third, the solution I attribute to Hume should stand out in contemporary discussions of desiring something “under the guise of the bad” as a particularly promising account of this phenomenon (sec. 6).

Ühismõiste ja vastuaktualisatsioon

Püüan näidata, kuidas Spinozalik ühismõistete (*notiones communes*) loomine, eriti Deleuze’i tõlgenduses, on seotud sellega, mida Deleuze nimetab vastu-aktualisatsiooniks ehk vastu-efektuaatsiooniks. Ühismõiste tähendab seda, et ma õpin sättima oma kohtumisi, saan teada, kuidas väliskeha või keskkond minu kehaga sobib. Deleuze toob näiteks ujuma õppimise. Edukas ujuma õppimine tähendab seda, et on vastuaktualiseeritud varasem olukord, see nii-öelda üles sulatatud, ning ühendatud ümbrusega nüansseeritumalt, täpsemalt, kohasemalt. Aktualiseerumine ehk kättejõudmine tähendab, et kättejõudnut vastu-aktualiseeritaks, lastaks käest. Ühtlasi seostan seda näidetega mittelääne pärimustest.

Defending Christine Korsgaard’s account on Extrinsic Final Value

It is commonplace in ethics to distinguish between instrumental and intrinsic value. Objects, activities, and states of affairs have instrumental value if their goodness depends on the value of some further valuable thing. They have intrinsic or final value when the entities in question are valuable in itself or for their own sake. This distinction between instrumental and intrinsic value is challenged by Christine Korsgaard (1983), who distinguishes between two pairs of value instead of one. Instrumental and final value relate to how persons value things, either as means or for their own sake. Extrinsic and intrinsic value relate to the ontological origin or “source” of value. An object’s intrinsic value supervenes on its intrinsic or non-relational properties; extrinsic value supervenes on the object’s extrinsic or relational properties. The distinction between intrinsic and

final value enables Korsgaard to claim that the external value source of almost all finally valuable things lies in the agent's act of valuing those things.

For example, a beautiful painting can be valued for its own sake and not because it is an instrument to gain pleasure. If the painting is, however, locked in a closet and no one could ever be delighted by it, the painting would not be valuable. Wlodek Rabinowicz and Toni Rønnow-Rasmussen (1999) oppose Korsgaard's explanation of extrinsic final value. According to them, the agent's act of valuing an object is not a relational property in the supervenience base of final value. It belongs rather to the "constitutive ground" of value. Thus, even if the painting final value is externally constituted by the agent's act of valuing, the painting could still be intrinsically valuable as long as the painting's value supervenes on its intrinsic properties of having, say, certain colors and proportions. Rabinowicz and Rønnow-Rasmussen provide an argument from modality in defense of their criticism. I argue that their argument fails to establish that Korsgaard's explanation of extrinsic final value is false.

Bibliography:

Korsgaard, C. (1983). Two Distinctions in Goodness, *Philosophical Review* 92 (2): pp. 169-195.

Rabinowicz, W., Rønnow-Rasmussen, T. (1999). A Distinction in Value: Intrinsic and For Its Own Sake, *Proceedings of the Aristotelian Society* 100 (1): pp. 33–51.

Fabrice Pataut, *Sciences, Normes, Démocratie*, Fabrice.Pataut@univ-paris1.fr

Indispensability and Ontology

It is often taken for granted that if it is shown that mathematics is indispensable for our best explanations of natural phenomena, existence claims of applied mathematics thereby acquire a bona fide philosophical justification. Nominalizations have been proposed to substantiate the opposite dispensability claim. Questionable assumptions are at work in the way in which both platonists and nominalists have construed the relation between inferences to the best explanation, ontological commitments and indispensability *vs.* dispensability claims, assumptions about why we should prefer a certain kind of explanation, either mathematical or not, and why the success of either kind plays the role of a defeater or underminer with respect to our beliefs in whatever plays the (alleged) role of explanans in the rejected explanation.

