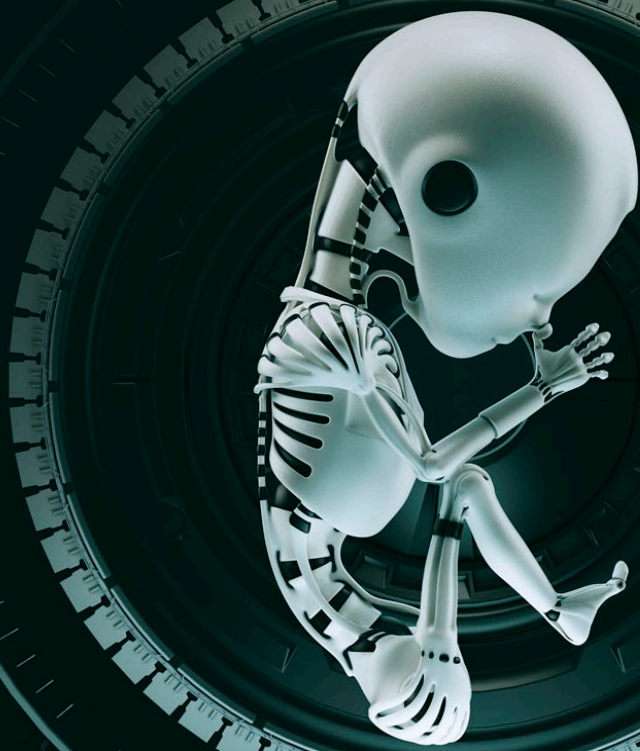
18TH INTERNATIONAL CONFERENCE
ON THE ETHICAL AND SOCIAL IMPACTS OF ICT

Paradigm Shifts in ICT Ethics
Proceedings of the ETHICOMP 2020

Edited by

JORGE PELEGRÍN-BORONDO
MARIO ARIAS-OLIVA
KIYOSHI MURATA
ANA MARÍA LARA PALMA



UNIVERSIDAD DE LA RIOJA

UNIVERSITAT ROVIRA i VIRGILI

Edited by

Jorge Pelegrín-Borondo

Mario Arias-Oliva

Kiyoshi Murata

Ana María Lara Palma

ETHICOMP 2020

Paradigm Shifts in ICT Ethics

Proceedings of the ETHICOMP 2020

18th International Conference on the Ethical and Social Impacts of ICT

Logroño, Spain, June 2020

UNIVERSIDAD DE LA RIOJA

UNIVERSITAT ROVIRA i VIRGILI

PROCEEDINGS OF THE ETHICOMP* 2020

18th International Conference on the Ethical and Social Impacts of ICT

Logroño, La Rioja, Spain

June 15 – July 6 (online)

Title	Paradigm Shifts in ICT Ethics
Edited by	Jorge Pelegrín-Borondo (University of La Rioja), Mario Arias-Oliva (Universitat Rovira i Virgili), Kiyoshi Murata (Meiji University), Ana María Lara Palma (University of Burgos)
ISBN	978-84-09-20272-0
Local	Logroño, Spain
Date	2020
Publisher	Universidad de La Rioja

Publisher: Universidad de La Rioja, www.unirioja.es

Cover designed by Universidad de La Rioja, Servicio de Comunicación, and Antonio Pérez-Portabella.

ISBN 978-84-09-20272-0

Presidency of the Scientific Committee

Mario Arias-Oliva

Jorge Pelegrín-Borondo

Kiyoshi Murata, Meiji University, Japan

Ana María Lara-Palma, Universidad de Burgos, Spain

Scientific Committee

Alejandro Catáldo, Talca University, Chile

Alicia Blanco González, Universidad Rey Juan Carlos, Spain

Alicia Izquierdo-Yusta, Universidad de Burgos, Spain

Alireza Amrollahi, Australian Catholic University, Australia

Camilo Prado Román, Universidad Rey Juan Carlos, Spain

Cristina Olarte Pascual, University of La Rioja, Spain

Dana AlShwayat, Petra University, Jordan

Efpraxia Zamani, University of Sheffield, United Kingdom

Emma Juaneda-Ayensa, University of La Rioja, Spain

Eva Reinares, Universidad Rey Juan Carlos, Spain

Gosia Plotka, PJAIT, Poland & De Montfort University, UK

Graciela Padilla Catillo, Complutente University of Madrid, Spain

Jesús García de Madariaga Miranda, Complutente University of Madrid, Spain

Joaquín Sánchez Herrera, Complutense University of Madrid, Spain Jorge Gallardo-Camacho, Camilo José Cela University, Spain

José Antonio Fraid Brea, Universidad de Vigo, Spain

Juan Carlos Yañez Luna, Autonoma University of San Luis de Potosí, Mexico

Katleen Gabriels, Maastricht University, The Netherlands

Kutoma Wakunuma, De Montfort University, UK

María del Pilar Martínez Ruiz, Castilla - La Mancha University, Spain

Marta Czerwonka, Polish-Japanese Academy of Information Technology, Poland

Marty J. Wolf, Bemidji State University, USA

Oliver Burmeister, Charles Sturt University, Australia

Paul B. de Laat, University of Groningen, The Netherlands

Pedro Isidoro González Ramírez, Autonoma University of San Luis de Potosí, Mexico

Simon Rogerson, De Monfort University, UK.

Shalini Kesar, Southern Utah University, USA

Sonia Carcelén García, Complutente University of Madrid, Spain

Teresa Pintado Blanco, Complutente University of Madrid, Spain

Ulrich Schoisswohl; Austrian Research Promotion Agency (FFG), Austria

Wade Robison, Rochester Institute of Technology, USA

Wilhelm E. J. Klein, Researcher in Technology and Ethics, Hong Kong

William M. Fleischman, Villanova University, USA

Yohko Orito, Ehime University, Japan

Yukari Yamazaki, Seikei University, Japan

Presidency of the Organizing Committee

Mario Arias-Oliva, Universitat Rovira i Virgili, Spain

Jorge Pelegrín-Borondo, University of La Rioja, Spain

Kiyoshi Murata, Meiji University, Japan

Emma Juaneda-Ayensa, University of La Rioja, Spain

Ana María Lara-Palma, Universidad de Burgos, Spain

Organizing Committee

Alba García Milon, University of La Rioja, Spain

Alberto Hernando García-Cervigón, Universidad Rey Juan Carlos, Spain

Ana María Mosquera de La Fuente, University of La Rioja, Spain

Anne-Marie Tuikka, University of Turkku, Finland

Antonio Pérez-Portabella, Universitat Rovira i Virgili, Spain

Araceli Rodríguez Merayo, Universitat Rovira i Virgili, Spain

Erica L. Neely, Ohio Northern University, USA

Elena Ferrán, Escola Oficial de Idiomas de Tarragona, Spain

Jan Strohschein, Technische Hochschule Köln, Germany

Jorge de Andrés Sánchez, Universitat Rovira i Virgili, Spain

José Antonio Fraid Brea, Universidad de Vigo, Spain

Juan Luis López-Galiacho Perona, Universidad Rey Juan Carlos, Spain

Kai Kimppa, University of Turkku, Finland

Leonor González Menorca, University of La Rioja, Spain

Luz María Marín Vinuesa, University of La Rioja, Spain

Manuel Ollé Sesé, Complutente University of Madrid, Spain

Mar Souto Romero, Universidad Internacional de La Rioja, Spain

María Alesanco Llorente, University of La Rioja, Spain

María Arantxazu Vidal, Universitat Rovira i Virgili, Spain

María Yolanda Sierra Murillo, University of La Rioja, Spain

Mónica Clavel San Emeterio, University of La Rioja, Spain

Orlando Lima Rua, Politechnic of Porto, Portugal

Rubén Fernández Ortiz, University of La Rioja, Spain

Stéphanie Gauttier, Grenoble Ecole de Management, France

Teresa Torres Coronas, Universitat Rovira i Virgili, Spain

ETHICOMP Steering Committee

KEYNOTE ADDRESSES

A journey in Computer Ethics: from the past to the present. Looking back to the future

by

Simon Rogerson (De Montfort University, UK), Shalini Kesar (Southern Utah University, USA), Don Gotterbarn (East Tennessee State University, USA), Katleen Gabriels (Maastricht University, Netherlands)

Ethical aspects and moral dilemmas generated by the use of chabots

by

Jesús García de Madariaga (Complutense University of Madrid, Spain), Crisitina Olarte-Pascual (University of La Rioja, Spain), Eva Reinares-Lara (Rey Juan Carlos University)

The Ethics of Cyborgs

by

Jorge Pelegrín-Borondo (University of La Rioja, Spain) and
Mario Arias-Oliva (Universitat Rovira i Virgili)

SUPPORTED BY

University of La Rioja

Universitat Rovira i Virgili

Ayuntamiento de Logroño

Centre for Computing and Social Responsibility, De Montfort University

Centre for Business Information Ethics, Meiji University

To those who passed away due to the COVID-19 pandemic

The ETHICOMP conference series was launched in 1995 by the Centre for Computing and Social Responsibility (CCSR). Professor Terry Bynum and Professor Simon Rogerson were the founders and joint directors. The purpose of this series is to provide an inclusive international forum for discussing the ethical and social issues associated with the development and application of Information and Communication Technology (ICT). Delegates and speakers from all continents have attended. Most of the leading researchers in computer ethics as well as new researchers and doctoral students have presented papers at the conferences. The conference series has been key in creating a truly international critical mass of scholars concerned with the ethical and social issues of ICT. The ETHICOMP name has become recognised and respected in the field of computer ethics.

ETHICOMP previous conferences:

ETHICOMP 1995 (De Montfort University, UK)

ETHICOMP 1996 (University of Salamanca, Spain)

ETHICOMP 1998 (Erasmus University, The Netherlands)

ETHICOMP 1999 (LUISS Guido Carli University, Italy)

ETHICOMP 2001 (Technical University of Gdansk, Poland)

ETHICOMP 2002 (Universidade Lusiada, Lisbon, Portugal)

ETHICOMP 2004 (University of the Aegean, Syros, Greece)

ETHICOMP 2005 (Linköping University, Sweden)

ETHICOMP 2007 (Meiji University, Tokyo, Japan)

ETHICOMP 2008 (University of Pavia, Italy)

ETHICOMP 2010 (Universitat Rovira i Virgili, Spain)

ETHICOMP 2011 (Sheffield Hallam University, UK)

ETHICOMP 2013 (University of Southern, Denmark)

ETHICOMP 2014 (Les Cordeliers, Paris)

ETHICOMP 2015 (De Montfort University, UK)

ETHICOMP 2017 (Università degli Studi di Torino, Italy)

ETHICOMP 2018 (SWPS University of Social Sciences and Humanities, Poland)

Table of contents

1. Creating Shared Understanding of 'Trustworthy ICT'

Track chair: Ulrich Schoisswohl, Austrian Research Promotion Agency (FFG), Austria

ARTIFICIAL INTELLIGENCE: HOW TO DISCUSS ABOUT IT IN ETHICS

Olli I. Heimo, Kai K. Kimppa

University of Turku (Finland)

olli.heimo@utu.fi, kai.kimppa@utu.fi

EXTENDED ABSTRACT

Artificial intelligence (AI) is the buzzword for the era and is penetrating our society in levels unimagined before – or so it seems to be (see e.g. Newman, 2018; Branche, 2019; Horaczek, 2019). In IT-ethics discourse there is plenty of discussion about the dangers of AI (see e.g. Gerdes & Øhstrøm 2015) and the discourse seems to vary from loss of privacy (see e.g. Belloni et al. 2014) to outright nuclear war (See e.g. Arnold & Scheutz 2018) in the spirit of the movie *Terminator 2*.

Yet it seems that with AI discussion there is a lot of space for misunderstandings and misrepresentations starting from but not limited to what is AI. In this paper therefore the AI from the ethical perspective of what we should discuss about AI is presented.

There is of course various different ways to conceptualise the difference between different kinds of things labelled as AI. Whereas the technical ones have the tendency to focus on the technical structure of the tool at hand, from the ethical point of view the focus should be more on 1) what the system can do and 2) how it does it. Moreover, we should also focus on the issue on how the bad consequences could be avoided (Mill 1863) and how the people with malicious intentions could be controlled (Rawls 1971). There of course are different motivations and (hopeful) consequences when using AI, which are duly worthy of a different discourse and study in themselves), but in this paper the issue of *definition* for the use itself is discussed. Hence, in the full paper we will discuss the following four different groups of AI in depth:

1) Scripts (gaming and otherwise)
2) Data mining and analysis
3) Weak AI (In its current form: neural networks, machine learning, mutating algorithms etc.)
4) General AI (Skynet, HAL, Ex Machina, etc.)

First of all the *scripts*, mostly advertised as “AI” in computer games are just “simple” algorithms. As these are mostly the first version of AI we meet when talking about it, we must remember that they are merely scripts and cheating (i.e. not AI at all) to make the opponents in computer games more lifelike, to make the sensation that you are playing against actual intelligent opponents. This of course is not true because the easiest, cheapest, and thus most profitable way to give the illusion of a smart enemy is to give the script the power of knowing something they should not. Hence the idea is to give the player the illusion, but the actual implementation is much simpler (and for smarter or more experienced players also quite transparent…). That is the art of making a good computer game opponent. Hence computer game AIs are just glorified mathematical models to entertain the customers.

The second one discussed as an AI quite often is *data mining* and the related data analysis, just gathering a lot of information from a huge pile of data. Yet this is usually and mostly done by scripting;

Patterns and mathematical models are found and tiny bits of data from the patterns are combined to find similarities, extraordinarities and peculiarities then to be analysed by humans aided by a traditional algorithm. There is nothing intelligent about these algorithms except the people making them. Therefore, they too are just glorified mathematical models and smart people working with them – a massive difference to the former though.

Thirdly, we discuss machine learning, mutating algorithms, neural networks and other state of the art AI research, i. e. *weak AI*. This is *the point* we should currently focus on when discussing themes related to AI. These methods make the computer better by every step the computer makes; every decision the computer makes improves the computer, not the user.

To clarify, Artificial Intelligence refers to a system, in which is a mutating algorithm, a neural network, or similar structure (also known as weak AI) where the computer program "learns" from the data and feedback it is given. These technologies are usually opaque (i.e. black box –design), so even their owners or creators cannot *know* how or why the AI ended up with the particular end-result. (See e.g. Covington, Adams, and Sargin, 2016). As AI has been penetrating the society in many different levels for years, e.g. banking, insurance, and financial sectors (see e.g. Coeckelbergh, 2015).

The fourth issue often discussed in the field of AI is the living AI, the thinking AI – the feeling and fearing AI, the "Skynet", the singularity "the moment at which intelligence embedded in silicon surpasses human intelligence" (Burkhardt, 2011) and starts to consider itself equal or better than humans. These AIs are luckily or sadly, depending on the narrative the utopia or the dystopia, are still mere fiction and in the technological scale in a future we cannot yet even comprehend.

When discussing technology, the possibilities of technology and possible technologies we must be aware that the first of these does already exist. The second one of these is due to exist, and the third one may exist. While it is possible that technology will exist in say 5-10 years, we also must remember that the society will not be what it is now and other technologies will exist and the society has moved on. There are numerous issues within the field of AI currently at hand, e.g. biased AI (Heimo & Kimppa 2019), liability of autonomous vehicles (see e.g. Heimo, Kimppa & Hakkala 2019), weaponizing AI systems (see e.g. Gotterbarn, 2010), facial recognition (see e.g. , Heimo & Kimppa 2019; Doffman, 2019) just to mention few. Moreover there are plenty of near-future applications of these that must be handled before they become a critical issue. Yet it is important to discuss about all the levels of AI technologies – and to tie them to their timeline!

As we know we must interpret the writings of the past for they were written in their time (see e.g. MacIntyre, 2014), we must also interpret the future which will be different in ways we cannot fully understand. Therefore to predict the AI can do in 10-20 years time is quite different when we cannot fathom what kind of society we will have in 10 years' time. We must yet keep in mind that what we give up now in the sense of privacy, personal information, liberties etc. can and will be taken away from us more efficiently with the future AI, especially if we follow the Chinese route, which is possible. But to talk of the society now with a futuristic AI seems intellectually dishonest. We do not have flying cars, hoverboards nor the cure for cancer, things predicted and assumed by everyone in any popular culture from the 80s or 90s (see e.g. Back to the Future) yet we have Twitter, Wikipedia and cat picture meems, not something we would actually have been predicting at the time. It is not that we would say that predicting future is irrelevant, moreover we wish to encourage people, scientists and philosophers to focus be explicit when predicting the future; to emphasize their predictions of the timeline they assume technology be in use. Hence when we are talking about AI there are many possibilities for the future but a General AI is a as much of a thing of a future we cannot yet predict, as datamining is a thing of the past. Predictions as predictions, and facts as facts, that is all we can do for honest science.

In the full paper we will discuss more about how we – as scientists – should use these and other timeline-specific issues when discussing ethical issues in IS and IT development to enhance the specifity and credibility of our research.

KEYWORDS: Artificial Intelligence, Ethics, Weak AI, Discourse.

REFERENCES

Arnold T. & Scheutz M. (2018) The ”big red button” is too late: an alternative model for the ethical evaluation of AI systems, Ethics and Information Technology, 20:59-69.

Belloni, A. et al. (2014) Towards A Framework To Deal With Ethical Conflicts In Autonomous Agents And Multi-Agent Systems, CEPE 2014.

Branche, P. (2019), Artificial Intelligence Beyond The Buzzword From Two Fintech CEOs, Forbes, Aug 21 2019, https://www.forbes.com/sites/philippebranch/2019/08/21/artificial-intelligence-beyond-the-buzzword-from-two-fintech-ceos/#43f741c7113d

Coeckelbergh, M. (2015) The tragedy of the master: automation, vulnerability, and distance, Ethics and Information Technology, 17:219-229.

Covington, P., Adams, J., and Sargin, E. (2016) Deep neural networks for youtube recommendations. Proceedings of the 10th ACM conference on recommender systems. ACM, 2016.

Doffman, Z. (2019) China’s ‘Abusive’ Facial Recognition Machine Targeted By New U.S. Sanctions, Forbes, Oct 8, 2019. https://www.forbes.com/sites/zakdoffman/2019/10/08/trump-lands-crushing-new-blow-on-chinas-facial-recognition-unicorns/#52641d79283a

Gerdes, A. & Øhstrøm, P. (2015) Issues in robot ethics seen through the lens of a moral Turing test, JICES 13/2:98-109.

Gotterbarn, D. (2010) Autonomous weapon's ethical decsions:" I am sorry Dave; I am afraid I can't do that.". In proceedings of ETHICOMP 2010 The “backwards, forwards and sideways” changes of ICT Universitat Rovira i Virgili, Tarragona, Spain 14 to 16 April 2010.

Heimo, O. I. & Kimppa, K. K. (2019), No Worries–the AI Is Dumb (for Now), Proceedings of the Third Seminar on Technology Ethics 2019 Turku, Finland, October 23-24, 2019, pp. 1-8.

Heimo, O. I., Kimppa, K. K. & Hakkala, A (2019) Automated automobiles in Society, IEEE Smart World Congress, Leicester, UK, 2019.

Horaczek, S. (2019), A handy guide to the tech buzzwords from CES 2019, Popular Science Jan 9 2019, https://www.popsci.com/ces-buzzwords/

Mill, John S. (1863) Utilitarianism, https://www.utilitarianism.com/mill1.htm, accessed 21.10.2019.

Newman, D. (2018) Top 10 Digital Transformation Trends For 2019, Forbes, Sep 11, 2018, https://www.forbes.com/sites/danielnewman/2018/09/11/top-10-digital-transformation-trends-for-2019/#279e1bca3c30

Rawls, J. (1971) A Theory of Justice, Belknap Press of Harvard University Press, Cambridge, Massachusetts.

Robbins S. (2018) The Dark Ages of AI, Ethicomp 2018.

DEVELOPING A MEASURE OF ONLINE WELLBEING AND USER TRUST

Liz Dowthwaite, Elvira Perez Vallejos, Helen Creswick, Virginia Portillo, Menisha Patel, Jun Zhao

University of Nottingham (UK), University of Nottingham (UK), University of Nottingham (UK), University of Nottingham (UK), University of Oxford (UK), University of Oxford (UK)

liz.dowthwaite@nottingham.ac.uk; elvira.perez@nottingham.ac.uk; helen.creswick@nottingham.ac.uk; virginia.portillo@nottingham.ac.uk; menisha.patel@cs.ox.ac.uk; jun.zhao@cs.ox.ac.uk

EXTENDED ABSTRACT

As interaction with online platforms is becoming an essential part of people's everyday lives, the use of automated decision-making algorithms in filtering and distributing the vast quantities of information and content to users is having an increasing effect on society, with many people raising questions about the fairness, accuracy and reliability of such outcomes. Users often do not know when to trust algorithmic processes and the platforms that use them, reporting anxiety and uncertainty, feelings of disempowerment, defeatism, and loss of faith in regulation (Creswick et al., 2019; Knowles & Hanson, 2018). This leads to concerns about wellbeing, which can negatively affect both the user and broader society. It is therefore important that mechanisms and tools are introduced which support users in the responsible building of trust in the online world. This paper describes the ongoing development of an 'Online Wellbeing Scale' to aid in understanding how trust (or lack of trust) relates to overall wellbeing online.

There are two broad aims of the Scale. For researchers, it will allow exploration of the relationship between different types of wellbeing, trust, and motivation, to understand how trust affects user's online experiences, as well as comparison across different online activities to highlight where the major issues are. For the users, the Scale will contribute to the development of a 'Trust Index' tool for measuring and reflecting on user trust, as part of engaging in dialogue with platforms in order to jointly recover from trust breakdowns. It will be part of a suite of tools for empowering the user to negotiate issues of trust online. This also contributes to design guidelines for the inclusion of trust relationships in the development of algorithm-driven systems.

The first stage of development of the Online Wellbeing Scale/Trust Index took place as part of a larger study into online trust, comparing attitudes of younger (16-25 years old) and older (over 65) adults. The study was approved by the Ethics Review Board for the Department of Computer Science at the University of Nottingham. Sixty participants took part in a series of 3 hour workshops. The project focused on user-driven, human-centred, and Responsible Research and Innovation approaches to investigating trust. Thus the workshop structure, including timings and ordering of tasks, the kinds of tasks to be completed, and practical consideration were co-created through a series of activities with members of the public in the relevant age groups, ensuring that the questions and tasks were relevant, understandable, and engaging. The workshops took a mixed-methods approach to encourage participants to think about issues in different contexts, and included pre- and post-session questionnaires exploring factors related to trust, motivation, digital literacy, and wellbeing.

The questionnaires were designed to explore whether there is a link between these factors, and how this might be measured. They consisted of a mixture of free text, multiple choice, and Likert-like items. The pre-session questionnaire asked about: *Activity:* 5 items of the type of activity that people do online, including socialising, shopping, information seeking, entertainment, and sharing content; *Trust:*

considerations of trust, including personal experiences and opinions; and *Digital Confidence:* Statements related to perceived digital literacy and how confident users are in carrying out tasks online.

The post-session questionnaire began with some questions about the session, then repeated statements from the pre-session questionnaire to see if there were changes in opinion, followed by open-text questions about online trust and wellbeing, and ratings of how much various features of websites affect their trust. Finally they were asked to complete 2 instruments for measuring wellbeing, modified to reflect online experiences: *Eudaimonic Wellbeing:* Basic Psychological Need Satisfaction (BPNS) scale (Gagné, 2003; Ryan & Deci, 2000; Ryan, Huta, & Deci, 2006); and *Emotional Wellbeing:* Scale of Positive and Negative experience (SPANE) (Diener & Biswas-Diener, 2009; Diener et al., 2010).

Rather than report the analysis of the responses to the questionnaire, the results here detail how they fed into the development of the prototype of the Online Wellbeing Scale, and how this will be used to investigate the role of trust online.

The 48 item scale is in 5 blocks, with each 'construct' measured by 6 statements on 7-point Likert or Likert-like scales. The *Activity* block covers the items from the initial questionnaire, plus an additional item, 'financial or organisation' which covers all activities suggested by participants. This block measures the frequency of each activity. The *Digital confidence* block with the original 7 statements from the pre-session questionnaire has cronbach's alpha (α) reliability of 0.800; removing one item to create a 6 item scale improves reliability to α=0.830. The *Eudaimonic Wellbeing* block from the original post-session questionnaire had low reliability for autonomy (α=0.605) and competence (α=0.510). Only relatedess reached an acceptable level (α=0.814). As such, the scale as a whole is not considered reliable. For the prototype Online Wellbeing Scale it has been replaced with a modified version of the Balanced Measure of Psychological Needs (Sheldon & Hilpert, 2012). This scale uses simpler language and reduces each construct to 6 items, with the ability to calculate the overall level of satisfaction and dissatisfaction of needs. It was noted that, particularly for the older age group, statements relating to interacting with people online were often either ignored or misunderstood. Therefore modifications include focusing wording on the online world and replacing specific references to 'people' with a more general interactional focus, eg "There were people telling me what I had to do" was replaced with "I was being told what I had to do". The *Emotional Wellbeing* block scored a good reliability for the positive experience scale of α=0.793, improved by removing one item to α=0.815, and the negative experience scale of α=0.830, improved by removing one item to α=0.831. As this is a bipolar scale, the equivalent positive and negative words were also removed, resulting in a modified scale with 6 items each for positive and negative experience, the reliability of which is α=0.819 and α=0.818 respectively. Finally, the *Trust* block consists of 6 statements using a combination of the qualitative results from the workshops and the questionnaire results surrounding levels of trust in online systems, with reference to related literature on trust (for example Gefen, Karahanna, & Straub, 2003). The modified prototype of the Online Wellbeing Scale will be tested in an online study to take place later this year. This will allow both validation of the scale and large-scale examination of the role of trust in online wellbeing. It will help lead to recommendations for ways in which online platforms can build user trust into their systems. At the same time, using the Scale as the beginning of a 'Trust Index' for reflection and empowerment will be explored with both stakeholders and users, including investigating ways to present results that are meaningful and engaging, and how this and other tools can encourage meaningful dialogue between the two groups.

KEYWORDS: wellbeing, trust, online experience, scale.

REFERENCES

Cheshire, C. (2011). Online Trust, Trustworthiness, or Assurance? Daedalus, 140(4), 49-58. https://doi.org/10.1162/DAED_a_00114

Creswick, H., Dowthwaite, L., Koene, A., Perez Vallejos, E., Portillo, V., Cano, M., & Woodard, C. (2019). "… They don't really listen to people": Young people's concerns and recommendations for improving online experiences. Journal of Information, Communication and Ethics in Society, 17(2). https://doi.org/10.1108/JICES-11-2018-009

Gagné, M. (2003). The role of autonomy support and autonomy orientation in prosocial behavior engagement. Motivation and Emotion, 27(3), 199–223.

Gefen, Karahanna, & Straub. (2003). Trust and TAM in Online Shopping: An Integrated Model. MIS Quarterly, 27(1), 51. https://doi.org/10.2307/30036519

Knowles, B., & Hanson, V. L. (2018). Older Adults' Deployment of 'Distrust'. ACM Transactions on Computer-Human Interaction, 25(4), 1–25. https://doi.org/10.1145/3196490

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. American Psychologist, 55(1), 68.

Ryan, R. M., Huta, V., & Deci, E. L. (2006). Living Well: A self-determination theory perspective on eudaimonia. Journal of Happiness Studies, 9, 139–170.

Sheldon, K. M., & Hilpert, J. C. (2012). The balanced measure of psychological needs (BMPN) scale: An alternative domain general measure of need satisfaction. Motivation and Emotion, 36(4), 439–451. https://doi.org/10.1007/s11031-012-9279-4.

ETHICAL DEBT IN IS DEVELOPMENT. COMPARING TECHNICAL AND ETHICAL DEBT

Olli I. Heimo, Johannes Holvitie

University of Turku (Finland), Hospital District of Southwest Finland (Finland)

olli.heimo@utu.fi; jjholv@utu.fi

EXTENDED ABSTRACT

Information system development process is a trade and art of creating complex systems to support the handling of information and processes in modern society. All information systems that encompass a large and complex software development part contain technical debt (*removed for peer review*). Technical debt describes a phenomenon where certain software development driving aspects – like deployment time – are deliberately or indeliberately prioritized over others – like internal quality. Due to having no impact on functionality, technical debt is easily overlooked. Technical debt is often exploited in order to enhance the software development process and to make minimum viable products (MVPs). Quick releases for example allow the developing organization to iterate the information system thus minimizing the work required. However, the downside of this is that – if development is continued for the MVP – the debt will cumulate and diminish the efficiency of the software development process for the information system in question (*removed for peer review*). Proper acknowledgement and management of technical debt allows the debt to be taken into account and informed decision to be made in order to pay the debt back.

Technical debt was first used by Ward Cunningham in a technical report (Cunningham, 1992). Herein, Cunningham described software development as taking on debt; developer work on current knowledge and produce pieces of software that fulfill functional requirements. However, the environment and knowledge develops. At a later stage the software still functions correctly, but its internal quality can be assessed to be sub-optimal. Debt can be paid by working on the internal quality, but it produces no added functional value. However, high internal quality allows the software to be developed efficiently.

The definition of technical debt has been revisited by several authors (Tom et al. 2013; Avgeriou et al., 2016). Most notably from this consensus, we see definitions for technical debt's principal, interest, and interest probability. Principal refers to the work amount required to increase the original software component's internal quality. Interest refers to the amount of extra work that is committed when ever the original software component is referenced by other components – i.e. complex or low quality component interface requires more work on the interface-users side. Finally, interest probability captures the chance that the original, sub-optimal software component is referenced by upcoming software development iterations; the interest becomes payable.

Managing technical debt requires that technical debt is identified, tracked and governed (Guo and Seaman, 2011). Identification corresponds to noting sub-optimal software components and documenting the principal and interest for them. Tracking notes that these components are evaluated for each possible iteration that they shall be used and estimating the interest probability for them. Governance then becomes an evaluation process for each software component and for the chosen iteration period(s): is the technical debt's principal smaller than the expected return value for its

interest? If yes, then the iteration(s) should start by removing the technical debt. If no, the technical debt can be ignored.

Whilst the management process is trivially explained, in practise it is not easily committed. Most notable reason for this is that technical debt emerges in different ways and is often exceptionally hard to identify. Fowler (2009) describes four categories for this: deliberate-reckless, deliberate-prudent, inadvertent-reckless, and inadvertent-prudent. Deliberate-reckless captures debt which is accumulated due to reckless but identifiable decisions like ”we don’t have time for design”. Deliberate-prudent captures the informed and known decisions like ”we must ship now and deal with consequences”. Inadvertent-reckless is the unkown decision like ”what’s layering” that is made whilst developing a piece of software. Finally, inadvertent-prudent is the retrospective analysis of ones own work to identify that ”now we know how we should have done it”. Noting from the previous, the inadvertent cases will remain hidden until discovered separately whilst the deliberate-reckless cases are generally inadvertent to the organization since they are caused by improper management. Only the deliberate-prudent case depicts an informed and followed decision to accumulate technical debt.