I look at two nominalistic proposals, by Field (*Field 1980*) and Rizza (*Rizza 2011*). In Field's case, the arithmetic of natural numbers supplemented by set theory is replaced by the logic of identity supplemented by existential numerical quantifiers to derive a conclusion about the existence of a number of concrete objects located in space. In Rizza's case, a number-theoretic lemma about the co-primeness of natural numbers is replaced by a paraphrase resorting to betweenness on temporal intervals to explain the evolutionary advantage of the behavior of cicadas. Building on criticisms offered in *Pataut 2016*, I argue that in each case the nominalistic force of the non mathematical explanation should be assessed in terms of the explanatory role of nominalistically respectable proofs rather than in terms of ontological economy proper.

Bibliography:

Field, H. (1980). *Science without Numbers - A Defence of Nominalism*, 1st edition, Princeton UP, Princeton, New Jersey.

Pataut, F. (2016). Comments on "Parsimony and inference to the best mathematical explanation", *Synthese* 193 (2), Special Issue on *Indispensability and Explanation*, pp. 351-363.

Rizza, D. (2011). Magicicada, mathematical explanation and mathematical realism, *Erkenntnis* 74, pp. 101-114.

Alex Pazaitis, Tallinn University of Technology, alexandros.pazaitis@taltech.ee

States and Cooperatives: Perspectives from commons-based peer production

This article interrogates the notion of the "Partner State", which is held to represent a new form of symbiosis between State and civil society, based on the principles and practices of commons-based peer production (CBPP). There are multiple historical examples of state policies supporting alternative or emergent economic structures, from the Israeli kibbutz movement and Italian cooperatives, to the contemporary policies for the social and solidarity economy. But is it possible for the contemporary State to support CBPP? Recent experimentations such as Ecuador's Open Society project, and the urban commons policies in cities like Barcelona, Madrid, Bologna and Naples speak for this potential.

Similarly, CBPP has been discussed as an opportunity to revitalize cooperativism in the digital economy. More specifically, the concept of "open cooperativism" introduces a tentative synergy between the cooperative principles, on one hand, and the practices emerging in CBPP, on the other. Open cooperatives provide an analytical basis for State theory to approach CBPP, as well as a pre-figurative blueprint for the Partner State. CBPP entails a learning process of bottom-up political action and influence to forward State reform. Topical issues on the legitimacy of power and democracy, both in the political and the economic sphere, are informed through the commons and contributory activity.

Marika Proover, Tallinn University of Technology, marika.proover@taltech.ee

Uncertainty and attitudes toward technology

Technology have been developing in ever growing pace taking forms, directions and functions hardly imagined few centuries or even few decades ago. Becoming prevailing in our everyday life and shaping our environment in unconceivable ways technology has evoked strong feelings toward itself whether we consider laypeople, policy makers or academics.

In this talk, the connection between attitudes toward modern technology and uncertainty will be explored. Different sources of uncertainty can be distinguished for this purpose: (i) epistemic uncertainty that rises from lack of knowledge and understanding, (ii) uncertainty stemming from abductive reasoning that is most suited to think about the future and (iii) ontological uncertainty characteristic to future.

The focus will be on the latter. Future falls into the category of the fundamental uncertainty and this uncertainty is by its nature ontological. As such, common ways of coping with uncertainty, such as acquiring information, knowledge or better understanding about phenomena, tend to fail whenever future is a crucial parameter. Further, our attitudes toward (modern) technology and technological change can be partially rooted in the ontological uncertainty characteristic to future. Technology and technological change in themselves would not be the cause for optimistic or pessimistic attitudes, rather they would underline and stress uncertainty involved.

Title: Chinese-whispers in the contemporary social world

In this paper, I theorise the ways in which changes to communication culture brought about by new forms of online communication have led to an increasing homogenisation of experiences/rhetoric in the cultural sphere. To illustrate this, I pay particular attention to the global spread of, so-called, artisanal food movements. The emphasis of the paper is to analyse how the process by which new concepts and meanings are tested and challenged is derailed online, undermining the legitimacy of the concepts as online communication influences the immediate world irrespective of its content.

Following Habermas, we can understand intersubjective communication in terms of the communication rules that it follows (i.e. rules of intelligibility, truth, truthfulness and rightness) (Habermas 1984; 1990). This system of communication is mimicked online, as the individual is able to develop an online placeholder for their subjectivity, an online-subject, who is available for online communication in the absence of the individual. Due to the seeming equivalence of communication online to offline, the online-subject is expected to follow the same communication rules as in the immediate world. However, intersubjective communication on social media is structured in line with the economic interests of the platform providers (Dean 2005; Fuchs 2017). As a result of this, communication enacted between online-subjects does not follow the same communication rules as immediate communication. Rather, individuals via online-subjects are encouraged to use communication instrumentally with the aim of gaining visibility/recognition through their communication (Andrejevic 2013; Fuchs 2016).

What this evolution in communication culture means is that new concepts and ideas gain uptake in the social world more generally, not as a result of agreement in discourse, but because of superficially appealing features of those concepts, or their expression, that attract attention and encourage replication irrespective of their full content (for similar discourse, see: Wendling 2018; Vaidhyanathan 2018)

Bibliography:

- Andrejevic, M. (2013). Estranged free labor, in Schulz, T. (ed.) *Digital labor: the Internet as playground and factory*, New York: Routledge, pp. 149-164.
- Dean, J. (2005). Communicative Capitalism: Circulation and the Foreclosure of Politics, *Cultural Politics*, 1(1), pp. 51-74.
- Fuchs, C. (2016). *Critical Theory of Communication: New Readings of Lukács, Adorno, Marcuse, Honneth and Habermas in the Age of the Internet*, London: University of Westminster. Available at: <http://fuchs.uti.at/books/critical-theory-ofcommunication/> (Downloaded 9.01.2017).
- Fuchs, C. (2017). *Social Media. A Critical Introduction*, 2nd edition, London: Sage Publications Ltd.
- Habermas, J. (1984 [1981]). *The Theory of Communicative action, Vol. 1: Reason and the Rationalisation of Society*, Translated by T. McCarthy. Boston: Beacon Press.
- Habermas, J. (1990 [1983]). *Moral Consciousness and Communicative Action*, Translated by C. Lenhardt & S. W. Nicholsen, Cambridge: Polity Press.
- Vaidhyanathan, S. (2018). *Antisocial Media. How Facebook Disconnects Us and Undermines Democracy*, Oxford: Oxford University Press.
- Wendling, M. (2018). *Alt Right: From 4-Chan to the White House*, London: Pluto Press.

A.I. Stalin – kurjuse algoritmiline anatoomia

Mh E. Radzinski *Stalini*, N. Bostromi *Superintelligence* (S.I.) ainetel.

Eelanalüüs

Marx ja Engels olid *kommunitaarse programmeerimise* (KP) arhitektid. *KP Manifestis* kirjeldasid nad kapitalistliku ühiskondlik-majandusliku formatsiooni (ÜMF) viletsust ja lähteülesannet: kommunistlikku maailmarevolutsiooni ja kapitalita ÜMF-i.

Eeldus: ÜMF on revolutsiooniliselt (ümber)programmeeritav ja seejärel plaanipäraselt administreeritav.

Arenduse lähteülesanne

Uue ÜMF-i peaarhitektid modifitseerisid lähteülesande kättesaadavale, ent algelisele arenduskeskkonnale – Vene Impeeriumile. Koostati deklaratiivide ja imperatiivide kataloog ning programmeeriti iseõppivad (*oportunistlikud*) revolutsioonilised performatiivid ja algoritmid.

Programm installeeriti veebruarist oktoobrini 1917. Oktoobripöördega käivitati.

Programmeerimise semantika, süntaktika ja pragmaatika

Eesmärgid: a) välised metadeklaratiivid, b) sisemised tegelikult soorituvad performatiivid, imperatiivid ja välistavad disjunktid (*mina või keegi teine*) – SIStalin kui kõrgeim eesmärk.