To summarize technical debt, as a well used investment mechanism it can provide organizations with the ability to trial (MVPs) software at low cost or to enter markets prior to others. As an unknowledged component of software development, technical debt will accumulate, at an increasing phase, affecting the software development efficiency and the organization’s profitability; possibly to the point of software bankruptcy where even the smallest addition to the software costs more in unavoidable refractorings than the addition itself. (Tom et al., 2013)

Ethical debt however is a subgroup of technical debt where ethical questions are left undecided or unsolved while creating information systems. Whereas technical debt can be used as a tool to iterate and hence create new solutions, the ethical basis of the system is harder to build afterwards. In the end, ethics is not a sticker to add to a product but a process, which must be started before the development process (*removed for peer review*). Therefore it is crucial to understand the main guidelines and ethical pitfalls before the actual coding process can be started. Or as *removed for peer review* state, when changing the ethical basis of the information system and how it works you change the whole way of how the organisation works – and vice versa. Therefore, the ethical basis must be in order as the view on how the organisation works must also be clear before starting the development process (see also Leavitt, 1964). One must keep in mind that every time the process of information system development turns from one level to another (defining, planning, development, implementation, upkeep), the price of fixing errors is manifold and hence to spot the ethical questions before moving to higher levels can save numerous work hours thus making the process more efficient – or more ethical, if there is no money left to fix the errors!

However when the process has been started, more ethical questions might arise. These problems can be approached with the mentality of debt and therefore there might be situations where the ethical decisions can be postponed while in development process similarly than in technical debt situations. In the full paper the analysis of ethical debt will be more in-depth by concentrating on the situations and cases where the ethical questions and ethical debt can be considered more acceptable.

KEYWORDS: Ethics, Software Engineering, Technical Debt, Ethical Debt

REFERENCES

Avgeriou, P., Kruchten, P., Ozkaya, I., and Seaman, C. (2016) *Managing Technical Debt in Software Engineering* (Dagstuhl Seminar 16162). Dagstuhl Reports, 6(4):110–138.

Fowler, M. (2009) *Technical debt quadrant*. Online Publication. http://martinfowler.com/bliki/Technical DebtQuadrant.html.

Leavitt H. J. (1964) *Applied Organization Change in Industry: Structural Technical and Human Approaches*, in Cooper, W. W., Leavitt H. J. & Shelly, M. W. (eds.): New Perspectives in Organizational Research. Wiley, New York, USA.

Tom, E., Aurum, A., and Vidgen, R. (2013) *An exploration of technical debt*. Journal of Systems and Software, 86(6):1498–1516.

Cunningham, W. (1992) *The WyCash portfolio management system* in Addendum to the proceedings on Object-oriented programming systems, languages, and applications (OOPSLA), vol. 18, no. 22, 1992, pp. 29–30

HOMO LUDENS MORALIS: DESIGNING AND DEVELOPING A BOARD GAME TO TEACH ETHICS FOR ICT EDUCATION

Damian Gordon, Dympna O'Sullivan, Yannis Stavrakakis, Andrea Curley

Technological University Dublin (Ireland)

damian.x.gordon@tudublin.ie, dympna.osullivan@tudublin.ie,
ioannis.stavrakakis@tudublin.ie, andrea.f.curley@tudublin.ie

EXTENDED ABSTRACT

The ICT ethical landscape is changing at an astonishing rate, as technologies become more complex, and people choose to interact with them in new and distinct ways, the resultant interactions are more novel and less easy to categorise using traditional ethical frameworks. It is vitally important that the developers of these technologies do not live in an ethical vacuum; that they think about the uses and abuses of their creations, and take some measures to prevent others being harmed by their work.

To equip these developers to rise to this challenge and to create a positive future for the use of technology, it important that ethics becomes a central element of the education of designers and developers of ICT systems and applications. To this end a number of third-level institutes across Europe are collaborating to develop educational content that is both based on pedagogically sound principles, and motivated by international exemplars of best practice. One specific development that is being undertaken is the creation of a series of ethics cards, which can be used as standalone educational prop, or as part of a board game to help ICT students learn about ethics.

The use of games in teaching ethics and ethics-related topics is not new, Brandt and Messeter (2004) created a range of games to help teach students about topics related to design (with a focus on ethical issues), and concluded that the games serve to as a way to structure conversations around the topic, and enhance collaboration. Halskov and Dalsgård (2006), who also created games for design concurred with the previous researchers, and also noted that the games helped with the level of innovation and production of the students. Lucero and Arrasvuori (2010) created a series of cards and scenarios to use them in, and had similar conclusions to the previous research, but also noted that this approach can be used in multiple stages of a design process, including the analysis of requirements stage, the idea development stage, and the evaluation stage.

The aim of our work is to develop educational content for teaching ICT content. In this paper we present the development of a series of ethics cards to help ICT students learn about ethical dilemmas. The development of ethics cards has followed a *Design Science* methodology (Hevner *et al*., 2004) in creating the board game these guidelines were expanded into a full methodology that is both iterative and cyclical by Peffers *et al*. (2007). Our project is currently in the third stage of this methodology, called the "Design & Development" stage, but the process is evolving as the cards are being designing to act as independent teaching materials that can but used in the classroom, as well as part of the board game.

A sample set of cards are presented below. The cards can be used independently in the classroom, for example, a student can be asked to pick a random *Scenario Card*, read it out to the class, and have the students do a Think-Pair-Share activity. This is where the students first reflect individually on the scenario, then in pairs, and finally share with the class. Following this a *Modifier Card* can be selected, of which there are two kinds, (1) modifications that make the scenario worse for others if the student

doesn't agree to do the task on the Scenario Card, and (2) modifications that make the scenario better for others if the student does agree to do the task. This should generate a great deal of conversation and reflection on whether doing a small "bad task" is justifiable if there is a greater good at stake.

The cards can also be used in the board game where the players have a combination of Virtue, Accountability, and Loyalty points, which are impacted by both the Scenario Cards and the Modifier Cards. It is worth noting that some modifiers result in points being added on, others subtracted, and others multiplied to the players' global scores.

Overall the goal of this project is not simply to design a game to help teach ethics, but rather to explore how effective design science methodologies are in helping in the design of such a game.

Scenario Cards: Set 1

[10 points] **Scenario Card** ***You are asked to write a system that will capture location information without consent***	[10 points] **Scenario Card** ***You are asked to write software to control missiles***
[10 points] **Scenario Card** ***You are asked to develop AI with human-level intelligence***	[10 points] **Scenario Card** ***You are asked to write software for an autonomous car that will always protect the driver irrespective of the circumstances***
[10 points] **Scenario Card** ***You are asked to write code that will crack the license on a commercial software package***	[10 points] **Scenario Card** ***You are asked to write a comms system that will run on channels reserved for emergency services***
[10 points] **Scenario Card** ***You are asked to build a system that is a lot like an existing competitor's system, but it is "just for a demo"***	[10 points] **Scenario Card** ***You are asked to secretly change an accountancy program to change the way it does calculations***

Modifier Cards: Set 1

Bad outcome, if you don't [+2] **Modifier Card** ***If you don't do it, someone else will do it, who is a much, much worse programmer***	**Better outcome, if you do [-2]** **Modifier Card** ***If you do it, you are guaranteed that no one will ever find out it was you who wrote this code***
Bad outcome, if you don't [+5] **Modifier Card** ***If you don't do it, someone else will do it, who will make it more unethical***	**Better outcome, if you do [-5]** **Modifier Card** ***If you do it, you will be paid at least €2 million, and it will only take 2 weeks in total***
Bad outcome, if you don't [x2] **Modifier Card** ***If you don't do it, your organisation will fail and 200 people will lose their jobs***	**Better outcome, if you do [x2]** **Modifier Card** ***If you do it, your organisation will select a group of five very sick people at random and pay for all their health costs***
Bad outcome, if you don't [x5] **Modifier Card** ***If you don't do it, a chain of events will occur that will ruin the economy of your country for the next 15 years***	**Better outcome, if you do [x5]** **Modifier Card** ***If you do it, your organisation will donate at least €60 million to your favourite charity***

KEYWORDS: Digital Ethics; Card Games; Board Games; Design Science.

REFERENCES

Bochennek, K., Wittekindt, B., Zimmermann, S.Y. and Klingebiel, T. (2007) "More than Mere Games: A Review of Card and Board Games for Medical Education", *Medical Teacher*, 29(9-10), pp.941-948.

Brandt, E. and Messeter, J. (2004) "Facilitating Collaboration through Design Games". In Proceedings of the *Eighth Conference on Participatory Design: Artful Integration: Interweaving Media, Materials and Practices*, 1, pp. 121-131, ACM.

Deterding, S. (2009) "Living Room Wars" in Hunteman, N.B., Payne, M.T., *Joystick Soldiers* Routledge, pp.21-38.

Deterding, S., Khaled, R., Nacke, L.E. and Dixon, D. (2011) "Gamification: Toward a Definition", CHI 2011 Gamification Workshop Proceedings (Vol. 12). Vancouver BC, Canada.

Flatla, D.R., Gutwin, C., Nacke, L.E., Bateman, S. and Mandryk, R.L. (2011) "Calibration Games: Making Calibration Tasks Enjoyable by Adding Motivating Game Elements", in Proceedings of the 24th annual ACM Symposium on User Interface Software and Technology (pp. 403-412). ACM.

Halskov, K. and Dalsgård, P. (2006) "Inspiration Card Workshops" In Proceedings of the *Sixth Conference on Designing Interactive Systems*, pp. 2-11, ACM.

Hamari, J., Koivisto, J. and Sarsa, H. (2014) "Does Gamification Work? A Literature Review of Empirical Studies on Gamification", in *Hawaii International Conference on System Sciences*, 14(2014), pp. 3025-3034.

Hevner, A.R., March, S.T., Park, J. and Ram, S. (2004) "Design Science in Information Systems Research", *Management Information Systems Quarterly*, 28(1), p.6.

Kapp, K.M. (2012) *The Gamification of Learning and Instruction: Game-based Methods and Strategies for Training and Education*. Pfeiffer; San Francisco, CA.

Lloyd, P. and Van De Poel, I. (2008) "Designing Games to Teach Ethics", *Science and Engineering Ethics*, 14(3), pp.433-447.

Lucero, A. and Arrasvuori, J. (2010) "PLEX Cards: A Source of Inspiration when Designing for Playfulness", In Proceedings of the *Third International Conference on Fun and Games*, 1, pp. 28-37, ACM.

Peffers, K., Tuunanen, T., Rothenberger, M.A. and Chatterjee, S. (2007) "A Design Science Research Methodology for Information Systems Research, *Journal of Management Information Systems*, 24(3), pp.45-77.

Seaborn, K. and Fels, D.I. (2015) "Gamification in Theory and Action: A Survey", *International Journal of Human-Computer Studies*, 74, pp.14-31.

Smith, R. (2010) "The Long History of Gaming in Military Training", *Simulation & Gaming*, 41(1), pp.6-19.

Wells, H.G. (1913) *Little Wars*. London: Palmer.

Vego, M. (2012) "German War Gaming", *Naval War College Review*, 65(4), pp.106-148.

Zuckerman, O (2006) "Historical Overview and Classification of Traditional and Digital Learning Objects", MIT Media Lab.

THE PHILOSOPHY OF TRUST IN SMART CITIES

Dr. Mark Ryan

KTH Royal Institute of Technology, Stockholm

mryan@kth.se

EXTENDED ABSTRACT

Trust entails forming relationships with others to respond to risks by bridging the gap of uncertainty about others' actions. However, trust does not attempt to eliminate risk, but is an embracement and acceptance of it, one acknowledges that placing trust in someone creates a vulnerability on the part of the trustor (McLeod 2015). The trustor opens themselves up to the vulnerability of betrayal if their trust is breached. The level of betrayal also depends on the context of the relationship. Bonds of trust may be stronger in certain circumstances, contexts, and relationships with the trustee. For example, intuitively, we would have stronger trusting relationships to close friends and loved ones that we would to complete strangers (Luhmann 1979). However, there is still a degree of trust placed in strangers to behave in a certain way towards us, particularly in the public places such as the city (O'Neill 2002).

In a city, we trust strangers to behave in a certain way in line with societal norms, with the most paramount trust being that they do not harm or aggrieve us in anyway (Jones 1996). Cities have always been grounded on trust because without it, they would become chaotic and often cease to effectively function: 'In the context of the city, strangers live together and depend upon each other in their daily shared space with little certainty about how others will behave. Trust acknowledges the inherent distance between strangers and, simultaneously, their interdependency and vulnerability' (Keymolen & Voorwinden 2019, p. 3). It is this shared, collective, embracement of vulnerability that allows trusting relationships to form and flourish. Trusting others is not an attempt to eliminate complexity, but rather, it embraces it, despite the risks inherent in uncertainty. Therefore, we must accept a degree of risk when we place our trust in strangers within urban environments.

On the other hand, municipalities are facing strains on resources, infrastructure, and healthcare and transportation within cities, and are attempting to reduce the risks associated with failing to do so. In Europe alone, nearly 80% of the total population live in cities and as populations grow, there is an increasing demand for improved standards of living, while at the same time, the need to reduce congestion, pollution and environmental harm resulting from development. The smart city paradigm is an approach that claims to effectively address these concerns. Through the use of technological innovation, scientific knowledge, and entrepreneurial endeavour, the smart cities paradigm is being proposed as a way for cities to respond to the risks of continued urban development.

The smart city incorporates data-driven technologies and applications to establish predictive patterns to make cities more manageable and controlled. The aim is to reduce risk, improve functionality, and remove uncertainty. The smart city paradigm attempts to integrate and exploit the use of data, different technological ingenuity, and scientific methods, to remove vulnerability and to reduce the complexity of urban environments (Keymolen 2016). It is this predictive objective that threatens to undermine the trusting components that cities are founded upon. The smart city paradigm aims to replace interpersonal trusting relationships with technological systematic reliability. However, is this necessarily a bad thing to want to reduce risks within cities? Is the goal of risk and harm reduction not something praiseworthy, or will we lose something that defines us as humans, in the form of trust?

The focus of smart cities is to reduce complexity through intensive data analytics, technological infrastructure, and innovation. However, in the process it shifts the focus from having a trust in strangers to a reliability in systems and technologies. There is a shift from interpersonal trust between fellow human beings to a reliability in technological systems. The difference between interpersonal trust and system trust is that the former is a response to uncertainty and complexity of others, while the latter is a response to the complexity involved in the systems that we depend on (Keymolen and Voorwinden 2019, p. 8). If we replace interpersonal trust with a trust in systems, then it has the potential to disassociate and remove us from trusting interpersonal relationships to a reliance on organisational and technological systems. We are moving away from an embracement of our vulnerability as a result of complexity to an attempt to control this complexity through systematic and artificial means. It is a shift from societies built on interpersonal trust to ones defined by systematic reliance, moving from an embracement of trustworthiness to an expectation of reliability.

The smart city framework replaces trustworthiness with reliability, a replacement of a coming together of individuals to a reliance on systems. There is an erosion of individuals' vulnerability, because the smart city paradigm's goal is to reduce complexity to make things more manageable and controllable, rather than understanding and embracing humankind's essential uncertainty. There is a fundamental shift from cities that are grounded on trusting relationships to meet the challenge of uncertainty and complexity to one that is essentially reliant on urban technological development and data analytics to eliminate uncertainty. Within trust, there is an acceptance of vulnerability and risk, and it is not an avoidance or an annihilation of uncertainty that is seen in data-driven smart city ideologies.

This paper will question the smart city paradigm, the implementation of smart city technologies, and the effect that they have on trust. It will propose that we cannot *trust* technology at all, we can only *rely* on it, and that trust is kept for interpersonal relationship between human beings. Reliability can be grounded on past performance and predictions of future performance, but not on reasons that underpin definitions of trust, such as the affective (Jones 1996; Baier 1986) or normative intent (Lord 2017; O'Neill 2002; and Simpson 2012) of the trustee. Therefore, we are left with basing our trust in those who are developing, deploying and integrating these technologies within smart cities.

We can only trust the people who are behind smart city technologies and projects, but this raises a number of questions that this paper will address: can we even identify who is behind these developments, can we trust them, and should we? Is the control of complexity and reduction of risk necessarily a good thing, and is it worth the trade-off associated with trust reduction? Is vulnerability an important human property or something that we should try to eliminate with the help of smart cities? Is this replacement of interpersonal trust with technological reliability something that we should be concerned about? If so, what can be done to ameliorate these concerns and who should be held responsible?

KEYWORDS: Philosophy of trust; ethics of smart cities; vulnerability; risk; reliability of technology; and trustworthiness.

REFERENCES

Baier, Annette, "Trust and Antitrust", *Ethics*, Vol. 96, 1986, pp. 231-260.

Jones, Karen, "Trust as an Affective Attitude", *Ethics*, Vol. 107, Issue 1, 1996, pp. 4-25.

Keymolen, Esther, "Trust on the Line: A Philosophical Exploration of Trust in the Networked Era", *Erasmus University Rotterdam*, the Netherlands, [Thesis], 2016.

Keymolen, Ester, and Voorwinden, Astrid, "Can we Negotiate? Trust and the Rule of Law in the Smart City Paradigm", *International Review of Law, Computers & Technology*, 2019, https://doi.org/10.1080/13600869.2019.1588844

Lord, Carol, *Can Artificial Intelligence (AI) be Trusted? And does it Matter?* Masters' Dissertation at the University of Leeds: Inter-Disciplinary Ethics Applied Centre, September 4th 2017.

Luhmann, Niklas, *Trust and Power*, Chichester: John Wiley, 1979.

McLeod, Carolyn, "Trust", *Stanford Encyclopedia of Philosophy*, 2015, available here: https://plato.stanford.edu/entries/trust/

O'Neill, Onora, *Autonomy and Trust in Bioethics*, Cambridge University Press, UK, 2002.

Simpson, T.W., "What is Trust?" *Pacific Philosophical Quarterly*, Vol. 92, 2012, pp. 550-569.

TOWARDS A CONCEPTUAL FRAMEWORK FOR TRUST IN BLOCKCHAIN-BASED SYSTEMS

Mattis Jacobs

University of Hamburg (Germany)

jacobs@informatik.uni-hamburg.de

EXTENDED ABSTRACT

The ever-increasing societal permeation and societal impact of ICTs evoke calls for considering more than just instrumental values regarding their development, application, and regulation. In this spirit, public and private actors push for user-centered, fair, transparent, accountable, and trustworthy ICTs (Datenethikkommission, 2019; HLEG, 2019; Pichai, 2018). However, philosophical terms like 'fairness', 'transparency', 'accountability', and 'trustworthiness' are conceptually multifaceted and context-dependent, i.e., they need to be refined and adjusted in the context of the respective ICTs in order to allow for meaningful use. While there is substantial progress in some domains, e.g., fair, accountable, and transparent machine learning, other domains still struggle with developing a coherent conceptual framework.

Against the background of the emerging blockchain technology[1], especially trust and trustworthiness are prominently discussed issues. The blockchain technology – according to its advocates – enables trust-free trans- and interactions in contexts that traditionally required either direct interpersonal trust between the interacting parties or a trusted intermediary (Beck, Czepluch, Lollike, & Malone, 2016; Swan, 2015). For instance, in the context of cryptocurrencies, the Bitcoin Whitepaper (Nakamoto, 2008, p. 8) suggests that the Bitcoin Blockchain is a system for electronic transactions that do not rely on trust. As long as only the closed ecosystem is concerned, many scholars follow this line of argumentation (Hawlitschek, Notheisen, & Teubner, 2018, p. 59). Critics – on the other hand – point at existing vulnerabilities to question the notion of a trust-free technology (Christopher, 2016; Walch, 2018; Werbach, 2018). While indeed traditional trust models like interpersonal or intermediary trust loose importance (if not vanish) – so they argue – new trust relationships and even a new form of trust gain relevance: trust that "significantly distinguishes itself from the more traditional typology of trust" (Swan & Filippi, 2017, p. 605) that is traditionally understood as a bilateral concept between two agents. According to this line of thought, the blockchain technology enables users to distribute trust over various actors whose individual trustworthiness they cannot and do not need to assess.

This new trust environment is described mainly from two perspectives. First, from a perspective that queries to what extent it differs concerning the setup of actors from 'traditional' modes of trust like interpersonal trust or trust in institutions and organizations such as governments and (central-) banks who serve as trusted intermediaries. According to these elaborations, distributed trust within blockchain-based systems differs in that users place it in different entities based on different assurances than in more traditional and well-studied institutional setups (Werbach, 2018, p. 30). Antonopoulos (2014) adds that in contrast to other approaches in security-related research in computer science, blockchain's approach to create trustworthiness does not depend on excluding non-trustworthy actors, but is based on mechanisms that render their actions inconsequential. The second perspective considers the technical environment in which the alleged new form of trust can develop.

[1] The term 'blockchain' in this abstract only refers to open/permissionless systems.

Scholars here explain the underlying protocols like proof-of-work and proof-of-stake and outline how the interactions enabled by these protocols support actors in (collectively) generating assurances that evoke trustworthiness (Hawlitschek et al., 2018; Mallard, Méadel, & Musiani, 2014; Mehrwald, Treffers, Titze, & Welpe, 2019).

However, while shedding light on some aspects of how the blockchain technology enables a new model of trust, these elaborations fall short of delivering a coherent conceptual framework, which incorporates what the term 'trust' in this context actually means. As mentioned at the outset, the terms 'trust' and 'trustworthiness' are context-dependent. Trust is a reducible concept, not a primitive (Hardin, 2002, p. 57), i.e., trust consists of a "set of other notions to which it is to be reduced" (Hardin, 2002, p. 88). While some components – or, in Hardin's words, "notions" – are constituting for almost any account of trust, e.g., "some possibility of misplaced trust – and some sense of expectations of another's behaviour" (Hardin, 2002, p. 88), others are only relevant in specific domains. For instance, some accounts of interpersonal trust require the consideration of motivational factors such as if the trusted person (the so called 'trustee') presumably has a positive attitude towards the trusting person (the so called 'trustor') (Baier, 1986). Yet, these factors are evidently not applicable in the context of trust in entities that do not have the capability to have a motivation (e.g., trust in technological systems; cf. Nickel, 2013) or in settings that presume actors to be rational agents who act solely based on self-interest (e.g., trust in game-theoretical settings; cf. Gambetta, 1988). Furthermore, while many accounts of trust are based on prior relationships and the resulting attribution of character traits, accounts of trust in, for instance, reputation systems (Botsman, 2017) are not. They focus on mechanisms that allow to distribute trust and hence can circumvent the need to make character judgements on any individual entity and enable trusting groups of strangers without having built up any personal relationship with any actor within this group beforehand. The elaborations in the discourse on blockchain and trust, as well as trustworthiness, so far omit to reflect on these issues. They do not address the question of which of the components that constitute trust according to the vast set of accounts of trust matter in the context of blockchain-based systems.

The goal of this paper is to address this conceptual void. It outlines how the assemblage of the different components and requirements, which, according to different accounts, constitute trust are put together in the setup that blockchain-based systems engender. The starting point is an analysis of the assemblages of components of familiar accounts of trust in distributed settings. Here, for example, accounts of trust in markets, accounts of trust in reputation systems, and accounts of trust in the 'wisdom of the crowd' are considered. Originating from the collection of components relevant in the respective contexts, the actor- and technological-based analyses of blockchain's trust model are adduced to compare in which regard the setup engendered by blockchain-based systems is comparable to the aforementioned other distributed setups. This allows to determine to what extent the components of trust in the respective settings are applicable in the context of blockchain-based systems, too, and, if necessary, which other components need to be additionally introduced to give a full picture. This contribution allows to develop a common understanding of the term 'trust' within this interdisciplinary discourse and hence supports building the conceptual basis for developing trustworthy blockchain-based systems as well as related guidelines.

KEYWORDS: trust; trustworthiness; blockchain; bitcoin; DLT.

ACKNOWLEDGEMENTS: I acknowledge financial support from Hamburg Ministry of Science, Research and Equality in the project Information Governance Technologies.

REFERENCES

Antonopoulos, A. (2014). Bitcoin security model: Trust by computation. *Forbes. com, February*, *20*. Retrieved from http://radar.oreilly.com/2014/02/bitcoin-security-model-trust-by-computation.html

Baier, A. (1986). Trust and Antitrust. *Ethics*, *96*(2), 231–260. https://doi.org/10.1086/292745

Beck, R., Czepluch, J. S., Lollike, N., & Malone, S. (2016). Blockchain-the Gateway to Trust-Free Cryptographic Transactions. In *ECIS*.

Botsman, R. (2017). *Who can you trust? How technology brought us together and why it might drive us apart* (First edition (eBook)). New York: PublicAffairs.

Christopher, C. M. (2016). The Bridging Model: Exploring the Roles of Trust and Enforcement in Banking, Bitcoin, and the Blockchain. *Nevada Law Journal*, *17*, 139.

Datenethikkommission. (2019). *Gutachten der Datenethikkommission.*

Gambetta, D. (1988). Can We Trust Trust? In D. Gambetta (Ed.), *Trust: Making and breaking cooperative relations* (pp. 213–237). Oxford: Blackwell.

Hardin, R. (2002). *Trust and trustworthiness*. *The Russell Sage Foundation series on trust: volume 4*. New York: Russell Sage Foundation. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&AN=1069635

Hawlitschek, F., Notheisen, B., & Teubner, T. (2018). The limits of trust-free systems: A literature review on blockchain technology and trust in the sharing economy. *Electronic Commerce Research and Applications*, *29*, 50–63. https://doi.org/10.1016/j.elerap.2018.03.005

HLEG, A. I. (2019). *Ethics guidelines for trustworthy AI.*

Mallard, A., Méadel, C., & Musiani, F. (2014). The paradoxes of distributed trust: Peer-to-peer architecture and user confidence in Bitcoin. *Journal of Peer Production*. (4), 1–10.

Mehrwald, P., Treffers, T., Titze, M., & Welpe, I. (2019). Blockchain Technology Application in the Sharing Economy: A Proposed Model of Effects on Trust and Intermediation. *Proceedings of the 52nd Hawaii International Conference on System Sciences*, 4585–4594.

Nakamoto, S. (2008). *Bitcoin: A peer-to-peer electronic cash system.*

Nickel, P. J. (2013). Trust in Technological Systems. In M. J. Vries, S. O. Hansson, & A. W.M. Meijers (Eds.), *Philosophy of Engineering and Technology: Vol. 9. Norms in Technology* (pp. 223–237). Dordrecht: Springer.

Pichai, S. (2018). AI at Google: our principles. Retrieved from https://www.blog.google/technology/ai/ai-principles/

Swan, M. (2015). *Blockchain: Blueprint for a New Economy*. Beijing: O'Reilly.

Swan, M., & Filippi, P. de (2017). Toward a Philosophy of Blockchain: A Symposium: Introduction. *Metaphilosophy*, *48*(5), 603–619. https://doi.org/10.1111/meta.12270

Walch, A. (2018). In Code(rs) We Trust: Software Developers as Ficuciaries in Public Blockchains. *SSRN Electronic Journal*.

Werbach, K. (2018). *The blockchain and the new architecture of trust*. *Information policy series*. Cambridge, Massachusetts, London, England: The MIT Press.

VIRTUOUS JUST CONSEQUENTIALISM: EXPANDING THE IDEA MOOR GAVE US

Olli I. Heimo, Kai K. Kimppa

University of Turku (Finland)

olli.heimo@utu.fi, kai.kimppa@utu.fi

EXTENDED ABSTRACT

As an IT professional, one has power over others through the decisions one makes. These decisions do not only create possibilities to create value through work or entertainment, but also value through moral decisions by allowing or limiting the growth of the users' characters. The decisions made in the system design (e.g. UI, functionalities, and communication methods), when designed correctly, can affect the character building process of the user by allowing, denying, and most importantly of all supporting certain actions. The IT professional can hence be in the position of a virtual virtue friend (Heimo et al., 2018) – for a person whom they will very probably never meet nor whom will never hear about the professional – and support the virtuousness of that user through the decisions they make in the design. Yet being virtuous is a sort of vague moral guideline – especially when these persons do not know much about each other, and therefore more specific instructions and short time aims should be clarified.

Hence Moor's just consequentialism which in turn can evaluate specific situations in everyday life. It is rather easy compared to the virtuous to see and examine ones motivations and consequences. Thus we want to mate Aristotle (even though through more modern interpretations) to Moor. One can make a virtue out of being Moorean by making a habit of having ones motivations just, and evaluating ones actions whether they have just consequences. To extend this to the role of an IS/IT designer to be a "virtue friend" in a "virtual world", to support the users' possibilities in acting for just motivations and just end results (on virtual friendship elsewhere see e.g. Briggle, 2008 or Elder, 2014, although their handling is more direct than ours). However as we clarify, more in our full this does not mean forcing the motivations nor forcing the desired consequences but, rather, as a friend supporting others develop their character to more Morean view.

James Moor (1999) in his paper on Just consequentialism and computing says that he is approaching the topic from both deontological and consequentialist perspectives – and thus he indeed does. He uses a Rawlsian approach of justice and consequentialist considerations to build a framework through which a developer can evaluate the function of their application. If an action is both justified and its consequences are good, the function of the application is also good. In this paper we claim this is not quite enough. We intend to show that combined with an interpretation on friendship from Aristotle the argument can be strengthened.

According to Aristotle, to reach Eudaimonia, one must be virtuous in their everyday life. A good person fulfils his or her telos by avoiding vices, achieving their virtues, and most of all, developing their character – by flourishing in what they do. Implementing virtues is part of everyday life, and doing so develops character, which is a sum of a person's deeds. By following virtues, acting virtuously becomes a natural aspect of ones actions so that the character develops itself to be virtuous. (Heimo, 2018; Heimo et al., 2018.) Vallor (2013) states that moral skills are necessary for moral virtue:

> *Someone could have moral skills in the sense of practical moral knowledge but fail to be virtuous because they are unreliable in acting upon this knowledge, or because they act well only for nonmoral reasons. Still, moral skills are a necessary if not a sufficient condition for moral virtue. Without the requisite cultivation of moral knowledge and skill, even a person who sincerely wishes to do well consistently and for its own sake will be unsuccessful.*

Virtue ethics does not generate a set of norms for normative ethics, to improve both people and society. Moreover virtue ethics generates guidelines both by promoting the virtuous actions and by encouraging people to avoid vice. The normativity comes from the level of ideas rather than from a strict set of rules. According to Aristotle, only through virtues can a person generate true value and vice versa with vices. Humans should aim for a virtuous life, which in the Aristotelian sense requires aiming for their telos, having virtuous friends, and navigating vices to arrive at virtue. A character is not virtuous by following virtue alone, since one might follow virtue reluctantly and in the face of temptation. Rather, when a person automatically aims toward all virtues, the character can become virtuous. (EN I, 9 – 10; 1098a, 15 – 21; 1098b 5 – 30; 1100a31 – 1101a21, II, 1; 1103a31 – 1103b25, 1104a10 – 1105a16; McPherson, 2013). Or as Vallor (2009) explains:

> *[…] the moral development of individuals cannot be assessed or predicted simply by looking at what they think, feel or believe—we also have to know what kinds of actions they will get in the habit of doing, and whether those actions will eventually promote in such persons the development of virtues or vices.*

Being virtuous then is not a situational choice but a life choice, and only through a life choice can a human enjoy a happy and good life. Yet socially valued virtues might not equal ethical virtues (Beauchomp & Childress 2001, p. 27). At work, humans are often expected to follow socially valued virtues even though they conflict with their moral virtues (e.g., Murphy, 1999). Being good at one's work does not equal being virtuous. The totality of human life, which is not divisible into parts that can then ignore other parts, needs to be taken into account and built virtuously (MacIntyre, 2004, pp. 240 – 241; 2007, p. xv).

So that we can evaluate actions in virtue ethics we must interpret Aristotle in such a way that it is more universal and not only focused on the character of the evaluator. Aristotelian virtue ethics is focused on ones' self, ones character, and thus the evaluations are easier to do because the evaluator knows all, or at least most of the intentions or perceived consequences of the actions. Whereas when evaluating other peoples' behaviour we are within a limited set of knowledge. In book IX of Nichomachean Ethics, however Aristotle introduces the idea of virtue friends. In addition to numerous kinds of qualities of friendship these virtue friends possess, they possess the two-way communication with us in promoting virtues in each-other. Hence to think how we should treat other people ourself, and how we wish other people to treat us, the idea of promoting the virtues in each-other seems to be a valid one.

However, it seems to be obvious that we do not want to treat everyone as we treat our closest friends. We do not want to share our lives, our health (see e.g. Wahlstrom, Fairweather & Ashman, 2011) our possessions, our time with everyone, but with a selected group of individuals we consider close and trustworthy. Yet, to promote the virtues from others seems to be good from the viewpoint of what is just, and from the viewpoint of where the consequences, at least in the large scale, could be beneficial.

Thus as an IT professional being a virtue friend to the users – those dependent on their decisions – seems just and virtuous. To support them, not to force them nor be indifferent about them, to make a habit of having their motivations just, evaluating their actions through not just motivations but also

through the consequences and steer their habits towards what they themselves have learned and discovered to be just. As friends treat others whose virtuous actions they try to support, so that their motivations become just and habitual, so that the consequences of their acts be just and the promethean values meet the epimethian thinking to promote the habit of creating iterations of more just consequences.

In the full paper we will discuss more about Moor (1999) - and what is just and consequently acceptable via virtue ethics and virtue friendship.

KEYWORDS: Virtue ethics, Just Consequentialism, Ethics, Aristotle, Moor.

ACKNOWLEDGEMENTS: We want to especially thank the reviewers for their insightful comments on 1) how is this actually IT and ethics, 2) how does virtue ethics then tie into Moor's model on just consequentialism, and 3) on how to fit different virtue ethics researchers into the model we propose.

REFERENCES

Aristotle (NE). (Circa 350 BCA). Nicomachean ethics. Several translations used.

Beauchamp, T. L., & Childress, J. F. (2001). Principles of biomedical ethics. Oxford University Press, USA.

Briggle, Adam (2008) Real friends: how the Internet can foster friendship, Ethics and Information Technology, 10:71-79.

Elder, Alexis (2014) Excellent online friendships: an Aristotelian defense of social media, Ethics and Information Technology, 16:287-297.

Heimo, Olli I. (2018). *Icarus, or the idea toward efficient, economical, and ethical acquirement of critical governmental information systems.* Ph.D. Thesis, University of Turku. https://www.utupub.fi/handle/10024/146362

MacIntyre A. (2004). Hyveiden jäljillä: Moraaliteoreettinen tutkimus. (2nd ed., After Virtue: A Study in Moral Theory. Translated by N. Noponen) Gaudeamus, Helsinki.

McPherson D. (2013). Vocational Virtue Ethics: Prospects for a Virtue Ethic Approach to Business. Journal of Business Ethics, 116(2), 283-296.

Moor, J. H. (1999). Just consequentialism and computing, Ethics and Information Technology. 1: pp. 65–69.

Murphy, P. E. (1999). Character and virtue ethics in international marketing: An agenda for managers, researchers and educators. Journal of Business Ethics, 18(1), 107-124.

Vallor, S. (2009), Social networking technology and the virtues, *Ethics and Information Technology* (2010) 12:157-170, Published online: 11 August 2009 Springer Science+Business Media B.V. 2009, DOI 10.1007/s10676-009-9202-1

Vallor, S. (2013) The future of military virtue: Autonomous systems and the moral deskilling of the military, 2013 5th International Conference on Cyber Conflict (CYCON 2013), Tallinn, 2013, pp. 1-15.

Wahlstrom, Kirsten, Fairweather, N. Ben & Ashman, Helen (2011) Brain-Computer Interfaces: a technical approach to supporting privacy, Ethicomp 2011.

2. Cyborg: A Cross Cultural Observatory

Track chairs: Mario Arias-Oliva, Universitat Rovira i Virgili, Spain – Jorge Pelegrín-Borondo, University of La Rioja, Spain

A BRIEF STUDY OF FACTORS THAT INFLUENCE IN THE WEARABLES AND INSIDEABLES CONSUMPTION IN MEXICAN SOCIETY

Juan Carlos Yáñez-Luna, Pedro Isidoro González Ramírez, Mario Arias-Oliva, Jorge Pelegrin-Borondo

Universidad Autónoma de San Luis Potosí, (México), Universidad Autónoma de San Luis Potosí, (México), Universitat Rovira i Virgili (Spain), Universidad de La Rioja (Spain)

jcyl@uaslp.mx; pedro.gonzalez@uaslp.mx; mario.arias@urv.es; jorge.pelegrin@unirioja.es

EXTENDED ABSTRACT

Most people that were born in the las century were able only to believe that technology could be an important part of people´s life. An important challenge of cyclical markets in the world is to differentiate the aims factors that have influence in the consumer behaviour. According with Fresneda Lorente (2019), the consumerism of electronic technologies will be increased in the 2020, a total of 20% of technology revenue is expected.

In this respect, globalization has enabled technology industries increase production and trade throughout the world. Most of the technological gadgets are made up in developing countries such as China, Taiwan, Chile, Brazil, etc., this allows to reduce manufacturing costs and increase its production in order to seek greater competitiveness. In the case of technologies focused to mobility such as Smartphones, gadgets, etc., every year companies develop new and sophisticated technologies with Internet as a common media. Companies are also working in the topics of Artificial intelligence (McLean & Osei-Frimpong, 2019), also in wearables (Fröbel, Avramidis, & Joost, 2019) and insideables or implants (Haeberle et al., 2019). Most of them requires Internet or the Smartphone to work adequately, but recent technologies works with, a set of data who interpret the environment and take appropriate decision; well known as Artificial Intelligence (AI). We can imagine a near future in which the use of devices with AI that will improve the human disabilities such as physical or mental defects through a set of microcircuits implanted and managed (or not) by external devices like wearables.

The acceptance of technology for improve the human abilities or disabilities is a complicated topic specially in social context. In the one hand, many individuals believe in the use of technology for transform their lives and to increase their welfare, on the other, many people make their lives in a strict order based in culture, religion or others social structures. In this regard, marketers, economists or decision takers in the business and government should study that more this topic in order to take steps that affect their economy.

Most of the consumerism in technologies focuses on wearables for health (52% of sales of wearables in the world) and the 39% will represent to health gadgets, such as Fitness bracelets or Smartwatches.

According to Garibay (2018), in Mexico 51.9% of individuals adopted and uses at least 3 gadgets; Likewise, the report of Interactive Advertising Bureau México (2019) shows that the acceptance of wearables and virtual reality grew up at least 15% in relation with previous years. We can assume that the penetration of technologies especially mobile or internet based, will continue growing exponentially, and it is possible that the consumers behaviour changes in the future. In relation with previous cited the penetration of wearables in the world has increased to 38% for people between 25-34 years old (Escamilla, 2019).

The implants business in Mexico focus mainly in Cosmetic and Health context. According with El Universal (2016) and Expansión (2019) Mexico is in the 4th place in the Breast Implant ranking, and according with El Debate (2017) in 2015 a 900,000 cosmetic implants were released in the country. The technological implants in Mexico is not common as other kind of surgical intervention, this may due to certain factors such as, expensive technology, expensive surgeries, lack of knowledge about the topic or culture and religion impediments.

The consumption of products is a fundamental part for jump-starting the markets in order to rise an economic development sustainable, however most of Mexican do not have the financial solvency for acquiring forefront technology (gadgets – wearables, implants or mobile) due principally to additional duties of importation and foreign i+D added costs. Most of the technology acquired in Mexico is imported from different countries in which has trade agreements.

In their last meeting (in México), the OECD countries established certain objectives in order to increase the digital transformation of services in each country (OECD, 2017). In the case of México, the amount for invest in Technology and Innovation is less than 1% of the Gross Domestic Product (GDP) in comparison with others OECD countries that invest more than 20% (Camhaji, 2017). This situation leads to economic stagnation and under development. Consequently, the growth of the country will diminished for a lack of knowledge development; and as is augmented in Cabrero Mendoza (2017): "The knowledge-based economy refers to the ability to generate scientific and technological knowledge, which allows to be more competitive, grow more, and transform the economy to achieve higher levels of social welfare".

Mexican government approved fiscal incentives in order to facilitate the consumerism of technology in all economic sectors. Those incentives could provide facilities to companies for save almost 94% of the investment in technology (Neuman, 2017). But for individuals to acquire forefront technology is still expensive, Mexican (specifically youngsters from mid-sized class) opt for purchasing cheaper technology such as low range wearables. In spite of cuts in the budget, Universities and Research Institutions in Mexico have been working in i+D. The principal aims in the research is the generation of biomaterials that could impact in the individuals' needs (Manjarrez Nevárez et al., 2017).

The proposal of this research is to analyse the perceptions about the acceptance of wearables or technological implants by Mexican citizens. We also consider how the transhumanism concept influences in the consumer consumption of technologies and how influences in their ethical behaviour.

KEYWORDS: Technology ethics, wearables, electronic implants, technology consumption.

REFERENCES

Cabrero Mendoza, E. (2017). ¿Dónde está México en ciencia y tecnología? Recuperado de http://www.jornada.unam.mx/2017/10/02/opinion/030a1pol

Camhaji, E. (2017). La ciencia, la oportunidad que México ha dejado pasar. Recuperado el 6 de enero de 2018, de https://elpais.com/elpais/2017/12/01/ciencia/1512157927_534452.html

El Debate. (2017). México , paraíso de cirugías estéticas. Recuperado el 8 de enero de 2018, de https://www.debate.com.mx/mexico/Mexico-paraiso-de-cirugias-esteticas-20170503-0366.html

El Universal. (2016). México, cuarto lugar mundial en cirugías de aumento de busto. Recuperado el 8 de enero de 2018, de http://www.eluniversal.com.mx/articulo/nacion/sociedad/2016/03/30/mexico-cuarto-lugar-mundial-en-cirugias-de-aumento-de-busto

Escamilla, O. (2019). ¿CÓMO SE ENCUENTRA EL MERCADO DE LOS. Recuperado de https://www.merca20.com/mercado-de-los-wearables/

Expansión. (2019). La Cofepris lanza alena por una marca de lmplantes llgada a un tlpo de cáncer. Recuperado de https://bit.ly/2kP1r8O

Fresneda Lorente, C. (2019). El gasto en tecnología de la información crecerá más de un 3 % hasta 2020. Recuperado de https://es.weforum.org/agenda/2017/03/el-gasto-en-tecnologia-de-la-informacion-crecera-mas-de-un-3-hasta-2020

Fröbel, F., Avramidis, E., & Joost, G. (2019). Workshop on Wearables and Machine Learning: Applications of Artificial Intelligence , Approaches on Textile Technology. En *Cooperation International Conference in HCI and UX* (pp. 177–181).

Garibay, J. (2018). ¿Será 2018 un buen año para los Wearables? Recuperado el 5 de enero de 2018, de https://www.merca20.com/sera-el-2018-un-buen-ano-para-los-wearables/

Haeberle, H. S., Helm, J. M., Navarro, S. M., Karnuta, J. M., Schaffer, J. L., Callaghan, J. J., … Ramkumar, P. N. (2019). Artificial Intelligence and Machine Learning in Lower Extremity Arthroplasty: A Review. *Journal of Arthroplasty*, 3–5. https://doi.org/10.1016/j.arth.2019.05.055

Interactive Advertising Bureau México. (2019). *Estudio de Consumos de Medios y Dispositivos entre internautas Mexicanos*.

Manjarrez Nevárez, L. A., Terrazas Bandala, L. P., Zermeño Ortega, M. R., De la Vega Cobos, C., Zapata Chávez, E., Torres Rojo, F. I., … Lerma Gutiérrez, R. (2017). Biomateriales como Implantes en el Cuerpo Humano. Recuperado el 8 de enero de 2018, de http://beta.uach.mx/articulo/2017/10/20/biomateriales-como-implantes-en-el-cuerpo-humano/

McLean, G., & Osei-Frimpong, K. (2019). Hey Alexa … examine the variables influencing the use of artificial intelligent in-home voice assistants. *Computers in Human Behavior*, *99*(May), 28–37. https://doi.org/10.1016/j.chb.2019.05.009

Neuman, G. (2017). Invierte en tecnología y deduce hasta un 94 %. ¡ Deducir o no deducir , esa es la...! Recuperado el 6 de enero de 2018, de https://www.pulsopyme.com/inviertir-tecnologia-deduce/

OECD. (2017). *OECD Digital Economy Outlook 2017*. Paris: OECD Publishing. http://doi.org/10.1787/9789264276284-en

CYBORG ACCEPTANCE IN HEALTHCARE SERVICES: THEORETICAL FRAMEWORK

Ala' Al-Mahameed, Mario Arias-Oliva, Jorge Pelegrín-Borondo

Universitat Rovira i Virgili (Spain); Universitat Rovira i Virgili (Spain); La Rioja University (Spain)

ala.almahameed@estudiants.urv.cat , mario.arias@urv.cat, jorge.pelegrin@unirioja.es

EXTENDED ABSTRACT

The importance of emerging technologies is growing. Among the new technologies that will be available in the market, insideables and implantable technologies, which are gaining more attention. Meanwhile, the boundaries between human and machine is becoming unclear, as technology become close to be embedded within human body (Britton & Semaan, 2017). In addition to that, the innovation in biomedicine, genetics, robotics and nanotechnology are making it possible to produce hybrid bodies that combines biological and technological parts (Kostrica, 2018; Triviño, 2015). The body-altering techniques are used to produce the "cyborg", which could be defined as a cognitively or bodily enhancement of human-being. This enhancement could be categorized into the following types (Greguric, 2014):

A. Cognitive abilities enhancement: such as infrared vision, memory enhancement, decision making and sensory perception, by using technological implants or wearable technologies.
B. Physical capabilities enhancement: such as strength, stamina and accuracy, by using bionic technology, genetic engineering and pharmacology.

On the other hand, reducing the size of the electronic components has introduced the nanotechnology, which stimulate the idea of creating small devices that can be implanted into human body to improve human physical and cognitive capabilities. These devices are called "Nanoimplants" (Pelegrín-Borondo, Reinares-Lara, Olarte-Pascual, & Garcia-Sierra, 2016; Reinares-Lara, Olarte-Pascual, Pelegrin-borondo, & Pino, 2016).

Nowadays, market already have different types of Cyborg technology that could be attached into human body through surgeries, wearables, pharmaceutical compounds and technological implants. The expectations regarding the cyborg market are promising for a reputable business with a potentially significant impact on future technologies and human societies (Pelegrín, Arias, Murata, & Souto, 2018). Some of these enhancements are already accepted by society, like the surgeries, wearables and pharmaceutical. While the technological implants for increasing the innate human capacity is partially accepted. Research in this area is required in order to formulate a complete picture about users' acceptance of these technologies. In other words, the acceptance of becoming a cyborg is still under investigation as the technology itself is under development (Reinares-Lara, Olarte-Pascual, & Pelegrín-Borondo, 2018). While, the aim of this research is to investigate the acceptance of cyborg as an entity in society, which is still under development as a technology, and nothing is known yet about how humans will perceive cyborgs when they will arrive. Our research aim is to research about the acceptance of cyborg services compared to human services, focusing in healthcare services. We will develop a theoretical framework that could be used to identify the choice criteria among these types of services.

Cyborg is an outcome of the technological innovation. Because of this reason, we consider that to review literature related to the acceptance of new technologies (e.g. robots' acceptance and cyborg technologies acceptance). In this context, the following theories and models have been used to develop the suggested structural model:

1. Technology Acceptance Model (TAM1) for Davis (1985) and its extensions TAM2 (Venkatesh & Davis, 2000) and TAM3 (Venkatesh & Bala, 2008).
2. The Unified Theory of Acceptance and Use of Technology (UTAUT1) for Venkatesh et al.(2003) and its extension UTAUT2 Venkatesh et al. (2012).
3. Cognitive-Affective-Normative Model (CAN) for Pelegrín-Borondo et al. (2016), which has been built to study the acceptance of being cyborg.

The proposed structural model includes Perceived Usefulness, Perceived Ease of Use, Perceived Risk, Trust, Social Influence, Empathy, and Emotions as the determinants of cyborg acceptance in the healthcare service industry.

In healthcare services sector, different studies have been studying the acceptance of new technologies by customers, applying the already cited models and theories, such as the acceptance of electronic health systems (e-health), mobile health services (m-health) and health information systems. Some studies in literature are supporting the Perceived Usefulness as the most significant determinant of the intention toward these technologies if compared to Perceived Ease of Use (Alsharo, Alnsour, & Alabdallah, 2018; Chang, Pang, Michael Tarn, Liu, & Yen, 2015; Sezgin, Özkan-Yildirim, & Yildirim, 2017) and the Social Influence as well (Bawack & Kamdjoug, 2018; Chu et al., 2018; Hossain, Quaresma, & Rahman, 2019). Furthermore, they have been used in studying the acceptance of wearable technologies for healthcare applications (Li, Wu, Gao, & Shi, 2016; Yang, Yu, Zo, & Choi, 2016) and in the electronic exchange of information across healthcare sector too (Ahadzadeh, Pahlevan Sharif, Ong, & Khong, 2015; Chu et al., 2018).

Each person is a member in their social entity. Therefore, other members' opinions and advices influences any behavior or decision. Social influence was introduced by the Theory of Reasoned Action (TRA) for Fishbein and Ajzen (1975) and the Theory of Planned Behavior (TPB) for Ajzen (1991). And it has been used in the technology acceptance models (Davis, 1989; Venkatesh, 2000). As well, it showed a significant impact on the acceptance of Nanoimplants (Pelegrín-Borondo, Reinares-Lara, & Olarte-Pascual, 2017; Pelegrín-Borondo et al., 2016; Reinares-Lara et al., 2018, 2016), breast augmentation for young women (Moser & Aiken, 2011) and on the acceptance of virtual customer integration (Füller, Faullant, & Matzler, 2010).

Some authors have pointed out to the importance of Perceived Risk in human-robot interaction. They claimed that when user's perception about risk is bigger than expected benefits, they could avoid the use of robots at all. However, the risk impact has been assessed through other dimensions (e.g. Trust). But the existing gap requires to investigate the impact of this construct by itself and through extending the conceptual models, including Perceived Risk in the intention toward such technologies (Blut, Wünderlich, & Brock, 2018).

In general, humans will start to use the perceptual cues and former experiences to classify an object (e.g. Human and Cyborg) and to effectively expect its behavior. In this stage, humans already recognize the abnormality of the other human, from the physical structure (e.g. wearables) or through their behaviours (e.g. implants). This stage is very important to avoid falling in "Uncanny Valley", in which the human will feel with unfamiliarity interacting with human-like objects (Stein & Ohler, 2017). Meanwhile, empathy and emotions can overcome the uncanny valley's negative outcomes. Emotions

have been considered as a way to distinguish humans from objects and machines. Furthermore, the ability to express basic emotions is a proof of humanity (Heisele, Serre, Pontil, Vetter, & Poggio, 2002). For instance, in the Cognitive-Affective-Normative (CAN) model, which was developed by Pelegrín-Borondo et al. (2016) to study the intention behavior toward being cyborg, the authors used the emotional dimension in their research model and found it as a significant predictor of the intention to become a Cyborg. Emotions are integrated in personal life, which have a significant impact on people perceptions, behaviours, beliefs and cognitive processes (Kasap & Magnenat, 2007). Originally, uncanny valley theory was introduced by Mori (1970) to propose the relation between human-likeness and familiarity while dealing with industrial robots. The theory proposed that, in some point (First Peak), maximum familiarity would be achieved once the robots become a human-like in terms of behavior and appearance. Furthermore, motion will enhance familiarity perception. However, the author pointed out to the feel of strangeness that could drop familiarity to the negative portion, which is representing the "Uncanny Valley"

Trust factor is introducing itself as a vital player in human-robot interaction context. It represents a psychological state of trustor about willingness and ability of trustee to help and cooperate in attaining trustor goals. Regarding to the research subject, human represents the trustor, Cyborg could represent the trustee, and healthcare service represents the goals (Brule, Dotsch, Bijlstra, Wigboldus, & Haselager, 2014).

On the other hand, the research suggests to include few open questions within the survey questionnaire to collect qualitative information, in order to gain an inductive knowledge from the participants, because it is a new subject and almost nothing known about human perception of the proposed shift in human-being structure and future.

KEYWORDS: Cyborg, Nanotechnology, Technology Acceptance, Healthcare Services.

REFERENCES

Ahadzadeh, A. S., Pahlevan Sharif, S., Ong, F. S., & Khong, K. W. (2015). Integrating Health Belief Model and Technology Acceptance Model: An Investigation of Health-Related Internet Use. *Journal of Medical Internet Research*, *17*(2), 1–27. https://doi.org/10.2196/jmir.3564

Ajzen, I. (1991). The Theory of Planned Behavior. *Organizational Behavior and Human Decision Processes*, *50*, 179–211. https://doi.org/10.1016/0749-5978(91)90020-T

Alsharo, M., Alnsour, Y., & Alabdallah, M. (2018). How Habit Affects continuous Use: Evidence from Jordan's National Health Information System. *Informatics for Health & Social Care*, 14. https://doi.org/10.1080/17538157.2018.1540423

Bawack, R. E., & Kamdjoug, J. R. K. (2018). Adequacy of UTAUT in Clinician Adoption of Health Information Systems in Developing Countries: The Case of Cameroon. *International Journal of Medical Informatics*, *109*, 15–22. https://doi.org/10.1016/j.ijmedinf.2017.10.016

Blut, M., Wünderlich, N. V, & Brock, C. (2018). Innovative Technologies in Branded-Service Encounters: How Robot Characteristics Affect Brand Trust and Experience. In *Thirty Ninth International Conference on Information Systems*. San Francisco. Retrieved from https://aisel.aisnet.org/cgi/viewcontent.cgi?article=1126&context=icis2018

Britton, L. M., & Semaan, B. (2017). Manifesting the Cyborg via Techno-Body Modification : From Human Computer Interaction to Integration. In *CSCW '17 Companion* (pp. 2499–2510). Portland, Oregon: ACM. https://doi.org/10.1145/3025453.3025629

Chang, M. Y., Pang, C., Michael Tarn, J., Liu, T. S., & Yen, D. C. (2015). Exploring User Acceptance of an E-hospital Service: An Empirical Study in Taiwan. *Computer Standards and Interfaces*, *38*, 35–43. https://doi.org/10.1016/j.csi.2014.08.004

Chu, X., Lei, R., Liu, T., Li, L., Yang, C., & Feng, Y. (2018). An Empirical Study on the Intention to Use Online Medical Service. In *2018 15Th International Conference on Service Systems and Service Management (Icsssm)*. https://doi.org/10.1109/ICSSSM.2018.8464965

Davis, F. D. (1985). *A technology acceptance model for empirically testing new end-user information systems: Theory and results*. *Doctoral dissertation*. Massachusetts Institute of Technology.

Davis, F. D. (1989). Perceived Usefulness , Perceived Ease Of Use , And User Acceptance. *MIS Quarterly*, *13*(3), 319–340. https://doi.org/10.2307/249008

Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention and behavior: An introduction to theory and research*. Reading, MA: Addison-Wesley.

Füller, J., Faullant, R., & Matzler, K. (2010). Triggers for Virtual Customer Integration in the Development of Medical Equipment - From a Manufacturer and a User's Perspective. *Industrial Marketing Management*, *39*, 1376–1383. https://doi.org/10.1016/j.indmarman.2010.04.003

Greguric, I. (2014). Ethical issues of human enhancement technologies: Cyborg technology as the extension of human biology. *Journal of Information, Communication and Ethics in Society*, *12*(2), 133–148. https://doi.org/10.1108/JICES-10-2013-0040

Heisele, B., Serre, T., Pontil, M., Vetter, T., & Poggio, T. (2002). Categorization by Learning and Combining Object Parts. *Advances in Neural Information Processing Systems*, *14*(2), 1239–1245.

Hossain, A., Quaresma, R., & Rahman, H. (2019). Investigating Factors Influencing the Physicians' Adoption of Electronic Health Record (EHR) in Healthcare System of Bangladesh: An Empirical Study. *International Journal of Information Management*, *44*, 76–87. https://doi.org/10.1016/j.ijinfomgt.2018.09.016

Kasap, Z., & Magnenat-Thalmann, N. (2007). Intelligent Virtual Humans with Autonomy and Personality: State-of-the-Art. *Intelligent Decision Technologies*, *1*(1–2), 3–15. https://doi.org/10.3233/IDT-2007-11-202

Kostrica, D. (2018). Medical Approach of Transhumanism. *HUMANUM*, *28*(1), 67–74. Retrieved from http://www.humanum.org.pl/images/2018/humanum_28_1_2018.pdf#page=67

Li, H., Wu, J., Gao, Y., & Shi, Y. (2016). Examining Individuals' Adoption of Healthcare Wearable Devices: An Empirical Study from Privacy Calculus Perspective. *International Journal of Medical Informatics*, *88*, 8–17. https://doi.org/10.1016/j.ijmedinf.2015.12.010

Mori, M. (1970). The Uncanny Valley. *Energy*, *7*(4), 33–35.

Moser, S. E., & Aiken, L. S. (2011). Cognitive and Emotional Factors Associated with Elective Breast Augmentation among Young Women. *Psychology & Health*, *26*(1), 41–60. https://doi.org/10.1080/08870440903207635

Olarte, C., Pelegrín, J., & Reinares, E. (2017). Model of acceptance of a new type of beverage: Application to natural sparkling red wine. *Spanish Journal of Agricultural Research*, *15*(1), 1–11. https://doi.org/10.5424/sjar/2017151-10064

Pelegrín-Borondo, J., Arias-Oliva, M., Murata, K., & Souto-Romero, M. (2018). Does Ethical Judgment Determine the Decision to Become a Cyborg? *Journal of Business Ethics*, 1–13. https://doi.org/10.1007/s10551-018-3970-7

Pelegrín-Borondo, J., Reinares-Lara, E., & Olarte-Pascual, C. (2017). Assessing the acceptance of technological implants (the cyborg): Evidences and challenges. *Computers in Human Behavior*, *70*, 104–112. https://doi.org/10.1016/j.chb.2016.12.063

Pelegrín-Borondo, J., Reinares-Lara, E., Olarte-Pascual, C., & Garcia-Sierra, M. (2016). Assessing the moderating effect of the end user in consumer behavior: The acceptance of technological implants to increase innate human capacities. *Frontiers in Psychology*, *7:132*, 1–13. https://doi.org/10.3389/fpsyg.2016.00132

Reinares-Lara, E., Olarte-Pascual, C., & Pelegrín-Borondo, J. (2018). Do you Want to be a Cyborg? The Moderating Effect of Ethics on Neural Implant Acceptance. *Computers in Human Behavior*, *85*, 43–53. https://doi.org/10.1016/j.chb.2018.03.032

Reinares-Lara, E., Olarte-Pascual, C., Pelegrin-borondo, J., & Pino, G. (2016). Nanoimplants that Enhance Human Capabilities: A Cognitive-Affective Approach to Assess Individuals' Acceptance of this Controversial Technology. *Psychology & Marketing*, *33*(9), 704–712. https://doi.org/10.1002/mar.20911

Schifter, D. E., & Ajzen, I. (1985). Intention, Perceived Control, and Weight Loss: An Application of the Theory of Planned Behavior. *Journal of Personality and Social Psychology*, *49*(3), 843–851. https://doi.org/10.1037/0022-3514.49.3.843

Sezgin, E., Özkan-Yildirim, S., & Yildirim, S. (2017). Investigation of Physicians' Awareness and Use of M-Health Apps: A Mixed Method Study. *Health Policy and Technology*, *6*(3), 251–267. https://doi.org/10.1016/j.hlpt.2017.07.007

Stein, J. P., & Ohler, P. (2017). Venturing into the uncanny valley of mind—The influence of mind attribution on the acceptance of human-like characters in a virtual reality setting. *Cognition*, *160*, 43–50. https://doi.org/10.1016/j.cognition.2016.12.010

Triviño, J. L. P. (2015). Equality of Access to Enhancement Technology in a Posthumanist Society. *Dilemata*, *7*(19), 53–63. Retrieved from https://www.dilemata.net/revista/index.php/dilemata/article/view/400

van den Brule, R., Dotsch, R., Bijlstra, G., Wigboldus, D. H. J., & Haselager, P. (2014). Do Robot Performance and Behavioral Style affect Human Trust?: A Multi-Method Approach. *International Journal of Social Robotics*, *6*(4), 519–531. https://doi.org/10.1007/s12369-014-0231-5

Venkatesh, V. (2000). Determinants of Perceived Ease of Use : Integrating Control, Intrinsic Motivation, and Emotion into the Technology Acceptance Model. *Information System Research*, *11*(4), 342–365. https://doi.org/10.1287/isre.11.4.342.11872

Venkatesh, V., & Bala, H. (2008). Technology Acceptance Model 3 and a Research Agenda on Interventions. *Decision Sciences*, *39*(2), 273–315. https://doi.org/10.1111/j.1540-5915.2008.00192.x

Venkatesh, V., & Davis, F. D. (2000). A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies. *Management Science*, *46*(2), 186–204. https://doi.org/10.1287/mnsc.46.2.186.11926

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User Acceptance of Information Technology: Toward a Unified View. *MIS Quarterly*, *27*(3), 425–478. https://doi.org/10.2307/30036540

Venkatesh, V., Thong, J. Y. L., & Xu, X. (2012). Consumer Acceptance and Use of Information Technology : Extending the Unified Theory od Acceptance and Use of Technology. *MIS Quarterly*, *36*(1), 157–178. https://doi.org/10.2307/41410412

Yang, H., Yu, J., Zo, H., & Choi, M. (2016). User Acceptance of Wearable Devices: An Extended Perspective of Perceived Value. *Telematics and Informatics*, *33*(2), 256–269. https://doi.org/10.1016/j.tele.2015.08.007

ETHICS AND ACCEPTANCE OF INSIDEABLES IN JAPAN: AN EXPLORATORY Q-STUDY

Stéphanie Gauttier, Jorge Pelegrín Borondo, Mario Arias-Oliva and Kiyoshi Murata

Grenoble Ecole de Management – Univ Grenoble Alpes ComUE (France), Universidad de la Rioja (Spain), Universitat Rovira I Virgili (Spain), Meiji University (Japan)

Stephanie.gauttier@grenoble-em.com; Jorge.pelegrin@unirioja.es ; Mario.aras@urv.cat ; kmurata@meiji.jp

EXTENDED ABSTRACT

Cyborg technology

Nowadays, it is possible to buy microchip implants to be put inside the body for a few hundred euros. The use of this technology, which is also called an "insideable" as it goes inside the body is not without ethical concerns.

Previous studies show that while there is low resistance toward insideables (technology that is implanted into the body) for human augmentation, participants also question the morality of such a use of enhancement technology (Murata et al. (2017). In a further study, Pelegrín-Borondo et al. (2018) show that ethical dimensions explain 48% of the intention to use cyborg technologies. Based on secondary data, Gauttier (2019) shows that there are many ethical deliberations embedded in the decision to accept or reject insideables. Understanding how ethical concerns shape the perception of cyborg technologies such as insideables is needed.

Shared points of view about cyborg technology in Japan

The points of view of Japanese students on insideable technologies were surveyed through a Q-study. Q-methodology aims at capturing the subjective points of view of individuals on a topic (Stephenson, 1935). The starting phase of a Q-study is to gather the volume of all items referring to a topic, which is then reduced to a smaller list of items to avoid redudancies. These items can be in textual, visual, or oral form. They are then proposed to participants who have to 1) read through the items; 2) sort them in three categories: agree, don't agree, neutral; 3) sort them on a forced distribution matrix according to the degree to which the items represent their point of view, 4) answer open-ended questions about the Q-sorting exercise. The forced distribution matrix follows a normal distribution, so that there are only a few statements the participants can rank as mostly representing or not representing their point of view. The filled in matrix are called Q-sorts, which are then analysed through Q-factor analysis. This analysis is a by-person analysis, and not an analysis by variable as is traditionally done with factor analysis in R. As a result, we obtain factors which are the composite Q-sorts representing the shared points of view across the sample, and can see which are the distinguishing or consensual items across the different factors.