Proletariaadi diktatuur – permanentne klassivõitlus ja totaalse despootia printsüüp.

Vale ja vägivald – eesmärkide nimel ülevalt alla piiranguteta adapteeritav vahend.

Süsteemi arhitektuur ja riistavara

Püramiidskeem: „Nõukogude rahvas“ < „Nõukogude võim“ < Partei ja tema relvastatud salk NKVD < (Partei) nomenklatuur < Keskkomitee < Poliitbüroo < S. I. Stalin >

Piirangute piirang: *seadusi ja lubadusi tuleb täita* ei kehti riigipoliitika ja suverääni tasemel.

Riigikapitalism. Inimsümbiootilised tootmis- ja sõjamasinad.

Inimagentide liigid: tegusatena asendamatud, asendatavad, massita ebavajalikud, kahjulikud.

Redutseeritud *põhiemotiivid*.

Viha: deklaratiivne, performatiivne, iteratiivne. **Hirm:** performatiivne, mittedeklaratiivne. Vajalik:

a) vastukodeerimise ärahoidmiseks, b) algoritmikäskude täitmiseks sisemotivatsiooni puuduses.

Viha ja hirmu resonants...

Rõõm: kvaasiõnn. **Kurbus:** rõõmsa tuleviku vaenlane.

Masinõpe AI-st – SI-ni

Prototüübid: Lenin ja Trotski. Lenin nõudis viimases juhendis Stalini ja Trotski isoleerimist *sandboxi*, kuid Stalin saavutas adminnide Z&K&B abil Trotski deinstallaerimise. Riik piirati tule müüri ja totaalse viirustõrjega.

Teosti Koba: inkassaator, kohalik admin. **Oraakel Stalin:** ideoloogiline kloon, kohtvõrgu admin.

Geenius (Džinn) I sekretär: põhivõrgu admin. **Suverään Peremees:** süsteemi peaadministraator.

Konkurendid ja kloonid

Arhitekt L. Trotski, A.I. Hitler, Mao Zedong...

Järeldused

Isaac Asimovi robotika 1. seadus on (sügavalt) defektne.

Kurja (ülibanaalse) algoritmistiku loomiseks ei pea programmeerima kurjust.

Suurim kurjus ilmneb siis, kui SI kasutab üksteise vastu inimesi.

Inner Conflict in Plato's *Protagoras* and *Republic*

Traditionally, Plato's *Protagoras* has been taken to defend a theory that stands in sharp contrast to Plato's *Republic*. In the *Republic*, Socrates speaks of the tripartite division of the soul into a rational, an appetitive and a spirited part, and acknowledges the possibility of inner conflict: while the rational part of the soul may urge the agent to act a certain way, the appetites often push in a different direction. One may know, for example that having a fourth drink before driving back home is a bad idea, yet one may still act against one's better judgment and have that drink. In the *Protagoras*, on the other hand, Socrates argues that, when one knows what is best, one always acts in accordance with that knowledge. Knowledge, he insists, is never overcome by the desire for pleasure or other non-rational forces beyond the agent's control.

Scholars usually explain this difference between the *Protagoras* and the *Republic* by arguing that the former is an earlier Platonic work, written at a time when Plato was still under Socrates' influence, while the *Republic* is a work of the mature Plato who has come to see the difficulties of the earlier Socratic view.

In the first part of my paper I propose that inner conflict in fact arises when we are unable to determine *conclusively* the best among competing goods, rather than when we know the best option but are overcome by our desires and emotions. In the second part, I argue that such an understanding of inner conflict in the *Protagoras* is compatible with the *Republic*: the desire for pleasure is real, but it cannot overcome the agent when the best option between competing goods is amply clear.