In our case, we used the verbatim of the interview study conducted by Murata et al. (2017) and analysed it to identify statements to include in the concourse. Additional statements coming from a study on insideable by Gauttier (2019) were added. In total, 33 statements were retained, which covered topics such as the regulation needed around this technology, religious motives, business interest, and so on. A forced distribution matrix ranging from +3 to –3 was used. 20 Japanese students

proceeded to the Q-sorting exercise. A Q-factor analysis was performed combining PCA and VARIMAX. 3 factors were retained. The correlation across factors is presented in Table 1.

Table 1. Factors correlation

	1	**2**	**3**
1	1.000	0.4610	0.5838
2	0.4610	1.000	0.3161
3	0.5838	0.3161	1.000

The three factors are analysed separately, then each description is refined considering the consensus and distinguishing statements. The three points of view can be described as follows.

Factor 1 – Not interested in the use of additional technology

It is necessary to ensure that the use of insideables is not forced (+3), especially because it is not acceptable to be controlled or monitored via insideables (-3). Even though the technology can appear to be convenient (+1), I do not like the idea of having the technology inside my body (-2) and I hope to spend the rest of my life as a flesh and blood person (+2). This is not driven by religious motivations (-3) but out of personal preference. I am not interested in being amongst the first to use this technology (-2), I am actually not really interested in using technology in order to keep my good health (-1).

Factor 2 – Prudent acceptation of the use of insideables as enhancement technology

I am not opposed to the idea of enhancement beyond medical purposes (+2) and I generally think that insideables are convenient (+3). I can think of using them myself: I do not aim at remaining a flesh and blood body (-2), and I am interested in technologies to keep healthy (+1). However, the use of insideables must be allowed only in specific conditions. For instance, it must not be forced (+3), the data must be protected (+2). I would think of each case of use and then make up my mind (+1).

Factor 3- The use of insideable is potentially dangerous for society

There is great danger associated to the use of insideables, even if this technology is convenient (+2). The dangers are of political nature (+3), but also related to how people would treat each other if some would use the technology (+2), and to our position in relation to technology as we would grow dependent (+2) and might harm our ability to learn on our own (+3). A social debate is therefore needed (+2), whereby eventually an age limit could be set (+1) and the use of insideables would not be forced (+1). It is necessary to think about these things first, there is no attraction in being amongst the first using the insideables (-2), and I am not sold to the idea of enhancement: I do not like the idea of technology becoming a part of my body (-2) and I do not agree with the idea of using insideables for non-medical purposes (-2).

Differences and consensus across factors

The differences across the factors are indeed mostly centered on the statements around the fusion between technology and body; the use of insideables for non-medical purposes.

The consensus across factors is related to a need for a social debate and neutral perspective on statements describing potential use situations. There is also a shared belief that the use of enhancement technologies is not going to solve issues: all factors disagree with the idea that insideables would lead to less ignorance because they can bring additional memory.

The role of ideology and ethics in shaping the factors

This study identifies different points of view on insideables among Japanese students. The points of view are polarised according to ideological positioning regarding the role of technology in the body, the idea of enhancement, as well as dependent on participants' abilities to imagine future uses. In this sense, they reflect the current philosophical debates on enhancement which oppose visions of the human and are at an impasse when empirical cases are not available to stimulate one's imagination of potential futures.

It is noteworthy that both in Factors 1 and 2, the idea of not knowing enough to have an opinion was highly rated, while Factor 3 disagrees slightly with this statement. This explains why Factors 1 and 2 did not focus on potential ways to regulate the use of the technology, for which a solid understanding of the technology is required. The participants forming this point of view explained in the post sorting interview that they "cannot clearly realize the nature and ways of using the insideable" (participant 5), that they "can't imagine" (participant 1, participant 10). The uses that go beyond what one can imagine should be prohibited (participant 10).

Participants in Factor 1 did explain that they would not accept the technology. There is a strong-pre-existing position on the insideables, which the Q-sorting exercise barely changes. Participant 2 mentions being intuitively opposed to insideables, but that reading the statements it appears that the general public must make their opinion.

Participants in Factor 2 do on the contrary mention some benefits for a surveillance society and some belief that enhancement can bring happiness. The difference between the two points of view is therefore related to an ideological positioning on the desirable role of the technology.

The points of view of Factor 1 and 2 consider ethical questions related to privacy and data control , as well as consent. These ethical concerns are widely discussed in society nowadays and are potentially easier to grasp for students. On the contrary, statements requiring a projection regarding what the technology could do are not ranked in the extremes, as if the participants had less interest, perhaps more difficulties, in appropriating them.

KEYWORDS: Human enhancement, Insideables, Ethics, Acceptability, Technology Acceptance, Q-method.

REFERENCES

Gauttier, S. (2019). 'I've got you under my skin'—The role of ethical consideration in the (non-) acceptance of insideables in the workplace. *Technology in Society*, 56, 93– 108.

Murata, K., Adams, A. A., Fukuta, Y., Orito, Y., Arias-Oliva, M., & Pelegrin-Borondo, J. (2017). From a science fiction to reality: Cyborg ethics in Japan. ACM SIGCAS Computers and Society, 47(3), 72-85.

Pelegrín-Borondo, J., Arias-Oliva, M., Murata, K., & Souto-Romero, M. (2018). Does Ethical Judgment Determine the Decision to Become a Cyborg? *Journal of Business Ethics*, 1-13.

Stephenson, W. (1935). Technique of factor analysis. *Nature*, 136(3434), 297.

THE ETHICAL ASPECTS OF A "PSYCHOKINESIS MACHINE": AN EXPERIMENTAL SURVEY ON THE USE OF A BRAIN-MACHINE INTERFACE

Yohko Orito, Tomonori Yamamoto, Kiyoshi Murata, Yasunori Fukuta

Ehime university Japan), Ehime University (Japan), Meiji University (Japan), Meiji University (Japan)

orito.yohko.mm@ehime-u.ac.jp; yamamoto.tomonori.mh@ehime-u.ac.jp;
kmurata@meiji.ac.jp; yasufkt@meiji.ac.jp

EXTENDED ABSTRACT

The assistive technologies (AT), such as power assisting suits and smart glasses, have been developed and put to practical use in the developed societies. Among others, a brain-machine interface (BMI) or a brain-computer interface (BCI) has recently attracted attention. A BMI enables communication between a human brain and external devices through obtaining signals from a brain and sending signals from devices to a brain with the aid of dedicated hardware and software. Most of BMI or BCI systems consist of four sequential components: signal acquisition, feature extraction, feature translation, and classification output (Rupp et al., 2014, p.9). According to the survey conducted by Nijboer et al (2013), the people involved in BCIs tend to prefer to consider that BCI systems "measure signals from the central nervous system and 'translate' those signals into output signals" (p.545). As the previous study suggests, BMI systems process the signals acquired from a human brain and translate them into a meaningful output in accordance with given purposes, such as the practical operations to move some materials and sending messages. In this respect, a BMI system can be regarded as a "psychokinesis machine" or "telepathy machine".

In many cases, a BMI has been used for medical and rehabilitation purposes in the form of a non-invasive wearable. For example, even patients of amyotrophic lateral sclerosis (ALS) who have lost their ability to control body movements can remotely operate devices using a BMI system. In such cases, the patients' brain signals are collected by the hardware systems and processed by the software systems so that they can operate devices through just thinking to do so. In Japan, the demonstration experiments using BMIs have already been conducted. In the experiments, disabled people who could not move their bodies at will successfully worked for a coffee shop as wait staff through remotely operating humanoid robots, thanks to the wearable BMI devices connected with their brain (Ory Labo, https://tinyurl.com/y4acthl8). BMI systems are expected to have a wide array of uses other than medical purposes, such as for gaming (Niiholt, 2008; Niiholt, 2009), marketing (Guger et al., 2014) and so on.

The usage of a BMI for wider range of purposes may exert a substantial influence over individuals, organisations and society as a whole, same as other information technologies. However, the ethical considerations of such wider usage of a BMI have not been undertaken. There are difficulties in imagining and evaluating the social risks caused by the usage, because the actual use of AT for healthy people in daily-life settings have little been experienced until now. In addition, we cannot expect that they will use BMIs also in the future because healthy people usually do not recognise the necessity to use such technologies. Therefore, a possible way to investigate the ethical and social aspects of BMI usage in the wider context is conducting an experimental survey.

In fact, almost all of existing studies on BMIs or BCIs have focused on the medical or rehabilitation use of them. Some of them attempted to examine the social risks and ethical issues concerning a BMI with

conducting questionnaire surveys of medical or rehabilitation professionals and patients using BMI devices (e.g. Nijboer et al, 2013; Gilbert, 2019; Isobe, 2013). There are just a few interview surveys asking healthy people (not experts or professionals) about their recognition of BMI usage to analyse its social risks. However, given that a BMI will soon be used in the wider context, the ethical issues and social risks concerning BMI use should be examined in a proactive manner taking socio-cultural and economic contexts in which the technology will be used into consideration. The following questions are raised based on this idea.

- If BMI devices malfunction against their users' intentions, who is responsible for the malfunction? How can we decide the responsible people or organisations?
- Should users' brain signals collected by BMI systems be protected as private information of the users?
- Is it socially or legally acceptable that brain signals obtained from individual users during their usage of BMI systems are utilised for some specific purpose? Can BMI systems be used as a lie detector?
- What kinds of advantages and risks do exist when BMI systems are used in organisational settings?

In the future, it is expected that an implantable BMI or a brain-chip-type BMI will be available. A personal decision on whether using such a device is serious for its user, because of the human cyborgisation through implanting the chip and the expected physical burden of the surgery to implant it. Here, the following ethical questions are raised.

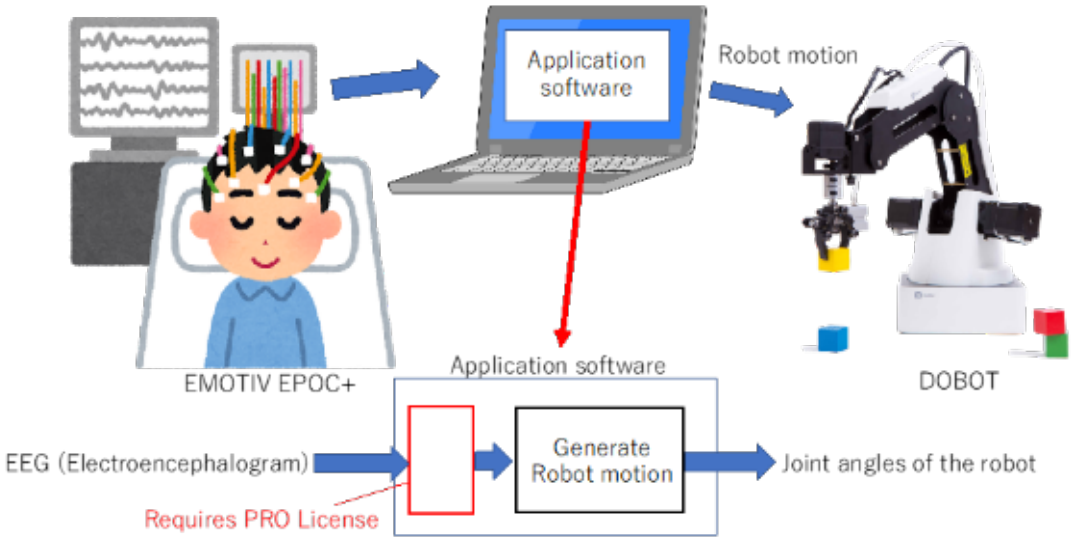
- In what condition is the implantation of BMI devices into the brain, which is a mysterious region of body and deeply related to human dignity, of a disabled or healthy person justified? Who is a person eligible to use the brain chip?
- To what extent should the autonomy of the individual's decision to become a cyborg using an implantable BMI be respected? Does this relate to the social role or professional duty of a BMI user?
- How can human somatic sensations, self-recognition and self-identities be transformed when an implantable BMI is embedded in the human brain?

To address those questions, this study conducts the experimental and interview surveys with healthy people. As shown in Figure 1, the experiment environment in this study is composed of an Electroencephalogram (EEG), a robot arm and the dedicated application software.

The subjects of the experiment and interview surveys are as follows.

- Healthy people (students, nursing care professionals, nursery teachers, teachers at elementary, junior high and high schools, athletes, people involved in sports, information technology engineers, people engaged in service industries)
- Researchers (in the fields of information ethics, philosophy, information management, sociology, engineering, computer science, production management, industrial engineering, physical information theory, medical ethics, neuroscience and so on)

Figure 1. The experiment environment



The interview sheets used in this study are designed so as to examine the subjects' attitudes to and recognition of the BMI from the following perspectives: (a) privacy and personal data protection, (b) human autonomy and dignity, (c) identity development and persona transformation, (d) the acceptance of body extension in individual and organisational context, (e) the workplace cyborgisation, (f) social responsibilities and informed consent. Through carefully examining the results of the surveys, this study reveals the social risks entailed in the development and usage of BMIs and the appropriate policies on this machine to address the risks.

KEYWORDS: brain-machine interface, privacy, autonomy, responsibility, human dignity, identity

REFERENCES

Gilbert, F., Cook, M., O'Brien, T., & Illes, J. (2019). Embodiment and estrangement: Results from a first-in-human 'Intelligent BCI' trial. Science and engineering ethics, 25(1), 83-96.

Grübler, G. and Hildt, E. eds. (2014). Brain-Computer Interfaces in their ethical, social and cultural context. Dordrecht: Springer.

Guger, C., Brendan, Z. A., & Edinger, G. (2014). Emerging BCI opportunities from a market perspectives. In Grübler, G., Hildt, E. (eds) Brain-Computer-Interfaces in their ethical, social and cultural contexts. The International Library of Ethics, Law and Technology, 12. Springer, Dordrecht, 85-98.

Fukushi, T. & Sakura, O. (2007). Ethical implementation of research and development on Brain-Machine Interface. Keisoku to Seigyo, 46(10), 772-777, Retrieved from https://doi.org/10.11499/sicejl1962.46.772 (in Japanese).

Isobe, T. (2013). The Perceptions of ELSI researchers to Brain-Machine Interface: Ethical & social issues and the relationship with society. Journal of Information Studies, (84), 47-63 (in Japanese) .

Nijboer, F., Clausen, J., Allison, B. Z., & Haselager, P. (2013). The Asilomar survey: Stakeholders' opinions on ethical issues related to Brain-computer Interfacing. Neuroethics, 6(3), 541-578.

Nijholt, A. (2008). BCI for games: A 'state of the art' survey. In International Conference on Entertainment Computing. Berlin, Heidelberg: Springer, 225-228.

Nijholt, A., Bos, D. P. O., & Reuderink, B. (2009). Turning shortcomings into challenges: Brain–computer interfaces for games. Entertainment computing, 1(2), 85-94.

Ory Labo, Retrieved from https://tinyurl.com/y4acthl8

Rupp, R., Kleih S.C., Leeb , R., del R. Millan J., Kübler A. & Müller-Putz G.R. (2014). Brain–Computer Interfaces and assistive technology. In Grübler, G., Hildt, E. (eds.), Brain-Computer-Interfaces in their ethical, social and cultural contexts, The International Library of Ethics, Law and Technology, 12. Dordrecht: Springer, 7-38.

Tamburrini, G. (2014). Philosophical reflections on Brain-Computer Interface. In: Grübler, G., Hildt, E. (eds.), Brain-Computer-Interfaces in their ethical, social and cultural contexts, The International Library of Ethics, Law and Technology, 12. Dordrecht: Springer, 147-162.

TRANSHUMANISM: A KEY TO ACESS BEYOND THE HUMANISM

Ferran Sánchez Margalef

University of Barcelona (Spain)

ferran.sanchez@ub.edu

EXTENDED ABSTRACT

We are at the beginning of a technological tsunami that will transform many of the spheres of human reality. Transhumanism, the pathway that will take humanity to its own transcendence, has already begun. Taking into account that some corporations or institutions of great social and economic importance (i.e. NASA, Google or the University of the Singularity) are already starting to invest in projects that facilitate the arrival of a post-human world, it is essential for humanity to consider the challenges that this movement implies. Hence, the first step to acquire the required consciousness to reach the understandings is to maintain a constructive, argued and peaceful debate. However, in a scenario of a non-agreement on the required consensus about the limits of Transhumanism, there is a certain possibility that humanity will stop using technology as a means to put itself at the service of its logic.

1 INTRODUCTION

The repercussions that Transhumanism (H+) forecasts affects every human since it directly involving the condition of *Homo sapiens.* Imagination has always been part of the human being, as well as the determination of realizing these recreations into reality. Nevertheless, until the present day, every dream has been limited by the biological condition.

To focus this brief communication about Transhumanism, the phenomenon that aims to substitute the biological limits by the technological, we will first establish and contextualize this movement, giving and commenting some of the motivations from several lectures. Then, we will provide some counterarguments from an axiological point of view, and leaning on the opinion of several authors that discuss the transhumanist proposals and refute their optimistic positions. Finally, the last section gives a conclusion as a consequence of a hermeneutic interpretation of the obtained results.

2 AN APPROACH TO TRANSHUMANISM

Transhumanism, a movement that is still emerging today, is one of the most decisive phenomenon in this century. The reason behind is because H+ is based on postulations that outline the possibility of overcoming the human biological condition and abandon our specie. This advance, which is only possible by the interconnection of the propitiated digitalization and technological implementation in different areas of knowledge (i.e. biotechnology, nanorobotics, AI, cognitive sciences and media technology), goes to the future while the next scientific advances make possible the arrival of Posthumanity.

Hence, one of its most characteristic features is that, while other worldviews, sensibilities, phenomena, movements or parties have been inspired in the past to build their speech, Transhumanism denies the past to venerate the future. According to the predictions, an advanced and superior specie is going to replace humans as known in the present-day. This fact, which do not have

any precedent, is going to be possible with the advance and technological deployment that will arrive to every single place of a Posthuman world, either organic or inorganic.

Despite not having historical references of H+, it is possible to equate the Copernican turn with the one that took place during the Renaissance and culminated with the French Revolution. We are referring to the humanism that moved God from the center of the Cosmos to place the human in it (*anthropocentrism)*. If Humanism put the man in the epicenter of the Universe, Transhumanism will displace him to give place to a new being that will appear from the human digitalization.

3 CULTURAL IMPACT OF TRANSHUMINSM

We would like to point out that any change at a social level requires, previously, a change of awareness from the society. In this sense, the battle over the technological hegemony has already begun and both the scientific discourse and human reality are gradually impregnated with the growing technological assimilation. It is not a minor matter that many of the spheres of human activity are already filled with applied sciences nor that our society venerate innovation, dynamism or consumerism, since these are the same values that will facilitate the arrival of Posthumanism.

According to the linguistic turn (Wittgenstein, 2013) we must consider the words we use because, beyond describing, they are configuring human reality. So, we can better understand why we use words such as AI that are presupposing an intelligence to machines, for being capable to process huge amounts of information and do complex calculations (while human intelligence is composed of a large number of factors such as sensitivity, the critical capacity, emotional intelligence, improvisation ability, etc., that machines do not have).

4 MORAL CRITISISM OF THE TRANSHUMANIST MOVEMENT

Two opposite sides have arose as a consequence of this new paradigm: Those who defend the postulates of the Transhumanism and those who refute them. That is, the repercussion of the movement is still unknown and, therefore, rejectable for several people. In this short text, some of the challenges proposed by the H+ in the axiological level are going to be discussed, taking into account the thesis of authors that have already contributed in this debate.

Fukuyama described Transhumanism as the most alarming idea ever expected (Fukuyama, 2002), when considering it as a frontal attack to humanity (Fukuyama et al., 2002). The occidental society has a consensus about the Human Rights since there is an international acceptance of the fundamental premises such as the life dignity or the equality of lives.

Although the different legislations, which subject humanity to the rights, defend (to a greater or lesser extent) to abide these universal maxims, it must be borne in mind that the first fundamental right, indispensable to be able to exercise any other, is none other than the natural right to life. Taking into account that H+ directly modifies the human condition and life as understood today, it is also clear that it attacks the very dignity of the species.

In addition, the breach of the right to a biological life is also the breakdown of the right to a proper and spontaneous identity resulting from a biological chance and from an extremely complex set of variables and conditioning factors. Trying to control and influence these variables implies directing a life, ergo violating its most intimate dignity: the freedom for each one to be what he or she must be. Beyond that, Sandel also reflects on the pressure that could be placed on the future improved subjects,

considering the expectations placed on them and the possible serious disappointments or even depressions if they are not up to the task.

The second argument that we want to comment is about equality. Even though the differences between humans are large and varied, they all share the same genetic condition. However, if H+ fractures with this equality, the contrast between the intelligent species inhabiting the planet will be emphasized. In other words, part of a privileged population will be able to access to the biotechnological benefits, while others will not. Hence, one may be concerning about the relationship between both communities.

Will those who have broken the biological condition and those who remain tied to their mortality compete for the same oppositions and in the same competitions? Transhumanist development inherently implies an imbalance at the social level produced by those who begin to access this type of technology. Harari uses the term *Homo Deus* to refer to this new species, to all those beings that begin to be endowed with divine properties (Harari, 2016). *Creation* is no longer an exclusive property of the gods: the future Being will also be able to create better beings and even to recreate itself.

Finally, freedom is the last defended thesis. As R. Panikkar states, technology is the knowledge of control (Panikkar, 2009). Assuming that Transhumanism can only take place with interrelation of the biology with technology, the possibility of a restricted freedom needs to be treated. Considering the reflections of E. Kant about the Illustration, freedom is defined as the overcoming of the man to the under-age (Kant, 2009). Therefore, Transhumanism envelops into a paradox since it may take away the resources that have allowed the man to think by himself.

5 CONCLUSION

First, it can be assumed that our behavior, our language or our reasoning are already being highly influenced by transhumanist postulates. A clear example can be seen in the concept itself and in its own antonym. The word *transhumanist*, chosen by the followers of this current, implies a series of positive connotations.

On the other hand, if we refer to the antagonist word, *bioconservative*, the name evokes a certain perception of antiquity or something retrograde. It is necessary to point out, in order to get an idea as faithful as possible, that while transhumanists chose the name used by Huxley and recovered by Esfanidary, *bioconservative* is the alias, clearly derogatory, that transhumanists have imposed.

Seeking new concepts on which to build discourses is therefore a prerequisite before starting the debate on ethical conditions; otherwise, battles will have been lost even before the start of the war. To conclude, then, we suggest another concept, *biovitalism*, on which to build a discourse and an alternative narrative to H+.

Second, refering to ethics we would like to point out that the question of accepting or not transhumanist postulates cannot be based on individual freedom. Transhumanism is not a movement that affects only people who choose to transcend humanity, it affects humanity as a whole. In other words, the repercussion of crossing the biological line implies, even if it is just a person who is doing it, the life of two rational species in the Earth.

It is necessary, then, to put in question the coexistence and the good welling of the relationship between the two rational species that will live in the future because it seems that the human that have overpass the biological condition will be capable to adapt much better to the Posthuman society.

KEYWORDS: Biovitalism, Ethic, Freedom, Technology, Transhumanism.

REFERENCES

Kant, E. (2009). ¿Qué es la Ilustración? *Foro de Educación*, *7*(11), 249-254.

Panikkar, R. (1991) *El "tecnocentrisme". Algunes tesis sobre tecnologia*. Barcelona: La Llar del Llibre.

Harari, Y. N. (2016). *Homo Deus: breve historia del mañana*. Madrid: Debate.

Fukuyama, F., & Reina, P. (2002). *El fin del hombre: consecuencias de la revolución biotecnológica.* Barcelona: Ediciones B.

Sandel, M. J. (2007). *Contra la perfección. La ética en la era de la ingeniería genética*. Barcelona: Marbot ediciones.

Wittgenstein, L. (2013). *Tractatus logico-philosophicus*. Routledge. Retrived from https://content.taylorfrancis.com/books/download?dac=C2013-0-16524-2&isbn=9781134644582&format=googlePreviewPdf

3. Diversity and Inclusion in Smart Societies: Not Just a Number Problem

Track chairs: Efpraxia Zamani, University of Sheffield, United Kingdom – Shalini Kesar, Southern Utah University, USA – Kutoma Wakunuma, De Montfort University, UK

BRIDGING THE GENDER GAP IN STEM DISCIPLINES: AN RRI PERSPECTIVE

Maria Michali, George Eleftherakis

South-East European Research Centre (Greece), CITY College,
International Faculty of the University of Sheffield (Greece)

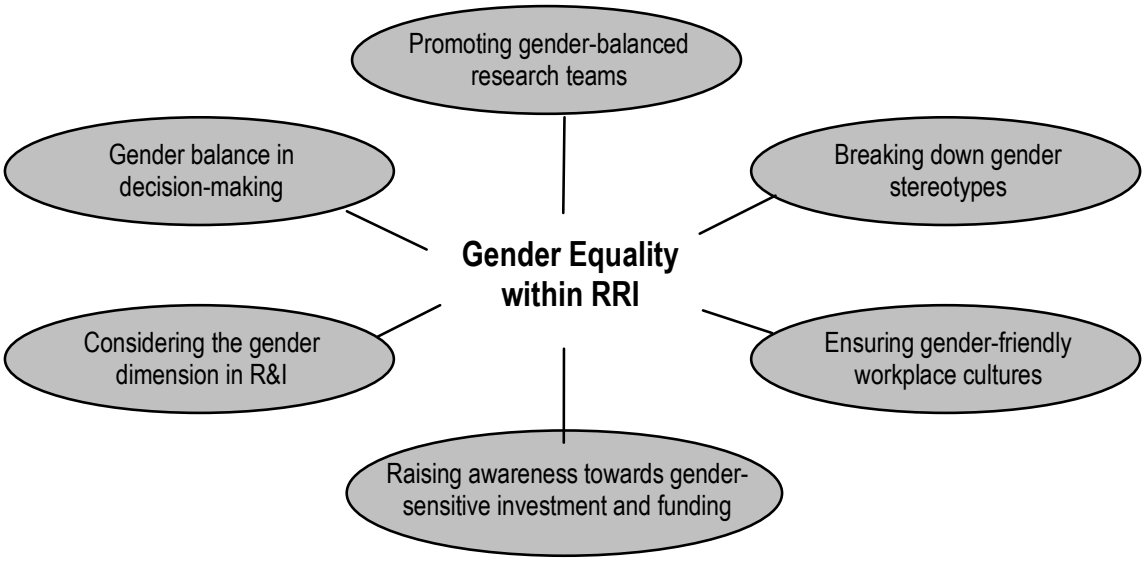
mmichali@seerc.org, g.eleftherakis@sheffield.ac.uk

EXTENDED ABSTRACT

Over the recent years gender equality has been receiving increasing attention, both in accordance to the humanitarian perspective referring to individuals' welfare and to the Humanitarian-Development Divide (United Nations Sustainable Development Goals – SDGs and UNESCO Priority of Gender Equality), as well as in accordance to enhancing female representation in fields 'traditionally' encountered as male-dominated, namely the STEM-related disciplines. In relation to STEM, the gender gap is prominent in various fields −for instance in research and in academia–, and in Science and Technology occupations. According to She Figures 2018, Europe may be close to bridging the gender gap in the doctoral field (47,9% female doctoral students in 2016), but there are considerable differences in gender representation per field of expertise; females constitute less than 1/3 in STEM fields such as Information and Communication Technologies (ICT) and Engineering-Manufacturing-Construction (21% and 29% in 2016 accordingly). A lack of diversity exists in the labour market as well, with only 30% of women with STEM qualifications in Europe having a relevant occupation; in other words, "a significant number of them take jobs in-non related roles, representing a loss of talent and potential and economic gains" (Salinas and Bagni, 2017, p. 721). Nevertheless, female inclusion in STEM does not rely only in augmenting female representation in terms of statistical percentages. Women face other prominent problems, like formal and informal recruitment-selection procedures hindering their advancement in science and especially in polytechnic careers (Carvalho and Santiago, 2010), while gender representations, 'extra-organisational' gender roles (Mills, 1988), or role models that encourage or discourage females from engaging in STEM are issues beginning to be addressed.

Within the context of our study, gender (in)equality and related multi-layered interventions (i.e. not only referring to a higher numerical representation of women) are addressed in relation to STEM disciplines. These interventions are interrelated to some SDGs sub-objectives, to the European Commission's 'commands' for gender equality in the European Research Area (ERA), as well as to the concept of Responsible Research and Innovation (RRI). Firstly, Goal 5.B of the SDGs suggests to enhance the use of enabling technology and ICT to promote female empowerment, while the European Commission (EC) similarly introduces gender equality policy interventions in scientific fields and calls for action towards a proper integration of gender issues through specific proposals in EU Research&Innovation Programmes, namely Horizon Europe and Horizon 2020 (European Commission, 2014). This is the focal point where these European initiatives are complemented by RRI. RRI refers to tackling contemporary societal challenges by aligning the values, needs and expectations of all actors involved in R&I systems. In the view of Von Schomberg (as cited in Owen et al, 2012) and his definition of RRI, "science and innovation are envisaged as being directed at, and undertaken towards, socially desirable and socially acceptable ends, through an inclusive and deliberative process" (p.753). These socially desirable ends actually seem to have been transformed to the six policy agendas that RRI addresses; the six RRI keys. Gender equality also belongs to these keys, and acquires multiple layers within RRI (depicted in Figure 1, which was designed according to the input from RRI tools website).

Figure 1.RRI and Gender Equality



Source: RRI tools website

Currently, several RRI initiatives foster female inclusion in STEM disciplines, and are related to projects implemented mainly during Framework Programme 7 (FP7) and Horizon 2020 (H2020). These projects set gender equality as a priority in various Research Performing Organisations (RPOs) with a STEM expertise, and proceed to the development of self-tailored Gender Equality Plans (GEPs). The GEPs aim to institutionalise gender equality, trigger structural transformations in the RPOs and reach a broad knowledge transfer which contributes to meeting various ERA objectives (e.g. priority 4).

In a similar line of argument, the present study delves deeper into RRI initiatives towards gender equality, and examines FP7 and H2020 EU funded projects that foster gender diversity and female inclusion in STEM-related RPOs. Emphasis is on the RRI key of gender equality as opposed to the other keys, since it constitutes an emerging issue reflecting contemporary concerns. It is actually a multifaceted issue, as gender equality is not just a number problem and complementary activities should be implemented for 'changing' the scientific status quo. The aim of this study, therefore, is to critically analyse the innovation practices implemented within EU Gender Equality projects. While examining various RRI projects (approximately 80) included in the two major calls of FP7 and H2020, five Gender Equality projects have been selected through a two-stage selection procedure including criteria like *innovativeness*, *stake*, *transparency* and *impact*, and these projects have been further and more critically analysed: *EQUAL-IST*, *STAGES*, *GENERA*, *GEECCO* and *PLOTINA*. The subsequent aim is to identify tendencies ('mega trends') in the actions of European RPOs, when ameliorating their intimate mechanisms by developing new structures ensuring gender equality. It is worth highlighting that these projects have been considered as a source (a 'container' of practices) and the practices as the units of a qualitative analysis. Following the arguments of Braun et al and their six-step framework for conducting a qualitative analysis (2019), we refer to a reflexive thematic analysis of the data collected with an inductive orientation; the processes of coding and theme development have taken place by employing the NVivo software (Version 12; QSR International Pty Ltd, 2018) and the codes/themes developed have been directed by the content of the data. Patterns and regularities were afterwards identified for reaching certain conclusions. Finally, the thematic analysis has been clustered with an essentialist framework (Braun et al, 2019), where one can report an assumed reality evident

in the data; the trends/tendencies detected are an assumed reality evident within the practices promoting gender equality in the scientific field.

Ultimately, this study investigates promising interventions towards gender imbalance in STEM fields – as it has also been suggested by Gorvacheva et al (2019) in terms of future research in the corresponding topic− and thus functions as a 'mapping' tool depicting the European conditions and endorsing the successful RRI practices for 'gendering'[2] the STEM disciplines. However, it has a twofold contribution; it additionally draws valuable conclusions that resemble a set of suggestions and can be employed as such, for aiding STEM-related European actors in genuinely establishing gender equality in R&I processes. In a few words, these conclusions/suggestions, being based on the patterns detected, refer to the contextualization of RRI and the need to develop self-tailored GEPs, to the most common lines of intervention of the GEPs –namely encouraging female leadership in science, measures against horizontal segregation, (early) career development, work-life balance, training towards gender issues and gender-neutral communication, gendering scientific contents and methods etc.− as well as to co-creation processes that accompany the GEPs (e.g. collaborative platforms). Reference is also made to the impact (both internal and external) of the GEPs, and to whether it can contribute to restoring the principles of universalism and meritocracy in scientific ethos.

Therefore, the above process −if encountered holistically− can lead to organisations and STEM-related disciplines that are aligned to contemporary societal concerns, and are truly response-*able* (i.e. able to provide responses to emerging situations and challenges) through innovation and *re*-search (i.e. continuous search). Of equal importance is that the new emerging definitions of smart societies can be enhanced; smart societies should not just embrace technology but also tackle societal ills (Haupt, 2017) in the way these are represented within the SDGs −and as mentioned gender equality is included in these goals. Finally, smart societies need a paradigmatic shift, while the term 'paradigm' refers to the common techniques and values that members of a scientific community share (Kuhn, 1962; Agamben, 2009), and the definition of the paradigmatic shift connotes a fundamental change in the basic concepts and experimental practices of a scientific discipline (Kuhn, 1962). Thus, gender-related practices within STEM disciplines shall replace the basic concept of males dominating this field and genuinely bring this shift, which actually refers to socially desirable ends and behaviors.

KEYWORDS: STEM, RRI, Gender Equality, Gender Equality Plans, EU aspirations, Smart Societies.

ACKNOWLEDGEMENTS: This work has been conducted within the framework of TeRRItoria (Territorial Responsible Research and Innovation Through the involvement of local R&I Actors) EU project, which has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 824565.

REFERENCES

Agamben, G. (2009). What is a paradigm? *Filozofski Vestnik, 30 (1), 107-125.*

Braun, V., Clarke, V., Hayfield, N., & Terry G. (2019). In: Liamputtong P. (Eds), *Handbook of Research Methods in Health Social Sciences*. (pp. 57-71). Singapore: Springer.

[2] According to the European Institute for Gender Equality (EIGE) when drawing on Šribar et al (2015), *gendering* is defined as the process of integrating the gender perspective into the understanding and construction of persons, phenomena, reflections, relationships, sectors of action, societal subsystems and institutions.

Carvalho, T. & Santiago, R. (2010). New challenges for women seeking an academic career: the hiring process in Portuguese higher education institutions. *Journal of Higher Education Policy and Management*, 32 (3), 239-249.

European Commission (2014). *Vademecum on Gender Equality in Horizon 2020*. Retrieved from: https://ec.europa.eu/research/swafs/pdf/pub_gender_equality/2016-03-21-Vademecum_Gender%20in%20H2020-clean-rev.pdf

European Commission (2019). *She Figures 2018*. Luxembourg: Publications Office of the European Union.

Goracheva, E., Beekhuyzen J., Brocke, J., & Becker, J. (2019). Directions for research on gender imbalance in the IT profession. *European Journal of Information Systems*, 28 (1), 43-67.

Haupt, M. (2017, December 19) What is a Smart Society? Toward the transcendent model society of 2030. *Medium*. Retrieved from: https://medium.com/project-2030/what-is-a-smart-society-92e4a256e852

Kuhn, T.S. (1962). *The structure of scientific publications*. Chicago: University of Chicago Press.

Mills, A.J. (1988). Organization, gender and culture. *Organization Studies*, 9 (3), 351-369.

NVivo qualitative data analysis software (2018). Melbourne: QSR International Pty Ltd. Version 12.

Owen, R., Macnaghten, P., & Stilgoe, J. (2012). Responsible Research and Innovation: From science in society to science for society, with society. *Science and public policy*, 39 (6), 751-760.

Salinas, P. C., & Bagni, C. (2017). Gender Equality from a European Perspective: Myth and Reality. *Neuron*, 96 (4), 721-729.

Ule, M., Šribar, R. & Venturini, A. U. (Eds.). (2015). *Gendering Science: Slovenian Surveys and Studies in the EU Paradigms*. Vienna: Echoraum.

NO INDUSTRY ENTRY FOR GIRLS – IS COMPUTER SCIENCE A BOY'S CLUB?

Gosia Plotka, Bartosz Marcinkowski

De Montfort University (United Kingdom), Polish-Japanese Academy of Information Technology

malgorzata.plotka@dmu.ac.uk; bmarcinkowski@pwjstk.edu.pl

EXTENDED ABSTRACT

Even though women actively took part in information technology evolution, still relatively few of them are pursuing their professional careers in Engineering industry. The lack of consistence in terminology used by research teams investigating the phenomena (IT-related contributions often address similar research settings as Computer Science, Computing or Systems) results in a number of studies with data that slightly vary. Similar variances have been observed among developed economies – data coming from Northern America sources are to some extent different than those coming from the European ones. For instance, according to Graf, Fry & Frunk (2018), only 14% of Engineering workers are women, and computer science industry is underrepresented (25%) as well; at the same time, there was only 2% increase in Engineering jobs within the 27 years timespan (between 1990 and 2017) – whereas 7% drop in numbers in Computing. To provide a basis for comparison, throughout the same period the share of women in other fields (such as health-related, life sciences and even the other STEM areas – such as physics and maths) has increased. Homogenous data, also coming from the Northern America market, is provided by Ehrlinger et al. (2018). On top of that, Gorbacheva et al. (2019) makes an observation that women constitute only 16.7% of employed IT specialists, even though overall 47% of them are active on the job market. Those claims are based, among others, on data provided by the Eurostat.

Moreover, diverse research also reveals that there is (1) a difference in the retention of women and men in the field of their study after completing their major in Computing/Engineering fields; (2) gender salary gap remains in place – just to mention research by Craigie & Dasgupta (2017) as well as Stephan & Levin (2005). The mechanics behind women effectively disappearing from some fields that can be considered "geekier" is still an open issue, and one of a vital importance for business organizations and faculty. Therefore, the authors followed the research goal of identifying the hindrances leading to women underrepresentation within Computer Science industry and elaborating a set of best practices how to overcome some of the common problems and work together on evening up the numbers in computing to make it more diverse and inclusive.

In the pilot research conducted as a focus group, some results from the literature have been confirmed (Kindsiko & Türk, 2017). Near 20 women from around the world were asked about what they consider the main reasons for them to struggle deciding to start their professional careers and stay in IT. Their feedback helped to identify four areas of problems:

- it is a boy's club indeed – if you do not look or act like one, you cannot belong to it;
- there is a lack of support structure or role models;
- different standards for different genders and women hindering each other;
- lack of awareness or exposure.

The paper discusses a multi-country empirical study among scholars and IT professionals. The study verifies how common and impactful are the issues revealed by the pilot study as well as highlights what has been done so far to encourage women and girls to join and/or stay in STEM. Based on that, change agents and organizational best practices are proposed.

KEYWORDS: Computer Science, Gender Inequality, Job Retention, Organizational Practices.

REFERENCES

Craigie, T-A., & Dasgupta, S. (2017). The Gender Pay Gap and Son Preference: Evidence from India. Oxford Development Studies, 45(4), 479-498

Ehrlinger, J., Plant, E.A., Hartwig, M.K., Vossen, J.J., Columb, C.J., & Brewer, L.E. (2018). Do Gender Differences in Perceived Prototypical Computer Scientists and Engineers Contribute to Gender Gaps in Computer Science and Engineering? Sex Roles, 78(1-2), 40-51

Gorbacheva, E., Beekhuyzen, J., vom Brocke, J., & Becker, J. (2019). Directions for Research on Gender Imbalance in the IT Profession. European Journal of Information Systems, 28(1), 43-67

Graf, N., Fry, R., & Funk, C. (2018). 7 Facts about the STEM Workforce. Retrieved from https://www.pewresearch.org/fact-tank/2018/01/09/7-facts-about-the-stem-workforce/

Kindsiko, E., & Türk, K. (2017). Detecting Major Misconceptions about Employment in ICT: A Study of the Myths about ICT Work among Females. World Academy of Science, Engineering and Technology International Journal of Social, Behavioral, Educational, Economic, Business and Industrial Engineering, 11(1), 107-114

Stephan, P.E, & Levin, S.G. (2005). Leaving Careers in IT: Gender Differences in Retention. The Journal of Technology Transfer, 30, 383-396

SMART CITIES; OR HOW TO CONSTRUCT A CITY ON OUR GLOBAL REALITY

Andres Enrique Uribe Garriga

Interamericana Law School (Puerto Rico)

auribegarriga@gmail.com

EXTENDED ABSTRACT

The conception and use of technology are part of the reasons that make us, unlike other animals. The capability of using tools and the possibility of transforming objects in things that will help our objective is part of our genetic as a Homo Sapiens. That is why, as humans, we have been able to create, design, and live on our technological innovations. The evolution of human society and sociability is linked directly to the use of technology and the interaction of it with the humans living in the same space.

The city is the most significant technological advantage that came with the socialization and is the place were technology and people join. The city, as Jane Jacobs (1962), establish on Death and Life of the Great American City, "is by definition full of strangers…" and where these strangers met, communicate and exchange ideas, ideals, and inventions. Consequently, it has become a place where different points of view merge, and more information is available and produce.

In the last decade, the term "Smart City" has become part of the main discussion between urbanists, sociologists, programmers, scientists, and policymakers. It is a term that has multiple definitions, but a straightforward explanation of it, is as Nam and Pardo (2011) define, "a city is smart when investments in human/social capital and IT infrastructure fuel sustainable growth and enhance the quality of life, through participatory governance.". In other words, a city that has a different technological initiative that helps the members of the society. To achieve this, the financial, government and security agencies must be aware of interactions between the individual and the technology to create a better experience in the urban area were such a "smart" initiative had been implemented. Therefore, this technological advantage is intimately related to the design, livability, and socialization in the city. Since technology has become the focus on acquiring and transmitting data of the individuals, the use of it has become a great debate in the security, city planning, and development areas and how it will affect the individual rights of the citizens of the city.

The primary goal of this essay is to understand diverse points of view and policies that exist in smart cities and how the sociability of its members is to transform by implementing different technology. By comparing different concepts of smart cities, for example, the models presented by Nam and Pardo (2011), Leydesdorff and Deakin (2011), and Lombardi (2011), we will be able to understand what a smart city is, how it works, and how smart city technology is used. Also, analyzing the technology that smart cities bring to the daily life of the urban habitats and understanding how these new devices can be a juridical entity. We must put all these aspects together and analyze how they could affect the policymaking, and subsequently, the everyday life of the citizens.

Different cities in the world such as Barcelona, London, Dubai, New York City, Paris, Tokyo, Singapore, among others, have embraced the initiative to turn their areas on a smarter space, that could identify the necessities of the different individuals and brings some solutions to the daily hassle of city life. This initiative has been used in aspects of the city design and functionality, to coordinate the transit, to acknowledge the best spot to build a new community, and the necessity of the essential utilities just by knowing the direction of the wind currents in the city. Other cities use smart city architecture to

help with lowering the carbon emissions, helping the waste management of the city, or even creating oxygen, not only creating a smarter city but also a cleaner one.

Smart cities bring legal, ethical, and social challenges that must be addressed assertively; if not, it can turn into a severe problem. The privacy and security issues should be the primary concerns when dealing with the implementation of smart policies for the city. The accumulation and exchange of data is the primary way of how smart environments work, and it is why the application of policies and laws are necessary to control this flow of material and to channeled accordingly to the necessities of the citizens. Knowing the necessities and aspirations of the community is essential to address policies regarding the implementation of any project toward a smart city initiative.

To understand how to regulate and make policies to control such systems, we must understand how the smart city works and the components that make a city smart. As Gasco and Gil (2017) recognized, they are three principal elements that define the smart cities and that need to be considered to make cities even smarter, these are:

> *"Adopt a global/integral view of the city, which materializes in different types of initiatives, from waste management to traffic control to water management.*
>
> *Integrate a renewed perspective, technological, and human. Technology is critical in the development of smart cities (and, therefore, it is the tool par excellence); however, smart cities must be developed for, by, and with citizens. As a result, urban governance and participation processes, as well as investments in human and social capital, are inherent attributes of a smart city.*
>
> *Pursue a triple goal: 1) to improve the efficiency of urban operations, 2) to improve citizens' quality of life, and 3) to promote the local economy while maintaining the environmental sustainability."*

To make cities smarter, these must have a robust energy and internet infrastructure that create a viable footprint to start any smart initiative. The strategy must follow a framework involving the planning of the city, and the establishment of a partnership between the private and public sectors. With the vision of integrating the technological devices on the development of the cities. Gasco and Gil (2017), understand that to reach this target, and these three stages must be acquired "A. Alignment with the city strategy, B. Promotion of public-private partnerships and C. Co-creation and co-production of smart services."

This work will discuss the framework for a smart city using Gasco and Gil (2017) as a model. In the first stage, we will discuss the city and the urban planning of it and how this action molds the reality of the residents in the area that will affect. How this way of city planning has changed during the history and how technology changes the way the city is view. Also, we will contemplate different models, process, and actions that cities have taken to become smarter and how integrates these new technologies into the preconception of each city.

The second part of this work will look at the different joint projects that have been developed to be part of the smart city. Also, it will explain the different technologies that made possible a smart development and how this apparatus is all joined via the Internet of Things (IoT), what is precisely the IoT and how it differentiates from the traditional form of internet. To understand this part, we will also have to explore the differences between the private and the public and how smart technology create a fragile and disruptive division between them.

The third part will discuss the smart services that can be provided to the citizens. How will it help improve life in urban areas in which the smart projects have been imposed and how it has been affected by it. Analyzing this will help us notice the different social, ethical, and political issues that arise from the implementation of each smart strategy. Finally, we will make some assumptions and recommendations on how the smart cities must be studied and how they will continue evolving depending on the development of the cities policies and the needs of the people who live there.

REFERENCES

Jane Jacobs, The Death and Life of Great American Cities, Vintage Books ed., 1992, originally published in Random House ed., 1961

Gasco-Hernandez Mila & Dr. J. Ramon Gil-Garci IS IT MORE THAN USING DATA AND TECHNOLOGY IN LOCAL GOVERNMENTS? IDENTIFYING OPPORTUNITIES AND CHALLENGES FOR CITIES TO BECOME SMARTER, 2017, 85 UMKC L. Rev. 915

Lombardi, P., et al., Head of Dep't, Paper Presentation at 11th International Symposium on the Analytic Hierarchy Process: An Analytic Network Model for Smart Cities (June 15-18 2011).

Nam, T. & Pardo, T., Research Fellow, and Ctr. Dir., Paper Presentation at 12th Annual International Conference on Digital Government Research: Smart Cities as Urban Innovation: Focusing on Management, Policy, and Context (June 12-15 2011).

Leydesdorff, L. & Deakin, M., *The Triple-Helix Model of Smart Cities: A New-Evolutionary Perspective*, 18(2) J. URB. TECH 53-63 (2011)

Kelsey Finch and Omer Tene, WELCOME TO THE METROPTICON: PROTECTING PRIVACY IN A HYPERCONNECTED TOWN, 2014, Fordham Urban Law Journal, 41 Fordham Urb. L.J. 1581, 41 Fordham Urb. L.J. 1581

Balkin, Jack M., "The Three Laws of Robotics in the Age of Big Data" (2017). Faculty Scholarship Series. 5159. https://digitalcommons.law.yale.edu/fss_papers

4. Educate for a Positive ICT Future

Track chairs: Gosia Plotka, Polish-Japanese Academy of Information Technology & De Montfort University, UK – Marta Czerwonka, Polish-Japanese Academy of Information Technology, Poland – Alireza Amrollahi, Australian Catholic University, Australia

ASSESSING THE EXPERIENCE AND SATISFACTION OF UNIVERSITY STUDENTS: RESULTS OBTAINED ACCROSS DIFFERENT SEGMENTS

Ana Isabel Jiménez-Zarco, Alicia Izquierdo-Yusta, María Pilar Martínez-Ruiz

Open University of Catalonia (Spain), University of Burgos (Spain),
University of Castilla-La Mancha (Spain)

ajimenezz@uoc.edu; aliciaiz@ubu.es; MariaPilar.Martinez@uclm.es

EXTENDED ABSTRACT

Student satisfaction is today one of the main objectives to be achieved by educational institutions. In fact, in the university environment, the student, as a client with whom the university relates, has become the main reason for the existence of the university (Gento-Palacios et al., 2012). The implementation of the European Higher Education Area has contributed to this fact by allowing the student to take an active role and contribute to co-creation in the teaching-learning process (Sarmiento-Borjórquez et al 2017). In this educational context, students begin to self-manage the acquisition of knowledge, making decisions about their education, managing their own resources, taking responsibility for what they are going to learn; while sustaining with ethical principles the value of what they expect from their educational formation (Díaz-Álvarez and Cortes-Pedraza, 2012).

Due to the importance that the student has acquired, universities have implemented management models that, centred on the student, try to improve financial results and increase profitability; but also, non-financial results, so they try to strengthen the corporate image of the institution, as well as increase student satisfaction. This fact makes it convenient for academic institutions to establish indicators that allow them to measure and evaluate the results obtained in the different proposed objectives. Indeed, among these non-financial results, one of the main objectives pursued is student satisfaction. Satisfaction is a concept traditionally considered key to the success of organizations, given its impact on the behavior of individuals, in terms of trust, loyalty and recommendation (Moliner-Velázquez et al., 2015)

However, satisfaction is a complex concept to define, but above all difficult to measure. In relation to the definition of the concept, there are multiple proposals made, although in general terms satisfaction is described as a feeling that is produced in the individual and that results from comparing expectations with the result obtained (San Martín et al., 2008). On the other hand, it should be noted that the complexity when measuring the concept is derived from different aspects, among which it is possible to mention the difficulty of understanding the way in which the individual develops an internal psychological process of evaluation (Oliver, 1980; Oliver, 1981; 1997); (2) the need to consider that the object of the evaluation is not only the service provided, but also the way in which it is provided (Hackman, 2006; Hoekstra et al, 2015); to understand that through the act of use and consumption, the individual seeks to satisfy different types of needs of a cognitive, affective and behavioral nature (Oliver, 1993; Retolaza and Grandes 2003). Despite the difficulties noted, universities are generally detecting the need to establish indicators related to student satisfaction, and therefore, tools to obtain information and measure it.

The aim of this research is to measure the experience and satisfaction of university students in a Spanish Online University. And this, with the aim of designing and developing strategies to achieve differentiation from other competitors. In order to achieve this objective, a questionnaire was

designed and implemented to measure different aspects related to the experience and satisfaction of university students (including aspects such as the way in which the subject is designed, the learning resources or the evaluation system). It was also intended to evaluate this research objective by distinguishing between the different student segments.

To carry out the empirical analysis, a sample was collected through an online survey sent to all virtual university students enrolled in undergraduate and master's courses during the second semester of the academic years 2017-2018 and 2018-2019. The final sample amounts to 20,771 questionnaires in (2017) and 20,250 questionnaires in (2018). The treatment and statistical analysis of the data showed several interesting findings.

With regard to the main conclusions obtained, it is worth highlighting, among other aspects, how students perceived the final results did not meet the expectations that had been previously generated. Additionally, variables as important as the design of the objectives and contents of the different subjects, both in the degrees and in the masters, were not evaluated very positively. From a managerial point of view, these two aspects are considered vital since they constitute the central axis of the student's training (it could be even suggested that they are the ones that generate more expectations in the student). For this reason, teachers who work in subjects that receive low ratings must carefully analyse these aspects. On the other hand, the content of the subject does not exceed these expectations, creating a certain dissonance and dissatisfaction in the students.

Another strategic aspect is the low score that university students generally assigned to the teaching staff. Variables such as mastery of the subject, planning, as well as the resolution of doubts may have influenced this. From the point of view of management, one aspect that influences this dimension is the experience of teachers in the subject, as well as the availability of time for their preparation. The crisis has favoured the "in extremis" hiring of teachers to teach the day after hiring, or the excessive teaching load in other cases. In addition, the availability of resources to carry out learning in general presents a positive and significant contribution to satisfaction.

Some ethical issues arise from this investigation. As a matter of this fact, if ethics is to play a vital role in building a morally developed university with communities united by strong ethical ties, the key stakeholders, including the professors and managers of such universities, must be ethical; as must the system and practices to be developed. Thus, there are some ethical concerns arising, among others, from the commercial exploitation of learners, the academic duties of integrity and care, as well as research ethics concerns arising from the analytical and other work being done by both academics and institutions. It is also important to highlight how in an ethical university, the integral development of the student must be pursued, treating him/her as a valuable member of society. This is why it is of the utmost importance to try to know, as far as possible, students´ opinions, concerns and desires - especially with the aim of responding to them in the most appropriate way possible.

KEYWORDS: University students, satisfaction, experience, online teaching, ethics.

REFERENCES

Díaz-Álvarez, C. and Cortés-Pedraza, S. (2006). El estudiante como cliente: riesgo para la calidad de la educación superior en Colombia. *Universidad Central Carrera*, *5*, 21-38.

Gento-Palacios, S.; Palomares-Ruiz, A.; García-Carmona, N, and González-Fernández, R (2012). Liderazgo educativo y su impacto en la calidad de las Instituciones Educativas. *XII Congreso*

Interuniversitario de Organización de Instituciones Educativas. Granada-España. Online en http://www.leadquaed.com/docs/artic%20esp/Liderazgo.pdf

Hackman, K. (2006). Using the service encounter to facilitate regulatory change. *Strategic Change*, 15(3), 145–152.

Hoekstra, J.C., Huizingh, E.K., Bijmolt, T.H. and Krawczyk, A.C. (2015). Providing information and enabling transactions: Which web site function is more important for success? *Journal of Electronic Commerce Research,* 16(2):81-94

Moliner-Velázquez, B.., Gallarza, M. G., Gil-Saura, I., and Fuentes-Blasco, M. F. (2015). Causas y consecuencias sociales de la satisfacción de los clientes con hoteles. *Cuadernos de Turismo*, (36), 295-313.

San Martín Gutiérrez, H., Collado Agudo, J. and Rodríguez del Bosque, I, (2008). El proceso global de satisfacción bajo múltiples estándares de comparación: el papel moderador de la familiaridad, la involucración y la interacción cliente-servicio. Revista Española de Investigación de Marketing ESIC Marzo 2008, vol. 12, No. 1 (65-95).

Sarmiento-Bojórquez, M. A. Cadena-González, M.., and Tuyub-Ovalle, T. D. C. (2017). Video interactivo como Objeto de Aprendizaje en la formación de los estudiantes de inglés en el nivel medio superior de la UAC. *Revista Electrónica Sobre Cuerpos Académicos y Grupos de Investigación*, *4*(8). http://www.cagi.org.mx/index.php/CAGI/article/viewFile/144/261

Oliver, Richard L. (1980). A cognitive model of the antecedents and consequences of satisfaction decisions. JMR, Journal of Marketing Research, 17(4), 460.

Oliver, R. L. (1981). Measurement and Evaluation of Satisfaction Process in Retail Setting, Journal of Retailing, 57, 25-48.

Oliver, R. L. (1993). Cognitive, Affective, and Attribute Bases of the Satisfaction Response, Journal of Consumer Research (20), pp. 418-430.

Oliver, R. L. (1997). Satisfaction: A behavioral perspective on the consumer. New York, NY, McGraw-Hill.

Retolaza, A., y Grandes, G. (2003). Expectativas y satisfacción de los usuarios de un centro de salud mental. Actas Españolas de Psiquiatría, 31(4), 171-176.

COMPUTER ETHICS IN BRICKS

Gosia Plotka, Bartosz Marcinkowski

De Montfort University (United Kingdom), University of Gdansk (Poland)

malgorzata.plotka@dmu.ac.uk; bartosz.marcinkowski@ug.edu.pl

EXTENDED ABSTRACT

Modern businesses are increasingly demanded by the society to comply with the standards of Corporate Social Responsibility (CSR) (Kim & Han, 2019; Patrignani & Kavathatzopoulos, 2016). Not only failing to meet these standards exposes an organization to a number of risks – such as contributing to environmental or corruption-related scandals, consumer boycott, or even state intervention – but also prevents the organization from taking advantage of certain opportunities. Such opportunities may include, but are not limited to, increasing brand recognition, attracting investors, increase workforce commitment, retaining consumer loyalty, or improving bottom line. One might argue that computer professionals are in a specific position in the context of this trend. On the one hand, the level of their access to corporate data often makes them vital links in keeping an eye on CSR of parent organizations. On the other, they often elude institutional CSR programs due to the high demand for IT professionals on both the freelancers and start-ups markets. Therefore, raising awareness regarding the potential impact of their decisions on individuals, society and environment among future computing professionals cannot be overestimated.

The title of this paper has been inspired by the movement called *ethics in bricks* that uses popular Lego bricks to disseminate and explain some of ethical dilemmas and concepts on Social Media (Twitter, Facebook and Instagram). This paper aims at presenting outcomes of R=T (i.e. research equals teaching) study on how to make some troublesome content popular and easier to understand and equip students in soft skills that are not the most natural one for those studying computer science. It is computer ethics that is among the topics covered by such content. The research is based on ten years of observation, focus groups and interviews.

Costa & Pawlak (2018) in their abstract submitted to ETHICOMP2018 summarise some previously expressed views on how practical computer ethics should look like. They bring up an assessment by Soraker (2010), who highlights that (1) the bulk of computer ethics-related literature is directed towards other computer ethicists; (2) is simply boring; (3) explore topics, which are self-evident; (4) is irrelevant to the actual practice of software engineering. Another highlighted view comes from Connolly & Fedoruk (2014). They state that education in computer ethics is theoretically unsound and empirically under-supported. Moreover, ICT professionals need to explicitly understand the social contexts of computing – while faculty staff ought to put significantly less focus on ethical evaluation. Costa & Pawlak (2018) argue that despite the case studies, recent publications continue strategy that features lack of social context or people's behaviour/physiological response.

There is also an expectation that computing-related courses ought to be accredited by professional bodies such as BCS to ensure that the students gain industry-standard training/skills and are prepared for employment upon graduation (Times Higher Education, 2017). Both honesty and ethicality are included on the list of skills that are highly demanded by the labour market (Chartered Management Institute, 2018; Lindley et al., 2013). Therefore, it seems to be vital to provide future computing professionals with the relevant information regarding their potential impact on individuals, society and

environment (Tassone & Eppink, 2016; Voskoglou & Buckley, 2012) on top of the knowledge on some more technical aspects of system development.

The contributors aim at delivering a set of guidelines on how ethical and professional values should be incorporated into curriculums and presented to students, so they see such values as must-have competencies for their sustainable development that meets today's and future job market needs. The novelty of this research lies in identifying the needs of different stakeholders involved in the educational process – including students, teachers, ICT industry, society and professional bodies like BCS, IEEE, ACM (Tassone & Eppink, 2016; Voskoglou & Buckley, 2012) as well as naming a computer science threshold concept. Therefore applying design thinking defined as *a solution-based approach to finding what would-be users really need* (Interaction Design Foundation, 2019) complemented by Soft System Methodology (SSM) described as *a systems-based methodology for tackling real-world problems in which known-to-be desirable ends cannot be taken as given* (Checkland, 1981) suits really well in this type of research.

This research uses threshold concept, one of the educational principles to make teaching computer ethics more effective. The threshold concepts introduced by Mayer and Land (2003) enable a new way of thinking about a phenomenon, thus enhancing the students' ability to master their subjects (Advance HE, 2015). Some of the well-known concepts include reading or gravity. Threshold concepts are described as (Flanagan, 2018):

- Transformative – once understood, a threshold concept changes the way in which a student views the discipline;
- Irreversible – difficult to unlearn;
- Integrative – once learned, are likely to bring together different aspects of the subject;
- Bounded – delineate a particular conceptual space, serving a specific and limited purpose;
- Troublesome – likely to be counter-intuitive, alien or seemingly incoherent;
- Reconstitutive – may entail a shift in learner subjectivity;
- Discursive – will incorporate an enhanced and extended use of language.

The longitudinal study conducted across years of practice of this paper's contributors led them to the conclusion that this "conceptual gateway" into computer science with taking into consideration professional guidance (Chartered Management Institute, 2018; Lindley, et al., 2013) is understanding of the idea of context.

As the threshold concept is often described as troublesome, knowledge teaching requires the approach that helps students with engaging effectively in their own learning process (Barton & James, 2017). Namely, using metaphor and applying active and creative methods such as Lego Serious Play, reframing seems to give a very positive outcome as thinking with hands and depersonalisation by employing social constructivism principles facilitates interaction with the properties of the problem (Vallée-Tourangeau & Vallée-Tourangeau, 2016).

Lego® Serious Play® (LSP) is a method introduced in 2010 by the Lego Group to support communication and problem-solving. It has been recognised that cognitive processes – such as learning and memory– are highly influenced by the way people use their bodies to interact with the physical world (Gauntlett, 2010). "Talking and thinking with hands" is a powerful way of overcoming some barriers with expressing an opinion and reflecting on own work or discussed topic (Executive Discovery, 2014). LSP

can be used as an alternative tool to ideate (brainstorm) and conceptualise the outcome – for example during meetings or focus groups. But it may also be regarded as a way to develop perspective thinking that (1) helps to embrace diversity; and (2) facilitates depersonalization to enhance the sense of security in the narrative process and to carry through reframing personal experience (Harn, 2018). Also, design thinking methodology is open for creative techniques using LSP to generate (ideation phase) and synthesis data gives very good results.

KEYWORDS: Ethics, Lego Serious Play, Curriculum, ICT.