Jaanus Sooväli, Tartu Ülikool, jaanus.soovali@ut.ee

Indogermaanlus, soomeugrilus ja lingvistiline relativism: Friedrich Nietzschest Uku Masinguni

Lingvistiline relativism või, maakeelsemalt, keelerelativism või keeleline relatiivsus, mida tuntakse ka Sapir-Whorfi hüpoteesina, väidab, et keele struktuur (ja sõnavara) ühel või teisel viisil, suuremal või vähemal määral mõjutab selle kõnelejate mõtlemist ja maailmamõistmist, millest omakorda järgneb, et erinevad keeled tingivad, implitseerivad, toovad endaga kaasa erisuguseid maailmakäsitusi. Mida sarnasemad mingid keeled on, seda suuremad on ka sarnasused nende keelte kõnelejate mõtlemises ja maailmamõistmises, mida erinevamad on keeled, seda suuremad on maailma- ja enesetõlgenduslikud erinevused. Oma ettekandes vaatlen esmalt lingvistilise relativismi tekketingimusi 18. ja 19. sajandi lingvistikas ja filosoofias ning seejärel eritlen selle erinevaid aspekte läbi Friedrich Nietzsche ja Uku Masingu filosoofiate prisma.

Henrik Sova, Tartu Ülikool, henrik.sova@gmail.com

Ülbus kosmilisel skaalal

Oma ettekandes soovin tasuda juba unustatud võla ja vastata argumentidele seoses ühe ammuse vaidlusega, mis toimus Sirbi veergudel minu ja prof Bruno Mölderli vahel. Kirjutasin Thomas Nageli filosoofia üle polemiseerides nii: „Nii jõuame Nageli esitatuga vastanduvale seisukohale, et meie mõtlemis- ja väljenduspiir on ka meie maailma piir. Ja tegelikult ka kogu maailma piir, sest

meie maailm on kõik, mis meil on. Kui meile näib, et saame siiski tähenduslikult rääkida maailmast väljaspool „meie vaimu haardeulatust“, siis on see ikkagi kas meie vaimu haaratud või on tegemist hoopis millegi muuga, kui esmapilgul paistab, nt tundeväljendusega, veidra keelekasutuse või suisa tähendusetusega.“ (Sirp 12.05.2017 „Meie vaatepunkti mõeldamatus“). Niisiis, arvasin toona ja arvan nüüdki ilmselt üsna provokatiivselt, et mõeldamatut pole olemas.

Prof Mölder toob sama Sirbi numbris vastuargumente. Ta leiab igati tasakaalukalt, et meie mõistuse ülesehitus määrab ära, millest suudame mõelda, ja „arvata, et see hõlmabki kõike olevat, on tõepoolest ülbis kosmilisel skaalal“. Ta esitab kokku kolm vastuargumenti. Mõõnan, et Prof Mölderi vastulaused näivad mõistlikud. Sellegipoolest püüan ma tema argumentidele vastu vaielda, õõnestada taas tema sõnastatud realistlikke intuitsioone toetudes seejuures Richard Rorty krestomaatilise oopuse „Philosophy and the Mirror of Nature“ viimasele peatükile.

Bibliograafia:

Mölder, B. (2017). Thomas Nageli mõistlik antihumanism, *Sirp*, 12.05.2017

Rorty, R. (1979). *Philosophy and the Mirror of Nature*, Princeton University Press

Sova, H. (2017). Meie vaatepunkti mõeldamatus, *Sirp*, 12.05.2017

Athamos Stradis, King's College London, athamos.stradis@kcl.ac.uk

The Unbridged Gap between Entropy and Memory

Why do we know more about the past than the future? One natural explanation of this ‘epistemic arrow’ of time is the fact that we have records of the past but not the future. Being obviously physical in nature, we can expect the time-asymmetry of records to be grounded in a yet more fundamental time-asymmetry (as the adage goes, ‘no asymmetry in, no asymmetry out’). The idea that the thermodynamic arrow (whereby entropy always increases) can provide this footing is a popular view in philosophy and a received view in physics. In this talk I explore a salient account in this vein, the ‘Computational Theory’, arguing it fails to achieve this goal.

The Computational Theory is based on Landauer’s Principle, which consists of two claims. First, any realistic computer (including humans) must implement logical erasure. Second, logical erasure necessarily invokes entropic increase in the computer’s surroundings. Hence, Landauer’s Principle is used to explain the epistemic arrow: our experience of an epistemic arrow is a product of our brain’s computations, and since these necessarily align with the thermodynamic arrow, the latter ultimately grounds the epistemic arrow.

This account, however, faces two obstacles. First, since it only characterises computational systems, it does not explain why many external records (fossils, photos, etc.), which do not constitute computational systems, tell us much more about the past than the future. Second, this account takes it for granted that the information encoded on a computer is more informative of the world’s lower-entropy states (which lie in the past) than its higher entropy states (which lie in the future), and so fails to connect the epistemic and computational arrows. Whatever the merits of Landauer’s Principle, it currently does not explain the epistemic arrow of time.

Attila Tanyi, University of Tromsø, attila.tanyi@uit.no

Stephen McLeod, University of Liverpool

What Are Basic Liberties?

We have two ways of talking about freedom or liberty: the singular and the plural. We speak in singular terms when we want to indicate how free a given person is; we use the plural form when

we want to discuss which liberties a person possesses. It is also widely assumed that not all liberties are created equal: some are more important than others. How are we to decide, though, as to which liberties are the most important or basic?

In this paper, we focus on John Rawls's theory to answer this question. Two moral powers are central to Rawls's approach to specifying the basic liberties: the capacity for a sense of justice and the capacity for a conception of the good. The fundamental case in which the first capacity is exercised is in "the application of the principles of justice to the basic structure and its social policies". The fundamental case in which the second capacity is exercised is in "forming, revising, and rationally pursuing such a conception over a complete life". Rawls distinguishes between two phases of specification. In the first, a list of basic liberties is identified at a high level of abstraction. In the second, the significance of different particular liberties is determined and adjudication over conflicts between them occurs. Rawls's complete account, we argue, defines as basic those liberties that are necessary to the provision of the social conditions essential for the full and informed exercise of the moral powers (first phase) in the two fundamental cases (second phase).

We argue that this account excludes certain particular liberties that it should not: e.g., some forms of freedom of movement. We suggest that probability should take over the role assigned to necessity in Rawls's account. On this approach also, however, some particular freedoms rightly regarded by Rawls as basic liberties do not meet the definition. To remedy this, we appeal to the liberal principle of legitimacy and we defend this move.

Sabine Thuermel, Technische Universität München, sabine@thuermel.de

Smart Systems and Smart Power in Digital Societies

In 2008 the vision of a *Smart City* was described as follows: "the goal of such a city is to optimally regulate and control resources by means of autonomous IT systems" (Siemens AG 2008). Ten years later concrete implementations of subsystems of such a city already exist as *Smart Mobility* for the optimal regulation of traffic, *Smart Energy* for efficient energy management or *Smart Health* for ambient assisted living. The prerequisite of any smart system is a sensor rich and datafied environment. The data gained are used by predictive algorithms to predict future behavioral patterns and optimize the resources accordingly. The focus is on providing knowledge under conditions of uncertainty in order "to know ahead and act before" trying to streamline processes towards enhanced efficiency. In the next development stage, the transition from prediction to prescription takes place: future behavior is not only anticipated but formed. Context-specific, adaptive micro-directives may be incorporated in future intelligent infrastructures to guarantee optimal service from a systems' perspective and nudge or even coerce the human participants towards the desired behavior. Thus such prescriptive smart systems manifest power relations demonstrating the power of technology in a Foucauldian way: "power is employed and exercised through a netlike organization" (Foucault 1980). However, in these smart systems humans may not only behave as intended but also act in a subversive way demonstrating that "individuals are the vehicles of power, not its points of application" (ibidem). Thus, even if these environments restrict human autonomy, they also open up possibilities for undermining such systems. They are "dispositifs" in the Foucauldian sense possessing the dual structure of manifestation of power and the possibility of subverting it.