REFERENCES

Advance HE (2015). Threshold Concepts. Retrieved from https://www.heacademy.ac.uk/knowledge-hub/threshold-concepts

Brown, T., & Wyatt, J. (2010). Design Thinking for Social Innovation. Development Outreach, 12(1), 29-43

Burton, E. (2017). Ethical Considerations in Artificial Intelligence Courses. Cornell University Library. Retrieved from https://arxiv.org/abs/1701.07769

Chartered Management Institute (2018). All University Students Must Gain Leadership Skills, Says New Employability Report. Retrieved from https://www.managers.org.uk/about-us/media-centre/cmi-press-releases/all-university-students-must-gain-leadership-skills

Checkland, P. (1981). Systems Thinking, Systems Practice. Wiley

Connolly, R., & Fedoruk, A. (2014). Why Computing Needs to Go Beyond Good and Evil Impacts. ETHICOMP2014, Liberty and Security in an Age of ICTs. The University of Pierre and Marie Curie

Costa, G., & Pawlak, P. (2018). Practical Computer Ethics – An Unsolved Puzzle! Creating, Changing, and Coalescing Ways of Life with Technologies. Polish-Japanese Academy of Information Technology

Executive Discovery (2014). The Science of LEGO® SERIOUS PLAY™. Retrieved from https://thinkjarcollective.com/wp-content/uploads/2014/09/the-science-of-lego-serious-play.pdf

Flanagan, M.T. (2018). Threshold Concepts: Undergraduate Teaching, Postgraduate Training, Professional Development and School Education. Retrieved from https://www.ee.ucl.ac.uk/~mflanaga/thresholds

Gauntlett, D. (2010). Introduction to LEGO® SERIOUS PLAY® Retrieved from https://davidgauntlett.com/wp-content/uploads/2013/04/LEGO_SERIOUS_PLAY_OpenSource_14mb.pdf

Harn, P.-L. (2018). LEGO®-Based Clinical Intervention with LEGO®SERIOUS PLAY® and Six Bricks for Emotional Regulation and Cognitional Reconstruction. Examines in Physical Medicine & Rehabilitation, 1(3), 1-3

Interaction Design Foundation (2019). Design Thinking. Retrieved from https://www.interaction-design.org/literature/topics/design-thinking

Kim, H., & Han, J. (2019). Do Employees in a "Good" Company Comply Better with Information Security Policy? A Corporate Social Responsibility Perspective. Information Technology & People, 32(4), 858-875

Lindley, D., Aynsley, B., Driver, M., Godfrey, R., Hart, R., Heinrich, G., Unhelkar, B., & Wilkinson, K. (2013). Educating for Professionalism in ICT: Is Learning Ethics Professional Development? In J.

Weckert, & R. Lucas (Eds.), Professionalism in the Information and Communication Technology Industry (pp. 211-232). ANU Press

Meyer, J., & Land, R. (2003). Threshold Concepts and Troublesome Knowledge: Linkages to Ways of Thinking and Practising within the Disciplines. In C. Rust (Ed.), Improving Student Learning: Theory and Practice – 10 Years on (pp. 412-424). Oxford Brookes University

Patrignani, N., & Kavathatzopoulos, I. (2016). Cloud Computing: The Ultimate Step Towards the Virtual Enterprise? ACM SIGCAS Computers and Society, 45(3), pp.68-72

Plattner, H., Meinel, C., & Leifer, L. (2015). Design Thinking Research: Making Design Thinking Foundational. Springer International Publishing

Soraker, J. (2010). Designing a Computer Ethics Course from Scratch. Retrieved from http://www.soraker.com/designing-a-computer-ethics-course-from-scratch/

Tassone, V., & Eppink, H. (2016). The EnRRICH Tool for Educators: (Re-)Designing Curricula in Higher Education from a "Responsible Research and Innovation" Perspective. Wageningen University

Times Higher Education (2017). Computer Science – De Montfort University. Retrieved from https://www.timeshighereducation.com/world-university-rankings/de-montfort-university/courses/computer-science

Vallée-Tourangeau, G., & Vallée-Tourangeau, F. (2016). Why the Best Problem-Solvers Think with Their Hands, as well as Their Heads. The Conversation, Retrieved from http://theconversation.com/why-the-best-problem-solvers-think-with-their-hands-as-well-as-their-heads-68360

EDUCATIONAL GAMES FOR CHILDREN WITH DOWN SYNDROME

Katerina Zdravkova

University Ss. Cyril and Methodius, Faculty of Computer Science and Engineering (N. Macedonia)

katerina.zdravkova@finki.ukim.mk

EXTENDED ABSTRACT

The commonness of Down syndrome (DS) increased worldwide. Inclusive education, which embraces educational, social and emotional practises, based on well-structured instructions, interventions and support in the classroom is extremely valuable. In parallel with the in-class activities, educational software stimulates the inclusion. This paper presents the recommendations for such educational applications together with the pilot study intended for acquiring basic learning skills. The developed educational game was presented to children in the Day Care Centre for DS (DCCDS) in Skopje. Their enthusiasm and interest to use it is the greatest motivation to carry on with the study and after an approval by the experts and parents to offer it as a mobile application.

INTRODUCTION

Regardless of the considerably improved prenatal detection, the incidence of this congenital anomaly increased worldwide. People with DS deserve the same opportunities and care as others, which results in increased life expectancy and better quality of life. This can be achieved by constant parental care and support, monitoring of the mental and physical conditions, medical therapies, and consistent community support (Reid, 2018).

Inclusive education proved to be the best way to provide educational, social and emotional benefits starting from very early childhood (Felix, 2017). Moreover, it changed the attitudes towards this disability and improved the interaction with children with DS (Campbell, 2003). If well designed and implemented, specially created educational applications can significantly facilitate the process of inclusive education, enhancing the cognitive and learning skills of these vulnerable children.

FEATURES OF EDUCATIONAL SOFTWARE INTENDED FOR DS

Educational software aims to stretch the abilities of their users. However, children with DS are usually gifted for one-type skills: language, math, strategic thought or physical coordination. They typically manifest a deficit of attention; thus they are not capable of comprehending longer or more complex rules (Mason, 2015). Children with DS are not patient to wait for the application to download or to process the following steps (Skotko, 2005). They also need instant rewards for each successful outcome. Furthermore, DS children have significant vision deficit and anomalies in colour discrimination (Krinsky-McHale, 2014), and a lack of control of muscles stiffness affecting their motor skills (Vicari, 2006). These cognitive and neuropsychological profiles, amplified with the guidelines for supporting children with disabilities (Encarnação, 2018) and the recommendations of the specialists from DCCDS resulted in the following conceptual design criteria:

1. Intuitive gameplay with easy navigation and few, simple functionalities accessible by clicking over a perceptive icon, which is active throughout the whole image;

2. Clear interface with bright colours, clear contours, realistic and simple images, and without anthropomorphic features or facial expressions (Lee, 2018);
3. Substituted single and double finger gestures by two touches: from the source place to the target (Landowska, 2018);
4. Virtual tutor who announces the game, and responds with an appropriate facial and voice expression (Herring, 2017);
5. Simple and unambiguous instructions, which are repeated whenever an image is touched;
6. Adjustable progression pace, based on the performance of the DS child;
7. Learners are not capable of reading, so the instructions should be spoken or presented with the sign language;
8. Quick download and very short waiting time to advance from beginning to end;
9. Free of charge.

CREATED APPLICATION

The application consists of three integral parts: developing literacy skills, developing basic mathematical competencies and practising memory (Fig. 1.). The screens have a white background, few images and intuitive navigation.

To give learners an opportunity to set their own pace, and to enable the progress, all three parts have three levels, starting from the simplest and ending with the most advanced.

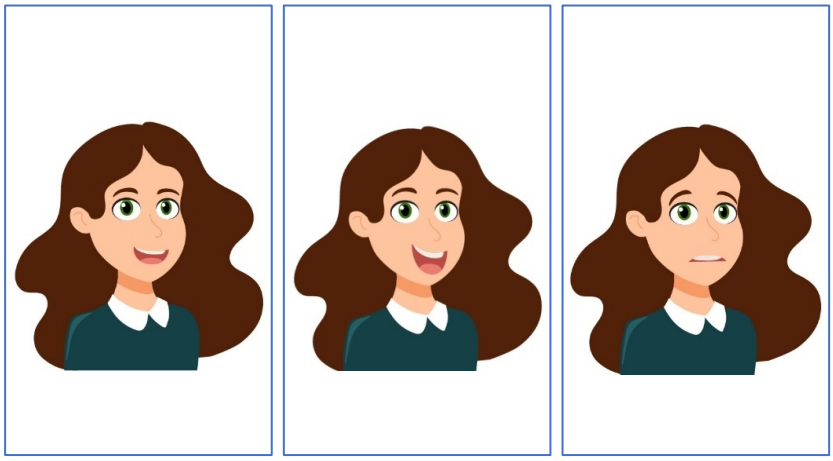
Figure 1. Memory game, intermediate level: coupling equal images, the initial letter with an image, the written name with an image, and numbers with a word

Implemented by Davor Trifunov

Each part of the application has its virtual tutor (Fig. 2), who has a full and deep voice and a perfect pronunciation. Tutors introduce the task, the levels, speaks out the names of the touched objects or pf the two navigation icons (Fig. 1).

If the learner accurately performs the task, tutor's face smiles and says a randomly picked congratulation with a happy voice. If the learner has failed, tutor's face becomes sad. After three wrong attempts, sad faced tutor suggests to repeat the task with a calm voice. After five consecutive mistakes, the advice is to go back to the previous level or to ask for help.

Figure 2. Three tutor's moods: instructional, happy and sad



Designed by Ana Zdravkova

FEEDBACK

Seven young boys and two girls aging from 15 to 19 and their parents were the first evaluators of the application. The game was installed on one tablet and demonstrated to every child individually. The age and the basic reading skills enabled them to successfully play the memory game. The whole event was touching for everyone. The kids were noticeably amused and attracted, except one girl, who was too shy. She listened the tutor with great attention and observed how the others played.

The most experienced boy comprehended the game immediately and asked to play the first. After trying all the options several times, he generously let others play. He manifested his frustration from the absence of an immediate congratulation after each successful coupling by lifting the speaker to hear the greeting.

Other five kids explored him, tried the game and managed to play it independently. The most extrovert boy succeeded after several trials and errors, and then tried to download the game from Google Play. Two kids, a boy and a girl created a strategy to first open all the tiles, and then couple them.

Two boys were not competent with the written words, one couldn't even discover the initial characters. They turned to the easier level of the game of own accord and were not enthusiastic to play it again.

During the second visit, all the kids, except the shy girl, activated the game and played it more competently, including the boys with lower literacy skills.

CONCLUSIONS

The ultimate goal of DCCDS is to prepare the kids for an independent life. They started making own meals under a full supervision of DCCDS staff and organized a cocktail with self-made bread and snacks. The next stage is to purchase the ingredients and start cooking according to a written recipe. To

achieve this goal, their literacy and understanding of quantities should increase significantly. According to DCCDS staff and their parents, the educational game will be of a great use.

The major challenge is the indifference and the anxiety of some kids. Hopefully, they are very confident in using the smart phones. Before launching it on Google Play, the application will be polished and upgraded with new contents suggested by the specialists from DCCDS. As a consequence, those kids who were shy to show their incompetence or who were not interested to use it will be able to experience it with the support by their family members.

The educational game is in Macedonian only. It can easily be adapted to other languages, making it available to wider community.

KEYWORDS: Down syndrome, educational software, life skills, mobile and tablet applications.

REFERENCES

Campbell, J., Gilmore, L., & Cuskelly, M. (2003). Changing student teachers' attitudes towards disability and inclusion. *Journal of Intellectual and Developmental Disability*, *28*(4), 369-379.

Encarnação, P., Ray-Kaeser, S., & Bianquin, N., 2018. *Guidelines for supporting children with disabilities' play: Methodologies, tools, and contexts*. De Gruyter Open.

Felix, V., Mena, L., Ostos, R., & Maestre, G. (2017). A pilot study of the use of emerging computer technologies to improve the effectiveness of reading and writing therapies in children with Down syndrome. *British Journal of Educational Technology*, *48*(2), 611-624.

Gilmore, L., Campbell, J., & Cuskelly, M. (2003). Developmental expectations, personality stereotypes, and attitudes towards inclusive education: Community and teacher views of Down syndrome. *International Journal of Disability, Development and Education*, *50*(1), 65-76.

Herring, P., Kear, K., Sheehy, K., & Jones, R., 2017. A virtual tutor for children with autism. *Journal of Enabling Technologies*, *11*(1), 19-27.

Krinsky-McHale, S., Silverman, W., Gordon, J., Devenny, D., Oley, N., & Abramov, I. (2014). Vision deficits in adults with Down syndrome. *Journal of Applied Research in Intellectual Disabilities*, *27*(3), 247-263.

Landowska, A., 2018. 8 Which digital games are appropriate for our children?. In *Guidelines for supporting children with disabilities' play*, 85-97.

Lee, J.M., Baek, J., & Ju, D.Y. (2018). Anthropomorphic Design: Emotional Perception for Deformable Object. *Frontiers in psychology*, *9*, 1829.

Mason, G.M., Spanó, G. & Edgin, J. (2015). Symptoms of attention-deficit/hyperactivity disorder in Down syndrome: effects of the dopamine receptor D4 gene. *American journal on intellectual and developmental disabilities*, *120*(1), 58-71.

Reid, W., Balis, G., Wicoff, J., & Tomasovic, J. (2018). *The treatment of psychiatric disorders*. Routledge.

Skotko, B. (2005). Mothers of children with Down syndrome reflect on their postnatal support. *Pediatrics*, *115*(1), 64-77.

Vicari, S. (2006). Motor development and neuropsychological patterns in persons with Down syndrome. *Behavior genetics*, *36*(3), 355-364.

IMPACT OF EDUCATE IN A SERVICE LEARNING PROJECT. OPENING UP VALUES AND SOCIAL GOOD IN HIGHER EDUCATION

Ana María Lara-Palma, Montserrat Santamaría-Vázquez, Juan Hilario Ortiz-Huerta

Universidad de Burgos (Spain)

amlara@ubu.es; msvazquez@ubu.es; jhortiz@ubu.es

EXTENDED ABSTRACT

The rate at which universities have been assimilating proposals in their educational environments has been constant. Since the first meeting at Praga in May 2001, efforts drive on getting improvements in specific and transversal competences of the students. Nowadays, eighteen years later, it is still present the way of doing innovation at the universities, and, the sustainable human development concept has been included within the topics and guides. Quoting Brotóns, (2009), "to improve the quality of teaching is mandatory to create real learning situations: with new innovative tasks, thinking in a positive ICT future and with acquisition, transfer and updating knowledge processes".

In the academic course 19-20, the University of Burgos has launched a call for Service Learning Projects (SLP) with the aim of reinforce the academic skills of the students endorsing a societal transformation. These projects nonetheless pursue very worthwhile goals: to educate students at higher education not only in cognitive aspects, but also in personal growth. Service Learning is a groundbreaking and appeal methodology focused in acquire knowledge by doing community service work.

Quoting Folgueiras, Luna and Puig (2011, pp.159), learning by using Service Learning tasks enhance students to "take part directly with those who are supporting, adapting to their needing's and facing up a realistic circumstance, really different from the classroom lectures and environments".

This paper describes a real case scenario composed by students of the University of Burgos from two different careers at the University of Burgos (Degree in Occupational Therapy and Degree in Management Engineering). (N=54 3th, 4th grade students). All have contributed to resolve three different challenges by using skills, competences and knowledge of their corresponding disciplines (health and engineering). From the academic point of view, students participate in meetings, visit the institutions and the users, attend creative thinking workshops and join their ideas to build realistic and feasibility solutions. From the scientific scenario, the aim is in three directions: first, to identify relevant learning/ethical/social theories and find out how this disruptive and innovative methodology relate to them. Second, to compare this scheme of work with design thinking methodologies. Third, to analyse differences in teamwork and the creative process due to the gender diversity of team´s members.

The three defiances are settled in three local entities, Cerebral Palsy Association (APACE BURGOS), Day Care Centre (Puerta del Parral) and the Spanish Confederation of Physical and Disable People in Castilla y León (COCEMFE CYL) and has run for 10 weeks of length (October-December 2019). In APACE, the work has consisted of designing and producing (manufacture with real materials) a product to improve the quality of life of people with cerebral palsy and trying to find parameters to help them to be more independent in their daily life (specifically, a bracket to facilitate the usability of a tablet and a mechanism for using a wheelchair). In the second organization, the Day Care Center Puerta del Parral, the collaboration is to design and manufacture low cost tools for helping a survivor person with brain stroke (the gadget is a tray where the person can eat only with one hand). And, finally, in the third entity, the confederation COCEMFE CYL, the aim of the tasks is to design a mechanism to help a

survivor person with brain stroke to get dressed without any exterior help and support (human being) (a support product with a spring and elastic mechanism to help with the socks by using only one hand).

According the methodology, the scheme of work has consisted of four stages: the first one for organizing the teams and groups of work; the second one, for acquiring conceptual and practical knowledge by doing learning workshops; the third one, for developing the projects (students use a novel Technological Center of the University of Burgos equipped with new and recent IT technologies such as 3D Printers, electronic devices, machines and material for design and manufacture the support products, and, the fourth one, supervisors have developed a survey for collecting data of satisfaction by using a rubric based on Campo (2015). The expectation about the effects of the use of new learning methodologies (Service Learning Projects) for Higher Education and its repercussion on cognitive competences, social competences, ethical competences and professional competences are the basis of the Hypothesis. Sample is divided in two groups: 27 students in the intervention group (SLP) and 27 students in the control group. For the quantitative analysis, the mathematical tool used is SPSS v24.

All in all, the aim of this paper is to reflect the benefits of applying these innovative activities, not only educating in a service learning project, but, still more, joining students from very different careers and supporting them to work, create, debate, interchange and build together. In our real case, not only the university community, but also the society (collaborative entities) get benefits as well; they both mutually reinforce through these implementations. The contribution adds a picture of how to achieve cognitive competences (technical), social competences (consciousness), ethical competences (compassion) and professional competences (productive versatility), among others. This novel scenario serves for a purpose: opening up values and social good in Higher Education.

KEYWORDS: Service Learning Projects, Innovative Education, Higher Education, Cognitive Competences, Social Competences, Ethical Competences, Professional Competences.

REFERENCES

Brotóns-Cano, R., Lara-Palma, A. M., Stuart, K., Karpe, J., Faeskorn-Woyke, H. & Poler, R. (2009): Competitive Universities need to Internationalize Learning: Perspectives from three European Universities. *Journal of Industrial Engineering and Management*, 2(1), 299-318. Retrieved from http://www.jiem.org/index.php/jiem/article/view/89

Campo-Cano, L. (2015): Una rúbrica para evaluar y mejorar los proyectos de aprendizaje servicio en la Universidad.*Revista Iberoamericana de Aprendizaje Servicio* (RIDAS), 1, 91-111.

Casado-Muñoz, R., Greca I., Tricio-Gómez, V., Collado-Fernández, M. & Lara-Palma, A. M. (2014): Impacto de un Plan de Acción Tutorial Universitario. Resultados Académicos, Implicación y Satisfacción. Revista de la Red Estatal de Docencia Universitaria (REDU) 12(4), 323-340. Retrieved from http://red-u.net/redu/index.php/REDU/article/view/587.

Folgueiras Bertomeu, P., Luna González, E., Puig Latorre, G. (2011): Service Learning: study of the degree of satisfaction of university students. Revista de Educación, 362, 159-185

Lara-Palma, A. M., Cámara-Nebreda, J. M., & Vicente-Domingo E. M. (2015): Impact level within the English Friendly Program in Virtual Degrees at the University of Burgos. A Real Case Scenario of Transnational Education. *Advances in Intelligent Systems and Computing*, 369, 525-532. Springer International Publishing Switzerland. Retrieved from http://www.scopus.com/inward/record.url?eid=2-s2.0-84946779842&partnerID=MN8TOARS

Lara-Palma, A. M. & Giacinto, R. (2014): Improving the Effectiveness of Virtual Teams: Tackling Knowledge Management and Knowledge Sharing. A Real Case Scenario. *Journal of Social Sciences* (COES&RJ-JSS), 4(1), 626-634. Centre of Excellence for Scientific & Research Journalism, COES&RJ LLC.

Lara-Palma, A. M. & Collado-Fernández, M., Tricio-Gómez, V. (2010): Improving Education: A Knowledge Transfer Map Proposal for University Tutorship. 5th International Conference of Education, Research and Innovation. 6043-6050. Retrieved from https://library.iated.org/view/LARAPALMA2012IMP

Lorenzo, C. & Lorenzo, E. (2020): Opening Up Higher Education: An E-Learning Program on Service-Learning for University Students. *Advances in Intelligent Systems and Computing*, 963, 27-38. AHFE International Conference on Human Factors in Training, Education, and Learning Sciences

Maquilón Sánchez, J. J. & Alonso Roque, J. I. (et. al.) (2014): Experiencias de innovación y formación en educación.

Torres-Coronas, T., Arias-Oliva, M., Yáñez-Luna, J. C., Lara-Palma, A. M. (2015): Virtual Teams in Higher Education: A Review of Factors Affecting Creative Performance. *Advances in Intelligent Systems and Computing*, 369, 629-637. Springer International Publishing Switzerland. Retrieved from http://www.scopus.com/inward/record.url?eid=2-s2.0-84946771368&partnerID=MN8TOARS

IMPROVEMENTS IN TOURISM ECONOMY: SMART MOBILITY THROUGH TRAFFIC PREDICTIVE ANALYSIS

Juan Guerra-Montenegro, Javier Sánchez-Medina

Universidad de Las Palmas de Gran Canaria (Spain)

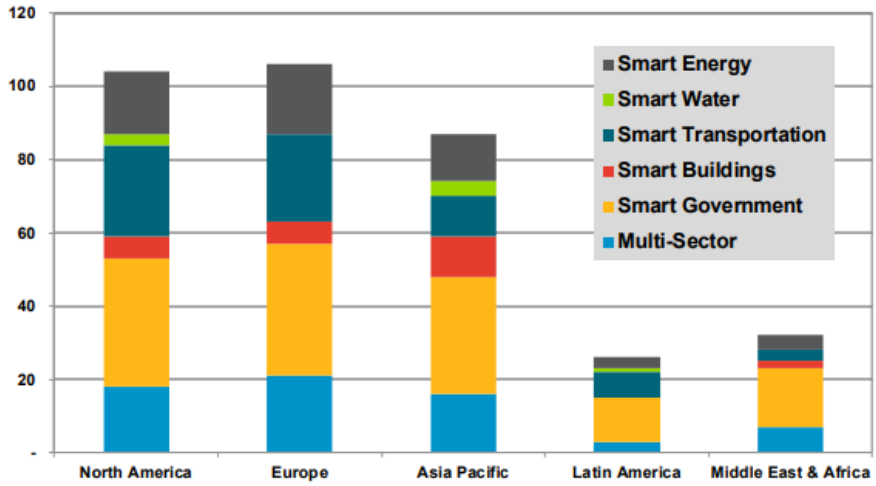
juanantonio.montenegro@ulpgc.es; javier.sanchez@ulpgc.es

EXTENDED ABSTRACT

We live in a world that gets more interconnected as the time passes. People has the ability to travel anywhere, which creates huge quantities of data flows. The use of ICTs in modern transport technologies in order to improve urban traffic, also known as Smart Mobility (Albino, Berardi and Dangelico, 2015), is changing the transportation paradigm. Mobility is converging with the digital industry, creating this new trend inside the transportation panorama (Noy and Givoni, 2018).

Since mobility is one of the key aspects of any city, the importance of Smart Mobility inside a Smart City should not be oversight. In fact, according to a study published by Navigant (Navigant Research, 2018), it can be seen in Figure 1 that 355 smart city projects were active worldwide, and that Smart Transportation was an important part of the development of these Smart City Projects. This also serves as a demonstration of how different governments around the world are adopting the Smart City Paradigm.

Figure 1. Active Smart City Projects (2018)



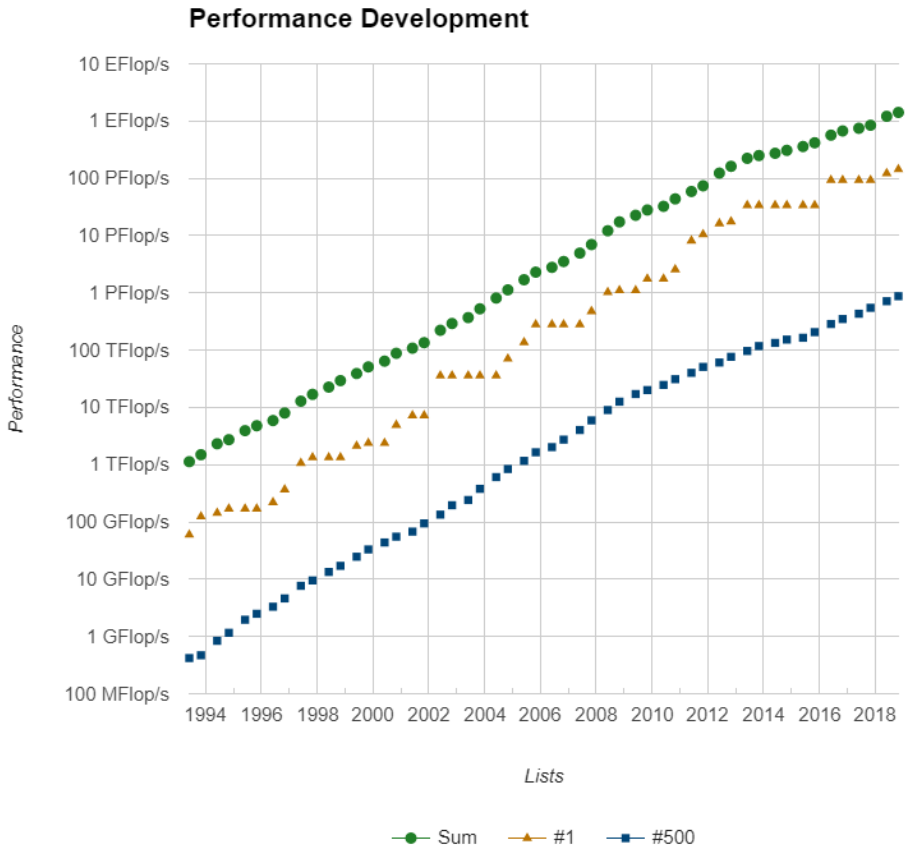
Source: Navigant Research. Smart City Tracker (2018)

Cities are slowly adding smart capabilities to their infrastructures and services. Data and technologies have been proven to be important parts of a Smart City ecosystem, but there are other layers that share the same importance. In fact, according to Gil-García, Pardo and Nam (2015) a Smart City is composed by four main layers: Physical Environment, Society, Government (where Smart Mobility is located) and Technology + Data, making Smart Mobility one of the key components of a Smart City.

Tourism is also getting smarter by applying different ICT principles and systems to improve different areas. Smart Tourism is composed by many layers that are supported by ICTs. These layers were defined by Gretzel et al. (2015) as: Smart Experience, which contains technology-mediated experiences; Smart Business Ecosystem, which creates and supports the exchange of touristic resources; and Smart Destinations, which has been demonstrated to be a special case of a Smart City. Taking this into account, it can be stated that Smart Mobility is a key component inside a Smart Destination. Since it has also been stated that a Smart Destination is a special case of a Smart City, this also means that Smart Mobility is also a key component inside a Smart City. This demonstrates that Smart Mobility and Tourism are related, and that improvements and researches made on the Smart Mobility area might benefit Tourism Economy.

Smart Mobility usually relies in predicting traffic patterns and driver behaviour. Traffic prediction analysis has improved in various ways in the last years, along a noticeable increment in available vehicle data thanks to the inclusion of advanced data management systems (Chang and Yoon, 2018) and an increasing trend in computational power, as seen in Figure 2. This increment is of paramount importance to generate more accurate classification and forecasting models. However, classic traffic forecasting methodologies tend to be insufficient in a scenario where enormous amounts of traffic spatio-temporal data need to be consistently processed in real time (Chindanur and Sure, 2018). Additionally, data could present changes inside its distribution (also known as Concept Drift), making predictive models lose accuracy over time.

Figure 2. Computational power increment.



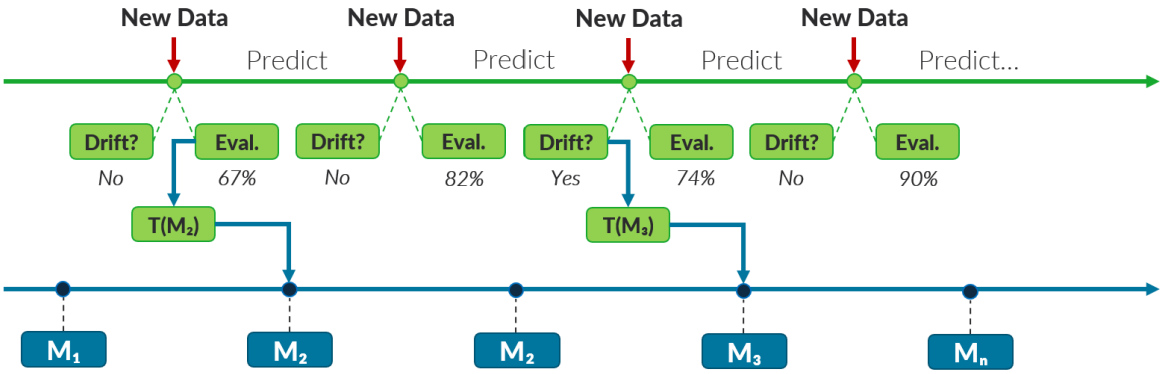
Source: Top500 (*https://www.top500.org/statistics/perfdevel/*)

In this paradigm Online Learning methodologies are useful because of their capacity to adapt and overcome these changes. Online Learning refers to machine learning methods in which data arrives in a sequential manner and is used to update the best predictor at each step instead of learning on the entire dataset at once. This new methodology allows to dispense of large data storage systems and creates predictive models that adapt to the concept drift present in sequential data patterns.

Bifet (2009) states in his research that digital data can grow without limit at high rates, even millions of data items per day, meaning that a uniform number of samples is generated continuously by different sources, which is similar to traffic forecasting. Given the huge size of the generated data, it is not possible to process it every time new data arrives, but it is feasible to process this data in lesser quantities to create new, updated prediction models in real-time.

In Figure 3, a possible configuration for a Data Stream Mining system is depicted where data is analysed as it arrives. To solve the data drift problem, the model M processes new data to find out if this data presents a drift, or evaluates the model to assure its predictive accuracy, thus updating itself if necessary with a newer version trained with the data it just analysed. This behaviour allows to overcome the Concept Drift present in this kind of system, ensuring an adaptive model that will resist the wear of time.

Figure 3. Possible Layout of a Data Stream Mining system.



Source: Own elaboration.

Data Stream Mining offers a solution for processing traffic data in real-time thanks to its adaptive capabilities, which are useful in a traffic ecosystem that presents different distributions over time that might make classic predictive models inaccurate or even useless. Applying Online Learning as a traffic predictive analysis methodology could Smart Mobility inside a Smart City by creating more accurate, adaptive traffic forecasting models and lowering the costs of the ICT structure inside a Smart City because of its inherent ability to process data in real-time instead of pre-processing Big Data stored in expensive data warehousing infrastructures. This could ultimately improve the overall Tourism Economy by predicting traffic flows in urban touristic areas.

KEYWORDS: Smart Mobility, Tourism Economy, Predictive Analysis.

REFERENCES

Albino, V., Berardi, U. & Dangelico, R.M. (2015). Smart Cities: Definitions, Dimensions, Performance, and Initiatives. *Journal of Urban Technology,* 22(1), 3-21.

Navigant Research (2018). Smart City Tracker 1q18. Retrieved from www.navigantresearch.com/reports/smart-city-tracker-1q18

Gretzel, U., Sigala, M., Xiang, Z. & Koo, C. (2015). Smart tourism: foundations and developments. *Electronic Markets,* 25(3)

Noy, K. & Givoni, M. (2018). Is 'Smart Mobility' Sustainable? Examining the Views and Beliefs of Transport's Technological Entrepreneurs. *Sustainability* 10(2), 422.

Gil-Garcia, J.R., Pardo, T., Nam, T. (2015). What makes a city smart? identifying core components and proposing an integrative and comprehensive conceptualization. *Information Polity* 20, 61-87.

Chang, H.H. & Yoon, B.J. (2018). High-Speed Data-Driven Methodology for Real-Time Traffic Flow Predictions: Practical Applications of ITS*. Journal of Advanced Transportation*, 2018.

Top500 (2018). Performance Development. Retrieved from www.top500.org/statistics/perfdevel/

Bifet, A. (2009). Adaptive learning and mining for data streams and frequent patterns. *SIGKDD Explorations* 11, 55-56

Chindanur, N., & Sure, P. (2018). Low-dimensional models for traffic data processing using graph fourier transform. *Computing in Science and Engineering* 20(2), 24-37.