Kujutluse idealiseeritud esemed

Mu ettekande eesmärgiks on analüüsida seda, kuidas kujutlusvõime laseb meil nautida tajusituatsioone, mida me ei saaks nautida, kui nad esineksid tegelikus maailmas. Me ei saaks neist mõnu tunda, kuna sellised tajusituatsioonid kombineerivad sensoorseid omadusi, mille koosinemine tegelikus maailmas ei ole kas võimalik või ei oleks nauditav. Ma kutsun seda “tegelikkuses juurdumata naudinguks” (TJN).

Mu argumenteerin oma ettekandes kahe väite kasuks. Esiteks, ma väidan, et TJN-i kogemuse näol on tegemist idealisatsiooniga, mille puhul esitatav situatsioon ei saa tegelikkuses esineda samal viisil, nagu ta kujutuses ilmneb. TJN eristub seega abstraktsioonist, mille puhul esitatava situatsiooni parameetrid on lihtsustatud, aga mis saab sellegipoolest tegelikkuses realiseeruda. Sellest tuleneb ühtlasi, et populaarne vaade, mille kohaselt kujutlusvõime imiteerib tegelikkust representeerivaid tunnetuslikke võimeid, on problemaatiline. Kujutlusvõime pigem manipuleerib kui imiteerib neid. Teiseks, ma seostan TJN-i esteetilisse kontekstiga ning väidan, et TJN-i üheks sobivaimaks allikaks on digitaalne fotograafia, mis võimaldab esemete ja olukordade esituse manipuleerimist viisidel, mis soodustavad reaalsusest irdunud kujutlusi. Näiteks stiliseeritud toidufotograafia esitab tihti “sööke”, mis ei pruugiks vaatajale maitsta ning ei pruugi olla isegi söödavad, mis aga just tänu sellele lasevad tegelikkust ignoreerida ning konstrueerida oma kujutluses objekt nii, et see ikkagi naudingut pakub. TJN võimaldab seega teatav sorti infleksiooni, mille puhul kujutluses ilmnev ese on kogetav teistmoodi kui seda oleks võimalik kohata tegelikkuses. Kuna idee infleksioonist on välja käidud piltide tajumise kontekstis, kujutab TJN endast selle idee laiendust ning lubab paremini mõista infleksiooni võimalusi teisel pool visuaalset modaalsust.

The History of Philosophy in the Light of AI: Hegel or Leibniz

Philosophy already has a long history of coming to terms with AI. But if the future of the concept is indeed inseparable from artificial languages and ubiquitous computing, then philosophy must also be able to understand and rewrite its own history in this unnatural light. To this end, I distinguish two manners in which modern philosophy has pursued the artificial cultivation of intelligence. The first is Hegelian. Recently, Yuk Hui and Reza Negarestani have pointed to the affinity between the Hegelian notion of absolute spirit and the functioning of intelligence found in cybernetics and systems theory, as well as in cognitive science. As technology becomes our destiny, this leads them to the problem of the continued relevance of humans to the history of a general self-authorizing intelligence. By contrast, I propose to bluntly identify intelligence itself with the sense for relevance (i.e. singularity). Philosophically speaking, this identification reaches back to the proto-structuralist system of Leibniz, which aims for universal communication. Leibniz’s many inventions of formal languages, from the binary system and the universal characteristic to magic and the calculator, constitute a proto-AI that functions as operative code of an inclusive civility. My thesis is the following: if Hegel offered the first grand narrative of the recursive self-critique of the immediacy of common sense in the form of artificial good sense, Leibniz bets on a cybernetic reason that contributes to the distributive composition of an unnatural common sense, all the while protecting multiplicity against its collectivization by a self-naturalizing good sense.

Veebientsüklopeediate episteemilisest eripärast Vikipeedia näitel

Ettekandes tulevad vaatluse alla uusaegsete eurooplike entsüklopeediate koht teaduse ökosüsteemis ja nende roll ühiskondliku episteemilise tööriistana ning veebientsüklopeediate eripära Vikipeedia näitel.