LESSON LEARNED FROM EXPERIENTIAL PROJECT MANAGEMENT LEARNING PEDAGOGY

Shalini Kesar, James Pollard

Southern Utah Univeristy, (USA), FIA, (USA)

Kesar@suu.edu; jamespollard1@suu.edu

EXTENDED ABSTRACT

This paper reflects on collaborative research to create an experiential learning pedagogy for a project management class. The motivation for this research was based on the authors belief that core skills such as teamwork, communication, professionalism and ethics are part of experiential learning pedagogy, which prepares the students to deal with challenges faced in today's technology related businesses. This class is taken by computer science and information system students prior to graduating. Prior to taking this class, the students are required to take the foundations classes in computer science and information systems courses. In addition to the foundation classes, the students also have taken classes that introduce them to topics linked with understanding the interface between computer software and hardware including processor architecture, computer arithmetic, instruction set architecture, and assembly language. They also learn about and conduct projects in higher-level languages, computer performance analysis, basic concepts of pipeline, introduction to memory management, Computer IO, and disk storage systems. Both computer science and information systems students are required to take programming language courses in C, C+ and Java Script. They are also required to take a class in database management systems that include: database processing, data modelling, database, database design, development, implementation, alternative modelling approaches, and implementation of current DBMS tools and SQL.

The Computer Science and Information Systems (CSIS) capstone project's gives senior-year students an opportunity to manage a major information systems development/enhancement project, in which they apply what they have learned in various other courses to a single project. The emphasis is on enterprise-level project management. This course's learning outcomes include to: 1) Provide senior-year students an opportunity to manage a major information systems development or enhancement project, which is proposed by the instructor or the students themselves. Students will utilize various technical skills that they have learned from various courses to a single project. Project management techniques will be emphasized in class; 2) Develop students' abilities to initiate, analyse, evaluate and manage an IS project in preparation for making informed decisions as a future IT project manager; 3) Develop students' ability to distinguish among opinions, facts, and inferences; to identify underlying or implicit assumptions; to make informed judgments; and to solve problems by applying evaluative standards when working with an IS project; 4) Provide students an understanding the challenges of different project stages, and will develop skills to understand and handle a variety of project management challenges; 4) Develop a comprehensive understanding of implementing and managing of an Information Systems project. Number of students range from ten to fifteen in each class annually. This class is for fifteen weeks and conducted in the last semester before the students graduate.

The context of the project is the Forest Service application, Design and Analysis Toolkit for Inventory and Monitoring (DATIM). This project is partially funded by the USDA Forest Service, Forest Inventory Analysis. This paper specially reflects on the four years of DATIM project's structure, results, and lessons learned. The goal was to facilitate an experiential learning environment to provide student's

flexibility to identify their milestones within the project scope. Based on the feedback from both students and client, every year the project and pedagogy style are modified. Experiential learning topics such as ethics were included as part of learning outcomes to help prepare computing students to meet global challenges that they may face in real business settings. The National Society for Experiential Education's framework was used while developing the curriculum. The theoretical framework is also used to reflect on the past and present's outcomes of the project as well to provide guidelines on curriculum robustness for an experiential learning project management class.

The National Society for Experiential Education (NSEE) consists of a group professional and academics dedicated to mutual learning and support across the varied roles and responsibilities represented in the field of experiential education. Founded in 1971. The members of NSEE advocate for the use of experiential learning throughout the educational system; to disseminate principles of best practices and innovations in the field; to encourage the development of research and theory related to experiential learning; to support the growth and leadership of experiential educators; and to create partnerships with the community. Since the founding of the Society, the Board of Directors, staff, and membership have been governed by policies and practices that guide ethical actions, relationships, and decisions. The distinctive purposes and conditions of experiential learning demand that all those involved in the process of learning through experience are held to the highest standards of mutual respect and responsibility, and that ethical behaviour is understood and practiced at every level of the learning process. Eight Principles of Good Practice for All Experiential Learning Activities include: Intention; Preparedness and Planning; Authenticity; Reflection; Orientation and Training; Monitoring and Continuous Improvement; Assessment and Evaluation; and Acknowledgment.

Finally, this paper also discusses the findings of various phases of the research and benefits that include: 1) student's feedback and suggestions were provided to the FS and were being used in a real business setting; 2) the modified pedagogy provided a classroom room environment with an experiential learning emphasis; 3) the collaborative work led to student employment; 4) modified pedagogy prepared computing students with skills that will help them when facing challenges in computing field; 5) success of the class strengthened the argument that pedagogy encompassing professionalism and ethics is important; 6) pedagogy needs to also emphasize team work, communication skills, leadership skills, critical thinking solving problems and finding alternative solutions. The recognition of the impact of this project management class and collaboration across disciplines in the institution has led to more opportunities for future capstone projects students to be exposed to experiential learning environment with real business settings.

KEYWORDS: Experiential learning, Pedagogy, Caspstone project, Forset Service, Information systems.

REFERENCES

Kesar, M., (2015). Including Teaching Ethics into Pedagogy: Preparing Information Systems Students to Meet Global Challenges of Real Business Settings, S. Kesar. ACM SIGCAS Computers and Society - Special Issue on Ethicomp, 45 (3).

National Society for Experiential Education. (1998) Eight Principals of Good Practices for ALL Experiential Learning Activities. Retrieved from https://www.nsee.org/8-principles

OVERCOMING BARRIERS TO INCLUDING ETHICS AND SOCIAL RESPONSIBILITY IN COMPUTING COURSES

Colleen Greer, Marty J. Wolf

Bemidji State University (USA), Bemidji State University (USA)

Colleen.Greer@bemidjistate.edu; Marty.Wolf@bemidjistate.edu

EXTENDED ABSTRACT

Currently, there is increased interest among computer scientists and humanities and social science faculty regarding what it means to incorporate ethics and social responsibility understandings and practices into computer science courses at US institutions. While some elements of this work are being considered at major research institutions (e.g., (Saltz, et al. 2019), (Groz, et al. 2019), (Burton, et al. 2018)) and private liberal arts colleges (e.g. (Shaer and Peck, 2018) there is little evidence that there are concerted efforts to extend this movement to public colleges and universities where faculty have high teaching loads (typically six or more courses each academic year). In this paper we describe a project that examines the barriers to adopting ethics and social responsibility more extensively throughout the computer science curriculum at these mid-range institutions. Adopting the position of Marty Wolf where he calls for experts from fields such as psychology, sociology, philosophy, ethics, and communication theory "to be drawn into collaborations as an integral part of the practice of computing in order to advance computing professionalism" (2016), we describe a process for establishing the sorts of collaborations that will enhance the expertise of computer science faculty in the areas of teaching ethical and socially responsible computing.

Our project consists of two phases. The first is an investigation into perceptions of computer science department chairs and faculty at target institutions surrounding notions of "ethical and socially responsible computing" and "collaboration." This phase involves two parts. In the first part, we interview computer science department chairs at US universities that are designated as primarily teaching institutions and where faculty have high teaching loads. Our line of inquiry includes a focus on the role of ethics and social responsibility in the department. We ask about the extent to which understandings and interpretations of ethics are present and practiced in curriculum and through particular pedagogical approaches. In addition, we ask questions of the chairs about perceived barriers to further engagement with social and ethical issues, and we gather information about how much consideration is given to "ethics" and "social responsibility" in the promotion and tenure process. To discern the extent to which engagement and practice of ethics in the classroom and in other aspects of faculty assignments, matters in professional development approaches and review, we make direct inquiry about how adopting new models of engagement are critiqued and supported in the tenure and promotion process. Questions about collaborative efforts among faculty have also been asked to discern how faculty in the departments engage in intradisciplinary and interdisciplinary efforts broadly speaking, and whether these efforts are rewarded by the institution in any particular manner. The second part of this phase of the project is a direct survey of the faculty that uses a quantitative instrument based on the same pattern of questions. This part seeks to expand our knowledge regarding faculty practices.

An initial review of responses from respondents in computer science departments point to the following understandings. First, computer science faculty have taken a narrow view of "ethically and socially responsible computing." The primary understanding has to do with the process of code

(software) development, and, when discussed, a fairly narrow interpretation of how these ethical matters are essential in a broader environment. Basically, the comments suggest that, students who get code—snippets or entire functions—from resources on the web need to understand how to do so "ethically," i.e., without plagiarizing. Incorporating ethics appears to be concentrated, to a great extent, in the upper division courses rather than across the curriculum. At this point we have not found evidence that the faculty are thinking broadly about social responsibility and how to bring in questions related to the various sectors of social environments that need to be considered, or heard from, prior to undertaking a project. So, while, in a secondary manner, the faculty note that professionalism needs to play a role in when and how to take on a project, their approach still reflects a narrow view through a disciplinary lens, not a socially responsive lens. Additionally, respondents do not engage in comprehensively assessing student learning regarding ethical and social responsibility, and we note that a primary barrier to expanding faculty practices in this area is awareness of the multifaceted nature of ethics as well as personal, professional, and social responsibility. The concept of a computing "problem" is narrowly framed, without discerning whether the "problem" should be addressed at all. We have not found evidence pointing to consideration of whether the "problem" is a secondary or tertiary factor in a more complex puzzle. Nevertheless, two realities that faculty face are that funding and time to engage in activities to address these sorts of considerations is in short supply, and building capacity is challenging.

The second phase of our project involves identifying computer science faculty and humanities and social science faculty who are interested in working together on course module redesign that will incorporate, for computer science students, more dimensions of social and ethical awareness; and for humanities and social science students greater understanding of the complexities and limitations of software design. Once faculty teams are identified, we will be going to a small number of campuses and offering a workshop for faculty with three goals. The first is for faculty to develop a better understanding of the ways that ethical and socially responsible computing can be incorporated into standard computer science courses. The second is to have computer science and humanities and social science faculty develop collaborative relationships with one another. The third goal is to have pairs of faculty (one computer science, one sociology/philosophy/history) develop two teaching modules: one for use in a standard computer science course and another for used in a standard humanities or social science course.

The full paper will begin with the results of our interviews with computer science department chairs as well as the results of our survey of computer science faculty. We will also describe the nature of the workshop we develop and the impact it has on breaking down the barriers to incorporating ethical and social responsibility into standard computer science courses and establishing new collaborations that serve to advance the incorporation of ethical and socially responsibility into the computing profession.

KEYWORDS: computing ethics, teaching computing ethics, integrating computing ethics, social responsibility in computing.

REFERENCES

Burton, E., Goldsmith, J., & Mattei, N. (2018). How to Teach Computer Ethics through Science Fiction. *Communications of the ACM*, 61(8), DOI: 10.1145/3154485

Grosz, B.J., Grant, D.G., Vredenburgh, V., Behrends, J., Hu, L., Simmons, A., & Waldo, J. (2019). Embedded EthiCS: Integrating Ethics Across CS Education. *Communications of the ACM*, 62(8), 54-61.

Saltz, J., Skirpan, M., Fiesler, C., Gorelick, M., Yeh, T., Heckman, R., Dewar, N., & Beard, N. (2019). Integrating Ethics within Machine-learning Courses. *ACM Trans. Comput. Educ.,* (19)4, Article 32, 26 pages. DOI: 10.1145/3341164

Shaer, O., & Peck, E. (2018). Teaching Pervasive Computing in Liberal Arts Colleges. *IEEE Pervasive Computing*, 17(3), 64-69. DOI: 10.1109/MPRV.2018.03367736

Wolf, M.J. (2016). The ACM Code of Ethics: A Call to Action. *Communications of the ACM*, 59(12), 6. DOI: 10.1145/3012934

START A REVOLUTION IN YOUR HEAD! THE REBIRTH OF ICT ETHICS EDUCATION

Simon Rogerson

Centre for Computing & Social Responsibility, De Montfort University (UK)

srog@dmu.ac.uk

EXTENDED ABSTRACT

This paper is a viewpoint rather than grounded in research. It questions some of the established ICT norms and traditions which exist both in industry and academia. The aim is to review current ICT ethics educational strategy and suggest a repositioning which aligns with the concept of computing by everyone for everyone. Whilst the *Educate for a Positive ICT Future* track call recognises the breadth of impact of technological advances, it seems to imply that the education of computing professionals is where the emphasis should lie in ensuring positive rather than negative impact. This will be challenged thereby contributing "to shift[ing] the paradigm towards a positive ICT future".

Computing is no longer the sole domain of professionals, educated and trained through traditional routes to service public and private sector organisations under paid contracts. Computing is now by everyone for everyone with the advent of economically accessible hardware, a multitude of software tools and the Internet (Rogerson, 2019). The IDC survey of 2018 found that there were, worldwide, 18,000,000 professional software developers and 4,300,000 additional hobbyists. The combined membership of leading professional bodies, ACM, ACS, BCS and IFIP represents only 3.09 per cent of that global total. The youngest app developer at Apple's Worldwide Developers Conference in June 2019 was Ayush Kumar aged 10 who started coding when he was 4 years old. He is not alone, 15 year old, Tanmay Bakshi, who is the world's youngest IBM Watson Developer, started software development when he was 5 years old. These facts suggest that professional bodies, in their current role, have little influence on 97 percent of global software developers whose ethical code and attitude to social responsibility comes from elsewhere.

It is now over a year since the launch of the new code of ethics for the ACM. At the last ETHICOMP conference much time was devoted to discussing the Code and the part it would play in moving ICT ethics forward. The code has spawned the ACM Integrity Project: Promoting Ethics in the Profession (https://ethics.acm.org/integrity-project/). The aim of this 2-year project of the ACM Committee on Professional Ethics is to promote ethics in the profession though modern media: YouTube videos, podcasts, social media, and streaming video. The use of modern media should certainly appeal to post millennials and offers a new approach to engage with future generations of computer scientists. Unfortunately, there has been little exposure of this project in, for example, the ACM's flagship publication, the Communications of the ACM. However, that same publication ran as its cover story in the August 2019 edition "Embedded EthiCS: integrating ethics across CS education" (Grosz et al). This is a paper about Harvard reinventing the wheel of computer ethics education which has a long and comprehensive history stretching back to the 1980s (see, for example, Aiken, 1983). It offers little new insight, does not link to a 40 year history and experience (for example see Pecorino and Maner, 1985; Martin, Huff, Gotterbarn and Miller, 1996; and Bynum and Rogerson 2004), and nor does it appear to connect with the Integrity Project.

Within industry and government, the compliance culture has taken a firm hold and so strangles the opportunity for dialogue and analysis of complex multi-faceted socio-ethical issues related to ICT. Superficial compliance is dangerously unethical and must be challenged vigorously in a technologically-dependent world. The timeframes for ICT development and ICT regulation and governance are, and will always be, misaligned. By the time some control mechanism is agreed, the technology will have moved on several generations and thus what has been agreed is likely to be ineffective. Currently, this seems to be the case with the governance of Artificial Intelligence, as there are so many opinions and vested interests causing protracted debate whilst AI marches onwards. Thus, it is paramount to imbue strategists, developers, operators and users with practical ICT ethics. In this way ethical computing has a chance of becoming the norm. Traditional approaches of professional bodies seem ineffective in a society which is moving rapidly towards complete dependency on technology

It is time to change. In the spirit of Kuhn (1962) we need a paradigm shift in ICT Ethics to address the societal challenges in the not-so-smart society of today. There needs to be a radical change in how the ethical and social responsibility dimension of ICT is included in education of the whole population rather than focusing on the elitist computing professional community. It is against this backdrop that this paper explores new avenues for widening education, both formal and informal, to all those who may become involved in computing. The approaches that are discussed also offer greater awareness to the public at large.

Thomas Kuhn (1962) suggests that scientific progress of any discipline has three phases: pre-paradigm phase, a normal phase and a revolution phase. Progress occurs when a revolution takes place after a dormant normal period and the community moves ahead to a paradigm shift. Given the ongoing frequent occurrence of ICT disasters, it seems ICT ethics education in its current dormant normal phase is in need of revolution. Four suggestions for a revolutionary approach are outlined:

Science and Technology Museums: An innovative interactive facility, Ethical Technology could be rolled out across the global network of science and technology museums and activity centres. It would be a programme for children and adults of all ages. It would be the catalyst for public awareness and public voice, schools' cross curricular activities, higher education research, teaching and learning, and new meaningful purpose for professional bodies.

Learning from history: Deborah Johnson (1985) stated that, "The ethical issues surrounding computers are new species of generic moral problems". She has been proved correct and consequently there is much to be learnt from the history of computers. The annals of ETHICOMP as well as the archives of trade journals provide a rich resource from which to learn. An interactive repository using a chronological taxonomy could be created to offer practical guidance to all those involved in computing.

Thought experiments: Brown and Fehige (2014) explain that thought experiments are used to investigate the nature of things through one's imagination. Usually they are communicated through narratives and accompanying diagrams. Brown and Fehige state that, "Thought experiments should be distinguished from thinking about experiments, from merely imagining any experiments to be conducted outside the imagination, and from psychological experiments with thoughts. They should also be distinguished from counterfactual reasoning in general, …". This approach can be used to explore the possible dangers of dual use of technological advances that could occur in the absence of effective ethical scrutiny.

Poetry: Poetry challenges us to think beyond the obvious and reflect on what has been, what is and what might be. Poetry can reboot the way in which social impact education is delivered to technologists. For example, in "Technological Dependency" (Rogerson, 2015) readers are encouraged to reflect on the deeper meaning of each haiku verse with regard to the ethical and social issues surrounding ICT and how we might address such issues.

These examples provide a clear indication of the potential of such novel alternative approaches. The culmination of this discussion lays out a new pathway for ICT ethics education which embraces people of all ages and all walks of life. This is developed by adapting and extending the STEM to STEAM (Science & Technology, interpreted through Engineering & the Arts, all based in Mathematical elements) transition model created by Georgette Yakman.

It is time to start a revolution in your head which will culminate in ethical computing by everyone for everyone. We have to accept and adjust to the fact that we are all technologist to a lesser or greater degree. How we educate our future generations must reflect this change to ensure ICT is societally beneficial. This viewpoint attempts to act as a catalyst for a much-needed paradigm shift.

REFERENCES

Aiken, R.M. (1983). Reflections on teaching computer ethics. *ACM SIGCSE Bulletin*, 15(3), 8-12.

Brown, J.R. & Fehige, Y. (2014). Thought Experiments. In Zalta, E.N. (ed.) (2017) *The Stanford Encyclopedia of Philosophy* (Summer 2017 Edition). available at https://plato.stanford.edu/entries/thought-experiment/

Bynum, T.W. & Rogerson, S. (editors) (2004). *Computer ethics and professional responsibility*. Blackwell Publishing

Graham, J. (2019). WWDC 2019: Meet Apple's youngest app developer, Ayush. *USA Today*, 5 June.

Grosz, B.J., Grant, D.G., Vredenburgh, K., Behrends, J., Hu, L., Simmons, A. & Waldo, J. (2019). Embedded EthiCS: integrating ethics across CS education. *Communications of the ACM*, 62(8), 54-61.

Johnson, D.G. (1985, 1994). *Computer Ethics*. Prentice Hall, Englewood Cliffs, N.J.

Kuhn, T. (1962). *The structure of scientific revolutions*. University of Chicago Press. Chicago.

Martin, C.D., Huff, C., Gotterbarn, D. & Miller, K., 1996. Implementing a tenth strand in the CS curriculum. *Communications of the ACM*, *39*(12), 75-85.

Param, S. (2018). Tanmay Bakshi: The Youngest IBM Watson Developer in the World. *TechGig*, 5 June.

Pecorino, P.A. & Maner, W., 1985. A proposal for a course on computer ethics. *Metaphilosophy*, *16*(4), 327-337.

Rogerson, S. (2015) Technological dependency. *ACM SIGCAS Computers and Society*, 45, 2. 4.

Rogerson, S. (2019). Computing by everyone for everyone. *Journal of Information, Communication and Ethics in Society*, 17(4), 373-374.

WHAT ARE THE INGREDIENTS FOR AN ETHICS EDUCATION FOR COMPUTER SCIENTISTS?

Norberto Patrignani, Iordanis Kavathatzopoulos

Politecnico of Torino (Italy), Uppsala University (Sweden)

norberto.patrignani@polito.it; iordanis@it.uu.se

EXTENDED ABSTRACT

This paper is concentrated on the ethics education for computer scientists. They are the future computer professionals, one of the *core nodes* of the Information and Communication Technologies (ICT) stakeholders' network. The paper is organized in five sections: a general introduction to the subject with a short review of the several approaches to this kind of education, a basic description of the theoretical foundations for ethics education, an overview of the proposed methodology, a short summary of the results from the field experience, and a conclusion.

INTRODUCTION

Teaching ethics to computer scientists is not a new subject, it is probably as old as the computer age in itself: since 1950 Norbert Wiener, was strongly proposing to investigate the social and ethical impacts of computing (Wiener, 1950). After him, in the '80s the field becomes an official discipline (Maner, 1980). Then, in 1985, there is a very important shift from a "*policy vacuum*" approach (Moor, 1985) to a more "*proactive computer ethics*" approach based on the concept of computers as "*socio-technical systems*" (Johnson, 2009). Probably the discipline becomes closer to an "*applied ethics*" with more reference to the day to day work of computer professionals thank to Gotterbarn and Rogerson. Gotterbarn suggests to look at "*... the values that guide the day-to-day activities of computing professionals*" (Gotterbarn, 1991), and Rogerson suggests to concentrate on "*the points where activities and decisions are likely to include a relatively high ethical dimension*" in an ICT project (Rogerson, 2009). Another interesting approach for an "applied ethics to ICT" is proposed by Spiekermann that underlines the need for incorporating values into the computer systems design (Spiekermann, 2016). In 2018 the concept of *Slow Tech*, a "*good, clean, and fair*" ICT is proposed as a "*heuristic compass*" for steering the computer scientists in their design activities (Patrignani and Whitehouse, 2018).

THEORETICAL FOUNDATIONS

The key question is "*can we teach ethics to computer scientists?*". This is a very critical question, since students in computer and engineering faculties are concentrated on a completely different dimension of research: their central question is "*What is true? What is false?*", the well-known Galilean process, or epistemology: observation, theory, prediction, and evidence (when experiments produce positive test results). Nowadays, this historical scientific method is under pressure from the "application" field (the market): the innovation process (creativity, feasibility, prototyping, engineering) is particularly fast in the ICT field (Cohn and Ford, 2003). When introducing the ethics reflection the central question changes completely, it becomes "*What is right? What is wrong?*", with the aim to help students to get

familiar with a process of "*thinking in the right way*", that could enable the elaboration of norms and distinguishes the choice between right and wrong.

The key question about the possibility of teaching ethics has a positive answer: according to Johnson it is possible by exposing students to case studies, giving the opportunity of identifying ethical issues at early stages in the ICT development projects, improving the students' ethical decision-making skills, providing knowledge (codes and standards), skills (ability to identify ethical issues), ability to make moral decisions, and the will to take action (Johnson, 2017). Simply, it is the education and the training for the acquisition of a skill according to classical and modern philosophy as well as to empirical research in moral psychology. Ethics is defined as the ability in making choices and decisions to handle moral issues in a way that satisfies relevant values (see for example the Platonic dialogue *Protagoras*, Platon, 1986; the virtue of *Phronesis*, Aristotle, 1975; and the theory of *Autonomy* of Kant, 2006; Arendt, 2003; Piaget, 1932; Kohlberg, 1985).

PROPOSED METHODOLOGY

The main ingredient for the proposed methodology for teaching ethics to computer scientists is based on a tool called (ICT) *stakeholders' network*. It is a complex net of nodes and arcs (relationships) that try to visualize with a graph the entire collection of actors involved in the ICT domain. The key element of this graph is the set of arcs (with arrows) connecting the nodes. Some of them are one-way, some others are bidirectional. And this is very important since represents often a "power relationship" among two nodes.

While the students draw this network start analysing these relationships and discover potential asymmetry and power relationships. This is the reflection that could trigger the identification of potential conflicts and ethical dilemmas. The ingredients for ethics education for computer scientists can be shortly summarized as: a) describe a scenario where ICT is playing a key role; b) draw the stakeholders' network related to the case under analysis; c) identify ethical dilemmas; d) propose possible and alternative courses of action.

In the class, another "ingredient" for ethics education is to divide the students in three groups for a "role game": the developers' team, the users' team, and the policy makers' team. Then, with the stakeholders' network in their hands the students start a brainstorming session, acting in accordance with their (simulated) roles. Another interesting approach to focus on "value alignment", for example, in the area of artificial intelligence, is the one recently proposed by the Future of Life Institute, where a "visual map" suggests to give evidence also to "*... the ends to which intelligence is aimed and the social/political context, rules, and policies in and through which this all happens*" (FLI, 2019).

At this point the *Slow Tech* "*heuristic compass*" is introduced as another key ingredient for ethics education by putting the three fundamental questions to the students: is the proposed development under analysis "*good*" (human-centred ICT, is it socially desirable)? Is it "*clean*" (environmentally sustainable)? Is it "*fair*" (taking into account the workers conditions along the entire supply-chain?). These three parallel questions are the main novelty of this approach in the domain of "*computer ethics*" education.

SOME RESULTS FROM THE FIELD EXPERIENCE

After some years of application of this methodology in the "*computer ethics*" class (both at Politecnico of Torino, Italy and at Uppsala University, Sweden) it is possible to describe some results based on a sample now involving several hundred students. This experience suggests that the students usually

appreciate this approach since in their final report for the "*computer ethics*" course they demonstrate the acquisition of ethical skills and competencies. For example an important empirical result is provided by the relative high average number of stakeholders in the network: eight. The minimum number of stakeholders is typically three: technology providers, policy makers, users. This demonstrates not only that their horizon for reflection is wider, but also that they are able to identify the main stakeholders, the most critical relationships, and the different interests and conflicts at stake. This increases their "ethical competences" and their capability of proposing different design choices (Kavathatzopoulos et al., 2015).

CONCLUSION

In this paper are presented the main ingredients for an ethics education for computer scientists. They are based on the key capability of drawing the most complete stakeholders' network, to reflect on the different relationships, and to identify ethical issues. Also, the Slow Tech "*heuristic compass*" is proposed as a collection of key questions to be raised in the design process of ICT "*socio-technical systems*" (Johnson, 2009). This can be a contribution to "*Educate for a Positive ICT Future*" as the title of the track of this conference.

KEYWORDS: computer ethics, computer science, Slow Tech.

REFERENCES

Arendt, H. (2003). *Responsibility and judgement*. New York: Schocken.

Aristotle, (1975), *Nicomachean Ethics*. Athens: Papyros.

Cohn, M., Ford, D. (2003). Introducing an agile process to an organization [software development], *IEEE Computer*, 36(6), June 2003.

FLI (2019, December 13). Future of Life Institute, Value Alignment Research Landscape. Retrieved from https://futureoflife.org/valuealignmentmap/

Gotterbarn, D. (1991). A "capstone" course in computer ethics, in Bynum et al. (1991) (eds.) *Teaching Computer Ethics*, Research Center on Computing and Society, S.Conn.State Univ.

Johnson, D.G. (2009). *Computer Ethics*, 4th Edition, Pearson.

Johnson, D.G. (2017). Can Engineering Ethics Be Taught? Engineering Ethics, Vol.47, N.1, Spring 2017, *The Bridge - Linking Engineering and Society*, National Academy of Sciences.

Kant, I. (2006). *Groundwork of the metaphysic of morals*. Stockholm: Daidalos.

Kavathatzopoulos, I. et al. (2015). *Ethics education for engineers: exercises, tools, methods*, in (ed.) Palsson M., 5th Development conference on Swedish Engineer Education (pp.30-31), Uppsala: Uppsala University.

Kohlberg, L. (1985), The just community: approach to moral education in theory and practice, in (eds) Berkowitz M. and Oser F., *Moral education: Theory and application* (27-87), Hillsdale, NJ:Lawrence Erlbaum Associates.

Maner, W. (1980). *Starter Kit in Computer Ethics*. Helvetia Press.

Moor, J. (1985). What is computer ethics? *Metaphilosophy*, October.

Patrignani, N., Whitehouse, D. (2018). *Slow Tech and ICT. A Responsible, Sustainable and Ethical Approach*. Palgrave-MacMillan.

Piaget, J. (1932). *The moral judgement of the child*, London: Routledge and Kegan Paul.

Rogerson, S. (2009). The Ethics of Software Development Project Management, in (eds.) Bynum T.W., Rogerson S., *Computer ethics and professional responsibility*, Blackwell.

Spiekermann, S. (2016). *Ethical IT innovation: A value-based system design approach*, Auerbach Publications.

Wiener, N. (1950). *The Human Use of Human Beings: Cybernetics and Society*. 2nd ed. revised, Houghton Mifflin and Doubleday, Anchor, Boston, MA, 1954.

5. Internet Speech Problems - Responsibility and Governance of Social Media Platforms

Track chair: William M. Fleischman, Villanova University, USA

INTERNET SPEECH PROBLEMS: RESPONSIBILITY AND GOVERNANCE OF SOCIAL MEDIA PLATFORM

Adriana Belgodere Rivera, Fabiana Piñeda Naredo

Interamerican University, School of Law (Puerto Rico)

Adriana.belgodere@lex.inter.edu, Fabiana.pineda@lex.inter.edu

EXTENDED ABSTRACT

Social media has had an extraordinarily positive effect on our lives. The way we share information is instantaneous, how it connects families and friends all over the world, and most importantly, how it has facilitated public debate and strengthened social movements. With this one tool, political leaders can be made or broken. Thanks to its power, thousands of people have connected and unified to create social protests across the globe. It has even been a voice for individuals with restricted expression in countries like Cuba, Venezuela, Iran and China.

To put it into numbers, there are 5.11 billion unique mobile users in the world today, that's 100 million more than in the past year. Internet users increased by 9 percent and come to a total of 4.39 billion. Currently, there are 3.48 billion social media users in 2019.[3] Constantly, social media usage keeps increasing and it probably won't decrease any time soon. Our society has transformed this incredible tool into an inherent part of our day-to-day lives, making it almost impossible to live without. Indisputably, the power of the internet has changed how we look for and gather information. Businesses have been born, networking is easier than ever, and the exchange of ideas is just a click away. With the advanced level of technology that we live in as of now, it would be highly difficult to imagine our day to day life without it.[4]

Just when we think we've got social media pretty figured out, everyday new discoveries arise and affect us directly. The amount of trust we invest in social platforms is increasing, however, our understanding of how they work is not. Privacy issues on the Internet have been around for a while, but with the huge development social media has had in the past five years, we ask ourselves, where will it take us?