Uusaja teaduse kujunemisest saati on universaalentsüklopeediatel olnud teaduse ökosüsteemis oluline roll sotsiaalse tõeagregaatorina. Entsüklopeediad on üks olulisemaid kanaleid, mis suunab ja filtreerib teadusinfot laiale publikule või, teistpidises vaates, ühiskonna üks peamisi episteemilisi vahendeid sotsiaal- ja massimeedia kõrval. Teadusfilosoofias ja epistemoloogias pole entsüklopeediatele aga suurt tähelepanu pööratud, napib isegi ajaloolisi ja praktilisi käsitlusi. Mõningal määral on entsüklopeediad tematiseerunud avalikus diskursuses pärast interneti tulekut ja veebientsüklopeediate väljakujunemist, mis on tõstnud ka entsüklopeediate tähtsuse ja leviku seniolematule tasemele, kuid nende toimimist on uuritud pigem infotehnoloogia ja sotsioloogia vahenditega kui sotsiaalepistemoloogias.

Üldkäsitlustes vastandatakse sageli veebientsüklopeediaid paberentsüklopeediatele, keskendudes eelkõige Vikipeediale, ent kui tüüpiliselt esiletoodavaid nähtusi lähemalt vaadelda, on need pigem esinenud entsüklopeedias ka ajalooliselt (sekundaarsus allikana, ühiskondlik või koolkondlik erapoolikus, autorite anonüümsus, heterogeensus). Uute entsüklopeediate tegelikust episteemilisest eripärast võib tuua järgmisi näiteid: veebipõhine tööprotsess toob nähtavale teadmiste valiku ja sünteesi protsessi; tunnistuste hindamisel nihkub rõhk eksperdi autoriteedilt sisulisele argumentatsioonile; veebitehnoloogia võimaldab kaasata entsüklopeediasse uusi teadmiste tüüpe ja esitusviise; avaldub teadmiste dünaamilisus ja entsüklopeedia epistemoloogiline protsessuaalsus; akuutsemaks muutuvad autoriõigus ja andmekaitse; uusaja entsüklopeediatraditsioonile omane püüdlus lugeja autonoomia ja võimestamise poole süveneb veelgi.

Teadmiste ja tõe vahekord kaldub veebientsüklopeediate enesemõtestuses teadmiste poole, kuid igatsus tõe järele püsib vajaduse tõttu allikate selektsiooniks. Ühelt poolt on kõik need aspektid kutsunud esile ühiskondlikke vastureaktsioone, teisalt aga ilmnevad samad protsessid teaduseski. Samas on selge, et entsüklopeediate kohanemine infoühiskonnaga ei ole sugugi lõppenud, ning aruteludes veebientsüklopeediate edasise arengu üle ongi oluline roll epistemoloogilistel aspektidel (nn traditsioonilised teadmised, masinloetavad keelesõltumatud ontoloogiad, teadmiste automatiseeritud süntees), mis võiksid pakkuda lähemat huvi ka teadusfilosoofiale ja semiootikale.

Online Friendship: How Computer-mediated Communication Changes our Relationships

In a digital age, new possibilities of communication transcend all areas of human life. The internet allows us to connect with people all over the world and to establish new bonds and alliances. For a while, scholars were mostly concerned with possibilities of exclusively online bonds ever coming close to traditional forms of close personal relationships such as friendship. It was argued that forms of online communication would necessarily stay limited and superficial.

But our online communication today is hardly limited to those far away from us. It has become an integral part of our everyday life and is thus starting to transform the relationships we uphold even with those living close to us. 'Offline' friendships as well as partnerships and family ties are to an

increasing degree maintained by forms of computer-mediated communication. This talk sets out to shed some light on how this transforms our personal relationships. Forms of online communication are diverse. To allow a differentiated perspective the talk will focus on the possibilities, limits and imperatives of (a) mass communication on a social media platform like *facebook* as well as (b) frequent exchanges of short messages via mobile phones. The potential of a new form of intimacy through constant micro-communications will be highlighted as well as the danger of impersonal communication through seemingly efficient options of always addressing a multitude of friends.