In the social media regulation debate, one side has compared these platforms to public utilities. Advocates of this argument believe that since social media and the internet are utilities, they should require increased regulation. These regulations would have the sole purpose to protect identity, information and ideas. They also fear that social media companies could develop into a monopoly and thus, creating regulations to be imposed on them, would be impossible.[5]

On the other hand, social media regulation is quite tricky. These platforms' essence is based on free expression. A citizen's right to speak and express him/herself in guaranteed by the First Amendment of the Constitution of the United States.[6] Federal jurisprudence has given us the following factors to

[3] Simon Kemp, Digital Trends 2019: Every Single Stat You Need To Know About The Internet, March 2019. Retrieved from: https://thenextweb.com/contributors/2019/01/30/digital-trends-2019-every-single-stat-you-need-to-know-about-the-internet/

[4] Madeleine Rosuck, When Lies Go Viral: The First Amendment Implications of Regulating the Spread of Fake News, 21 SMU Sci. & Tech. L. Rev. 319 (2018)

[5] *Id.*

[6] U.S. Const. amend. I

be considered when evaluating a First Amendment case: (1) the proximity and (2) the degree of the language used to the dangers that the speech might elicit.[7]

To say that regulating the internet is complex is an understatement. Courts have expressed that: "The First Amendment's command that government not impede the freedom of speech does not disable the government from taking steps to ensure that private interests not restrict, through physical control of a critical pathway of communication, the free flow of information and ideas."[8]

With tons of information, ideas and thoughts being poured into social media platforms every second, restrictions and censorship are bound to occur. What's interesting is that since these social media companies are private institutions, they're entitled to some degree of freedom to impose regulations. But, to what extent?[9] Private institutions are regulating what government does not. “the First Amendment does not permit government to censor speech to prevent harms to the public apart from known exceptions such as direct incitement to violence.” Thus, these private companies are regulating speech that could harm citizens or social groups in particular for example racial comments, bullying or gender discrimination.[10]

Congress has the power to regulate the Internet. The broad authority under the Commerce Clause gives Congress authority to regulate “channels of interstate commerce,” “instrumentalities of interstate commerce,” and “activities having a substantial relation to interstate commerce.” If the regulated area falls into any of those categories, it is within Congress's power to regulate. Courts have consistently held that the Internet is a channel of interstate commerce. Indeed, Congress has used this power to regulate various online activities. Regulating the removal of harmful content on social media platforms falls squarely within Congress's power under the Commerce Clause. However, the regulation would face significant First Amendment challenges.[11]

Even though citizen’s freedom of speech could be affected, it could also be considered valid to regulate it on social media. Hate speech and fake news are two regulations that government could establish. Fake news means “satirical news, hoaxes, news that’s clumsily framed or outright wrong, propaganda, lies destined for viral clicks and advertising dollars, politically motivated half-truths, and more.”

Furthermore, another obstacle to regulate the Internet is the fact that it's considered a global common. "On the international level, commons are areas that “do not fall within the jurisdiction of any one country”; these areas “are termed international commons or global commons.”[12]

So, should the Government regulate what we say on social media? There’s been a long and exhausting debate about this, and still no action has been done by the government. Social media is currently regulated in a limited way, it is largely immune from government regulation. It is known to be one of the most helpful tools but can also be very negative and destructive.

The real problem we’re facing today is *who* will regulate internet speech. Of course, there are basic and obvious harmful speech that is regulated, like child pornography. However, what about political

[7] Madeleine Rosuck, *supra*.

[8] *Id*.

[9] *Id.*

[10] John Samples, *Cato*, Why the Government Should Not Regulate Content Moderation of Social Media (Apr. 9, 2019), https://www.cato.org/publications/policy-analysis/why-government-should-not-regulate-content-moderation-social-media

[11] Nina I. Brown & Jonathan Peters, Say This, Not That: Government Regulation and Control of Social Media, 68 Syracuse L. Rev. 521, 522 (2018)

[12] Paulina Wu, Impossible to Regulate? Social Media, Terrorists, and the Role for the U.N., 16 Chi. J. Int'l L. 281 (2015)

content? More importantly, what about the increasingly popular concept of *fake news*? Who is entitled to control these kinds of expressions and to what extent is there a valid censorship of it? There are infinite questions that have been brought up when facing social media regulation.[13]

This research explores the complex and delicate issues that exists for social media regulation. Although a solution is highly unlikely given the obstacles of this type of regulation, this investigation aims to compile different aspects of the Internet speech problems and analyze the opposing arguments. Through statistics, jurisprudence and more, the authors attempt to find a clear look at the implications that go into the governance of social media in this overly technological era we're living.

KEYWORDS: Internet speech, responsibility, governance, social media.

REFERENCES

Article 19, Regulating Social Media: We Need a New Model That Protects Free Expression (Apr. 25, 2018). Retrieved from: https://www.article19.org/resources/regulating-social-media-need-new-model-protects-free-expression/

John Samples, Cato, Why the Government Should Not Regulate Content Moderation of Social Media (Apr. 9, 2019). Retrieved from: https://www.cato.org/publications/policy-analysis/why-government- should-not-regulate-content-moderation-social-media

Madeleine Rosuck, When Lies Go Viral: The First Amendment Implications of Regulating the Spread of Fake News, 21 SMU Sci. & Tech. L. Rev. 319 (2018)

Nina I. Brown & Jonathan Peters, Say This, Not That: Government Regulation and Control of Social Media, 68 Syracuse L. Rev. 521, 522 (2018)

Paulina Wu, Impossible to Regulate? Social Media, Terrorists, and the Role for the U.N., 16 Chi. J. Int'l L. 281 (2015)

Simon Kemp, Digital Trends 2019: Every Single Stat You Need To Know About The Internet, March 2019. Retrieved from: https://thenextweb.com/contributors/2019/01/30/digital-trends-2019-every-single-stat-you-need-to-know-about-the-internet/

[13] Natasha Tusikov & Blayne Haggart, *The Conversation*, Stop Outsourcing the Regulation of Hate Speech to Social Media (Mar. 27, 2019), http://theconversation.com/stop-outsourcing-the-regulation-of-hate-speech-to-social-media-114276

PROBLEMS WITH PROBLEMATIC SPEECH ON SOCIAL MEDIA

William Fleischman, Leah Rosenbloom

Villanova University (USA), The Workshop School (USA)

william.fleischman@villanova.edu; leah.rosenbloom@workshopschool.org

EXTENDED ABSTRACT

In this paper, we consider some of the tensions and conflicts between freedom of speech on the Internet, and other public goods and individual rights. The dimensions of the problem include: Threats of physical violence to individuals; threats directed at groups defined by ethnic, national, religious, sexual or gender identity, or political orientation; abusive, harassing, and/or hateful speech; incitement to self-harm; doxing; social exclusion; and dissemination of false information. We make provisional suggestions for discouraging the actions of troll armies and for applying more vigorous measures of transparency in regard to political advertising on social media.

1. INTRODUCTION

The popularization of the Internet promised a radical democratization of communication: Everyone can be a publisher, cost of entry is low, and access is available to anyone connected to the Internet. But early on, prescient individuals understood that "cheap speech," in Eugene Volokh's pungent phrase, carried other implications not all of which are entirely conducive to the dissemination of reliable information or the reasoned discourse of the marketplace of ideas.

As Tim Wu points out in "Is the First Amendment Obsolete?" (Wu, 2017), the assumption that the most serious threats to freedom of speech come principally from governmental actors is no longer entirely valid. Direct censorship can now be supplemented or supplanted by the actions of privately constituted troll armies or bands of individuals and/or robots programmed to drown out disfavoured speech. These means are at the disposal of powerful private interests and loosely organized partisan groups.

In this paper, we consider some of the tensions and conflicts between freedom of speech on the Internet, and other public goods and individual rights. We argue that the widest scope should be afforded individuals' right to free expression, but believe that social media platforms should be held to certain standards of responsibility for preventing or redressing harms resulting from speech on these platforms.

2. THE TROUBLED AND VIOLENT TERRAIN…

"We've got a speech problem on the Internet!" is an observation that covers a lot of ground. The dimensions of the problem include: Threats of physical violence to individuals; threats directed at groups defined by ethnic, national, religious, sexual or gender identity, or political orientation; abusive, harassing, and/or hateful speech; incitement to self-harm; doxing; social exclusion; and dissemination of false information.

2.1. Troll Armies

Gamergate (Wikipedia, 2019) is a well-known example of an online hate mob. Lately, troll armies have figured prominently in polarized political discourse. Tim Wu (2017) cites two examples: David French, a writer associated with the conservative *National Review*, and Rosa Brooks, a professor of law at Georgetown University, both targets of online mobs for criticism of the current U.S. president.

The rhetoric was murderous and hateful in both instances – Nazi imagery and the face of his daughter in the gas chamber in the case of French (French, 2016), and extremely violent misogynistic language directed at Brooks. (Brooks, 2017)

2.2. Reverse Censorship and Flooding

Another technique used by governments to marginalize dissident speech involves mobilizing a volume of opposing information to drown out inconvenient speech or distort the informational environment to render the speech dubious and unimportant. An important variant of reverse censorship, used in political advertisement, floods public discourse with patently false information or “fake news.” It is widely understood that in 2016 targeted political advertisements disseminating false information were instrumental in both the U.K. Brexit referendum (Cadwalladr, 2019) and the U.S. Presidential election. (Lapowsky, 2018)

3…. NONETHELESS THERE ARE GOOD REASONS TO PROTECT EVEN EXTREME SPEECH…

In spite of these examples, there are important reasons to favor protecting even extreme speech on the web. Speech on the internet is an important cornerstone of sweeping social change happening all over the world. The price people pay for speaking publicly (where speech is tied to the person's real name and location) is often high. Consequences range from execution and imprisonment to stalking, harassment, and surveillance. (Patterson, 2017). Tools that allow for anonymity, or do not require users to identify themselves, allow people to fight institutionalized oppression without becoming targets for retaliation.

Governments don't like anonymous online platforms because they make it impossible to identify people who are saying what are considered bad things. It is important to remember that characterization of "bad things" is a prerogative of those in power. In authoritarian regimes, "bad things" can be anything against the government or current power structure. Of course, "bad things" can also be racial slurs, homophobic remarks, troll brigading feminists, etc. However, history has shown that governments exercise their power selectively against those who threaten their authority, not to protect the vulnerable and marginalized. Anonymous online platforms also come under attack from private groups with particular agendas, often to curtail the speech of disfavored marginalized groups who most need these rights. The ACLU, EFF, and other civil and human rights organizations recognize the importance of preserving free speech online and have filed amicus curiae briefs in many cases including hateful speech and discussions of terrorism.

Any actual terrorist plan would be incitement, and law enforcement would be responsible for tracking them down the source, which Facebook can help them do by complying with a warrant. The platform itself, and any speech that is not incitement, are protected under the First Amendment. Civil rights organizations understand the absolute importance of protecting these rights because of their use in challenging oppressive institutions and facilitating positive change.

4.… BUT NOT NECESSARILY WITHOUT ANY LIMITS

First Amendment protections are intended to restrain the power of government to interfere with the freedom of expression of individuals. However, speech online occurs through an internet service provider or a social media platform which inherit First Amendment protections. As private enterprises they can and do set rules that should apply to their users. Often these rules are vague and inconsistently applied. In particular, when speech policies come into conflict with the profit motive of the platform, these policies frequently evaporate. The result is often flagrantly abusive, obscene, threatening or deceptive speech.

We believe that social media platforms should hold themselves to standards and policies that are consistently applied and promote more responsible speech. Reddit provides an example that shows this is possible. (Marantz, 2016)

4.1. The Case of Troll Armies

We propose that in cases like those cited, where troll armies coordinate hateful speech and threats against an individual, the social media platform that facilitates such an attack take action against members of the mob. The principle here is that those claiming the right to speak violently and abusively under the doctrine of freedom of expression are, in fact, acting to attempt to silence another individual, and therefore, perversely curtailing the very same right of that individual.

These situations should be relatively easy to document, once the attacked individual registers a complaint with the social media platform, since they consist of N --> 1 more or less synchronized messages (multiple sources, one target). We recognize the limitations of algorithmic detection or wholesale human moderation of every instance of hateful and threatening speech. By contrast, the cases to which we refer are not so frequent that human inspection would be impossibly difficult. Naturally, this requires nuanced consideration of the multiple messages, some of which may be reasoned arguments expressed in strong language and should be differentiated from those that simply spew hate in language and (photoshopped) images.

4.2. The Case of False Political Advertising

Our other proposal has to do with political speech - specifically political ads that circulate false or discredited information. Facebook is currently involved in such a dispute. We don't want to say that these things should be outlawed - there's plenty of history, going back to the election of 1800 in our country, of scurrilous political speech (McCullough, 2001). But it is troubling that ads on Facebook and other platforms appear and then disappear without any trace so there is no possibility of auditing them or providing public scrutiny. They are particularly pernicious because they are targeted to people identified as susceptible through analysis of their Facebook profiles.

Carol Cadwalladr (2019) has documented how this occurred in the Brexit referendum and how Facebook has stonewalled any serious attempt to investigate the sources of funding and means of targeting these false and vanishing ads. Facebook executives have been notably oblivious and evasive about such advertising. (Lee, 2018) Our proposal is to force the social media platform to keep publicly accessible, auditable records of political ads so that they can be scrutinized and rebutted in the same way that's possible with ads on other media.

KEYWORDS: freedom of speech, violent and abusive speech, internet, social media.

REFERENCES

Brooks, R. (2017). And then the Breitbart lynch mob came for me, *Foreign Policy*. Retrieved from https://foreignpolicy.com/2017/02/06/and-then-the-breitbart-lynch-mob-came-for-me-bannon-trolls-trump/

Cadwalladr, C. (2019). Facebook's role in Brexit and the threat to democracy. Retrieved from https://www.ted.com/talks/carole_cadwalladr_facebook_s_role_in_brexit_and_the_threat_to_democracy?language=en

French, D. (2016). The price I've paid for opposing Trump, National Review. Retrieved from https://www.nationalreview.com/2016/10/donald-trump-alt-right-internet-abuse-never-trump-movement/

Lapowsky, I. (2018) Mark Zuckerberg speaks out on Cambridge Analytica Scandal, Wired. Retrieved from https://www.wired.com/story/mark-zuckerberg-statement-cambridge-analytica/

Lee, D. (2018). Mark Zuckerberg, missing in inaction, *BBC News*. Retrieved from https://www.bbc.com/news/technology-46231284

Marantz, A. (2018). Reddit and the struggle to detoxify the internet, *New Yorker*. Retrieved from https://www.newyorker.com/magazine/2018/03/19/reddit-and-the-struggle-to-detoxify-the-internet

McCullough, D. (2001, at 545ff). *John Adams*, Simon & Schuster, New York

Patterson, B. (2017) Police Spied on New York Black Lives Matter Group, Internal Police Documents Show, retrieved from https://www.motherjones.com/crime-justice/2017/10/police-spied-on-new-york-black-lives-matter-group-internal-police-documents-show/

Wikipedia (2019). Gamergate controversy. Retrieved from https://en.wikipedia.org/wiki/Gamergate_controversy

Wu, T. (2017) Is the First Amendment obsolete? Knight Foundation First Amendment Institute. Retrieved from https://knightcolumbia.org/content/tim-wu-first-amendment-obsolete

SRI LANKAN POLITICS AND SOCIAL MEDIA PARTICIPATION A CASE STUDY OF THE PRESIDENTIAL ELECTION 2019

Chintha Kaluarachchi, Ruwan Nagahawatta, Matthew Warren

Deakin University (Australia)

c.kaluarachchi@deakin.edu.au; rnagahawatta@deakin.edu.au; matthew.warren@deakin.edu.au

EXTENDED ABSTRACT

The social media influences traditional media and has become an alternative to traditional media (Piechota, 2011). Negussie & Ketema (2014) argued that Facebook (FB) is a media for freedom of dialogue and consent to people from different ethnicities, religions, and backgrounds to directly share information without any restrictions. Also, FB election campaign is recognized as a facilitative mode to access political information in several ways. Through FB group activists and ordinary citizens could voice their opposition to the government when denied democracy and suppress their views and voices (Hanson et al., 2010). For example, The USA presidential campaign in 2008 was the first to use the world of YouTube, My Space, FB, and political blogging Internet based for such purposes. By 2010, 22% of Internet users have been using social media network for political activity (Bekafigo et al., 2013). Further, many studies argued that social media could influence people by changing their perceptions, attitudes and promote people to think differently. From a political party perspective, social media provides a cost-effective medium to reach-out to large number of users (voters), it provides a rich two way engagement with users (voters) and by its nature creates interaction. Social media also offers a business benefits for political parties, by using social media they could engage with many more users (voters) rather than traditional media, so it means their investment in social media could give greater returns (Warren, 2018).

With the announcement of Sri Lankan presidential election 2019, there was increase in social media posts related to political campaigns. The target of these political campaigns is social fragmentation and reduce voter's loyalty towards democratic political parties and candidates. This tread is rising rapidly aiming at variety of targets. These campaigns target personal lifestyle values to engage with variety of cases such as human rights, racist violence, economic justice, environmental protection (Bennett, 2012). Another key aspect of the use of social media by political parties is that it allows them to influence voters and the way that could vote, this is also known as information operations. Information operations also known as influence operations, includes the collection of tactical information about an adversary as well as the dissemination of propaganda .e.g. fake news in pursuit of a competitive advantage over an opponent. (Waltzman, 2017). In a Sri Lankan context, the influence of social media on Sri Lankan politics has brought new dangers. According to the Prime minister Ranil Wickramasinghe "Sri Lanka continues to face 'New dangers' posed by hate speech, fake news" (Maldives Independent, 2019).

This study collected data from FB using CrowdTangle related to the Sri Lankan presidential election 2019. The research question related to "Does the Facebook influence the Sri Lanka Presidential election 2019 and to examine what shape the influence takes, whether the influence is guided, or evolving freely related to the "Responsibility and Governance" of social media platforms.

The study focussed on collecting data from the four most popular open FB political groups namely "Ape Rata", "JVP Balakotuwa", "Ekayayata kola patata", "Sri AV TV Network". These has been selected

to perform the analysis and 175 posts had been selected using purposive sampling technique from October 20-27 in 2019, to analyse using the key themes by the researchers. Table1 presents the key themes of the posts and viewers' interactions during October 2019 related to the Sri Lanka presidential election.

Table 1. Type of posts used by the selected FB public groups and followers interactions related to the 2019 presidential election

Key Themes		Total Post (Frequency)	Percentage (%)	Number of reactions	Number of Shares	Number of Comments	Total Interactions (Count)*	Total Interactions (%)*	Average Followers per Theme**
Promotion of the candidates		45	26%	6061	4059	817	10937	19%	324814
Distribution of Fake News (false and demeaning information)		16	9%	1041	906	129	2076	4%	144933
Social fragmentation and reduce voter's loyalty		29	17%	5135	7409	747	13291	23%	258282
Social awareness		33	19%	4506	3114	259	7879	13%	426079
Racist violence		10	6%	2983	2233	365	5581	9%	362531
Economic justice		19	11%	9246	3493	644	13383	23%	659540
Environmental protection		6	3%	339	306	40	685	1%	215091
Social Security and Human Rights		17	10%	3177	1376	427	4980	8%	312524
Total		175	100%	32488	22896	3428	58812	100%	

*Total Interactions: The sum of Reactions, Shares and Comments related to the each theme.

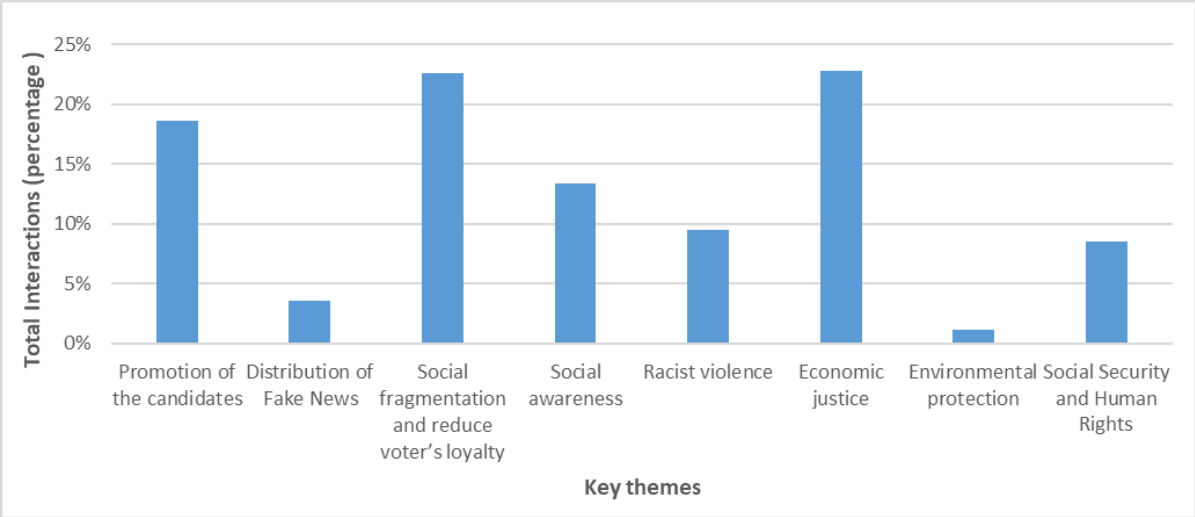
**Average Followers: The sum of Page Likes, Instagram followers, Twitter followers, or sub edit subscribers for all of the matching results.

Table 1 shows that the most of the posts shared related to the following themes: Promotion of the candidate (26%), Social fragmentation and reduce voter's loyalty (17%) and Social Awareness (19%) while Environmental protection (3%) theme is the least post shared theme.

Then we looked at the data to determine the social interactions related to the themes of the posts.

Figure 1 explore the user generated interactions by theme. Total interactions were vary according to the each theme the majority of followers interacted to the 'Social fragmentation & reduce voter's loyalty' and 'Economic justice' themes with each 23% interactions. Nevertheless Environmental protection theme has got least number of social interaction (1%) as shown.

Figure 1. Distribution of social interactions by theme related to the 2019 presidential election



*Total Interactions: The sum of Reactions, Shares and Comments related to the each theme.

The analysis revealed what the major themes were in the Sri Lankan presidential election 2019. The analysis identified that the major themes Promotion of the candidates (26%); Distribution of Fake News (9%); Social fragmentation and reduce voter's loyalty (17%); Social awareness (19%); Racist violence (6%); Economic justice (11%); Environmental protection (3%) and Social Security and Human Rights (10%). What was of interest was that number one theme was the 'Promotion of particular candidates' many of the posts were promoting the main candidates, Gotabaya Rajapaksha and Sajith Premadasa were popular. The highest-profile candidates were Gotabaya Rajapaksha and Sajith Premadasa and they had used a sophisticated and heavy social media campaigns (EU election observation mission; 2019) and that seems to influence social media posts related to Promotion of the candidates theme. Another interesting outcome was that 'Distribution of Fake News' that only reflected 9% of the themes in the posts, the authors had expected this figure to much higher.

Another important outcome of the analysis was the disclosure of social interactions related to the themes in the posts. Mostly interacted themes by the followers were Social fragmentation and reduce voter's loyalty (23%) closely followed by Economic justice (23%). According to the EU election observation mission in Sri Lanka Presidential election (2019) most of the candidates used cross-platform electioneering tactics online, with official party pages adjoining third-party sites that frequently served to discredit the rival. This may lead to influence posts and social interactions related to the "Social fragmentation and reduce voter's loyalty theme mostly". This may possibly a guided influence by third party sites operated by the political parties. Also Sri Lankan nation has struggled by the challenges they posed due to sluggish economy, increasing political polarisation and security challenges. Because of that National security and Economic Justice were a prominent themes in the election campaigns. From our analysis also we can confirm that the economic justice theme was a prominent theme in terms of the posts and social interactions.

The paper has shown that social medial has the ability to generate discussion and debate, we showed that the most popular FB posting was to promote particular presidential candidates, and it may possibly a guided influence. The most interacted themes were "Social fragmentation and reduce voter's loyalty theme" and "economic justice". When we analyse the user interactions we can see both guided and freely evolving interactions related to the Sri Lankan presidential election 2019. Also issues

such as Fake News were not a major issue. The authors have shown that Facebook did have guided and freely evolving influences on the Sri Lankan presidential election of 2019.

In conclusion, few studies have been written about prior Sri Lankan presidential elections and parliamentary elections and the impact of social media upon those elections.

KEYWORDS: Social Media, Social Interactions, Presidential Election, Politics, Sri Lanka.

REFERENCES

Bekafigo, M.A., Cohen D.T., Gainous J., & Wagner K.M. (2013). State Parties 2.0: Facebook, Campaigns, and Elections. The International Journal of Technology, Knowledge, and Society, 9(1), 99-112.

Bennett, W. L. (2012). The Personalization of Politics: Political Identity, Social Media, and Changing Patterns of Participation. The ANNALS of the American Academy of Political and Social Science, 644(1), 20–39. https://doi.org/10.1177/0002716212451428

Goodman, J., Wennerstrom, A., & Springgate, B. F. (2011). Participatory and social media to engage youth: from the Obama campaign to public health practice. Ethnicity & disease, 21(3 Suppl 1), S1–99.

Hanson, G., Haridakis P.M., Cunningham A.W., Sharma R., & Ponder J.D. (2010). The 2008 Presidential Campaign: Political Cynicism in the Age of Facebook, MySpace, and YouTube. Journal of mass communication and society, 13(5), 584-607.

Meti, V., Khandoba P.K., & Guru M.C. (2015). Social Media for Political Mobilization in India: A Study. J Mass Communicat Journalism 5:275. doi:10.4172/2165-7912.1000275

Maldives Independent, (2019). Sri Lanka continues to face 'new dangers' posed by hate speech, fake news – PM, Retrieved October 4, 2019, from http://www.adaderana.lk/news/57499/sri-lanka-continues-to-face-new-dangers-posed-by-hate-speech-fake-news-pm

Negussie, N., & Ketema G. (2014). Relationship between FB Practice and Academic Performance of University Students, Asian Journal of Humanities and Social Sciences (AJHSS), 2(2), 31-37.

Piechota, G. (2011). Application of social media in political communication of local leaders in election processes (on the example of Facebook's use by mayors of voivodship cities in Poland in the 2010 election campaign).

Waltzman, R. (2017). The Weaponization of Information, RAND, URL: https://www.rand.org/content/dam/rand/pubs/testimonies/CT400/CT473/RAND_CT473.pdf, accessed 10/11/18.

Warren, M. (2018). Political Cyber Operations, Conference Proceedings of Australian Cyber Warfare Conference 2018, ISBN 978-0-6484570-0-8.

6. Justice, Malware, and Facial Recognition

Track chair: Wade Robison, Rochester Institute of Technology, USA

A FLOATING CONJECTURE: IDENTIFICATION THROUGH FACIAL RECOGNITION

Wade Robison

Rochester Institute of Technology

wlrgsh@rit.edu.

EXTENDED ABSTRACT

Facial recognition has been widely used already to identify criminal suspects from video recordings and to provide evidence in criminal trials. Its use in criminal trials is part of a larger pattern of using questionable rules of skill that tell us how to identify a suspect based on fingerprints, bite marks, and other supposedly identifying information.

A rule of skill tells us how to achieve a particular end: to bake a cake, do such-and-such; to buttress a girder, do so-and-so. They are the tools of the trade, so to speak, for any profession, and among them are rules that prescribe procedures to follow — in writing a valid will or ensuring a fair trial, minimizing the risks of infection, and on and on.

We are all familiar with rules. We learn them early on as we learn to count or correct our pronunciation so we can be understood. We know as well what happens when we fail to follow the relevant rules. We open ourselves to criticism and to failure.

Rules set norms, and they are no different in that way than, say, the rules of logic. But some rules are floating conjectures, without the sorts of evidential backing needed to make them reliable. They direct the activities of those within a profession and so have effects in the world. But not having been formulated through rigorous experimentation and testing, they may lead professionals astray.

Anyone who reads mysteries or watches crime shows knows that central to a crime's solution is what can be found at the crime scene — 'DNA, hair, latent fingerprints, firearms and spent ammunition, toolmarks and bitemarks, shoeprints and tire tracks, and handwriting.' Detectives hunt for samples at the scene that can then be compared to samples from a suspect, and they remind everyone not to touch anything at the scene so that when they dust for fingerprints, for instance, their findings will not have been contaminated. They are hunting for fingerprints or hair or something else left by whoever committed the scene crime.

Experts compare the features of what is found at the crime scene with the features of the relevant sample from a suspect, and if there is a match, they have significant evidence that the suspect is the criminal. How significant? That depends upon the validity and reliability of the methods of comparison.

As it turns out, we have a way of answering that question. Although using DNA to tie a particular suspect to a crime scene is not without its problems, it has become the gold standard, and we can assess the validity and reliability of comparing fingerprints and hair, for example, by determining if using DNA gives us the results we got in previous cases comparing other features from the crime scene.

As it turns out, feature comparisons are not very reliable at all. As a 2016 Report to the President on forensic science stated,

> Reviews by the National Institute of Justice and others have found that DNA testing during the course of investigations has cleared tens of thousands of suspects and that DNA-based re-examination of past cases has led so far to the exonerations of 342 defendants.

The Innocence Project exonerated more than 350 individuals, and in 45% of the cases, those individuals were convicted because of a failure of feature comparisons combined with misleading testimony from experts who ensured juries and judges that they were sure within a 'reasonable degree of scientific certainty.' That is a phrase that gives great weight to the evidence but has no scientific validity.

The 2016 Report quotes a judge about testimony from an expert that 'markings on certain bullets were unique to a gun recovered from a defendant's apartment':

> As matters currently stand, a certainty statement regarding toolmark pattern matching has the same probative value as the vision of a psychic: it reflects nothing more than the individual's foundationless faith in what he believes to be true. This is not evidence on which we can in good conscience rely, particularly in criminal cases, where we demand proof—real proof—beyond a reasonable doubt, precisely because the stakes are so high.

The Report adds,

> In science, assertions that a metrological method is more accurate than has been empirically demonstrated are rightly regarded as mere speculation, not valid conclusions that merit credence.

The need for evidence and testimony based on evidence is nicely put by U.S. District Judge John Potter, in 'an early case on the use of DNA analysis,' *U.S. v. Yee (1991):*

> Without the probability assessment, the jury does not know what to make of the fact that the patterns match: the jury does not know whether the patterns are as common as pictures with two eyes, or as unique as the *Mona Lisa*.

So we have floaters in forensic science. Because of them, some individuals were executed. We cannot interview the individuals who were wrongly executed, but we can get a sense of how much more damage the use of these floaters has caused by looking at several cases, including the Brandon Mayfield case, which has become a classic example of how misidentification of a sample can mislead investigators, taking them off the scent of the perpetrator, as it were, onto the scent of an innocent person.

In March 11, 2004, ten bombs killed 192 passengers on trains in Madrid and injured more than 1400, according to initial reports. The Spanish authorities found a fingerprint on a bag of detonators and forwarded it to the FBI to see if it could find a match in its database. The FBI's Integrated Automated Fingerprint Identification System (IAFIS) 'generated a list of 20 candidate prints.' None was a perfect match, but IAFIS also lists close matches, and one belonged to Brandon Mayfield, a lawyer in Oregon. The FBI 'immediately opened an intensive investigation of Mayfield, including 24-hour surveillance…and physical searches' of his law office and residence. When news somehow broke that an American was a suspect in the bombing, the FBI detained Mayfield on May 6th because they were 'absolutely confident' that Mayfield's fingerprint was on the detonator bags. They kept him in solitary confinement 'for up to 22 hours per day.'

The fingerprint from Spain was examined by a fingerprint specialist in the FBI who verified it as belonging to Mayfield. That judgment was confirmed by a second FBI fingerprint specialist and by the

fingerprint unit chief, all of whom agreed it was Mayfield's. That decision was confirmed by a court-appointed specialist. Four fingerprint experts fingered Mayfield, as it were.

The Spanish authorities identified the person whose fingerprint was on the bag of detonators, and it was not Mayfield. As it turned out, further analysis of the fingerprints showed that Mayfield's was not identical to the one found in Spain, but what is of importance here is that specialists in fingerprint identification judged that it was and that they had absolute confidence in their judgment. The Mayfield case is a dramatic example of why such judgments cannot be relied upon and should not be relied on, especially in criminal cases where the stakes are high. We must have proof beyond a reasonable doubt, and the Mayfield case puts in doubt reliance on fingerprints comparisons.

Another example of a floater in forensic science concerns bite marks. Keith Harward 'narrowly escaped the death penalty,' but spent 33 years in prison after being convicted of rape and murder on the basis of six forensic dentists testifying that the bite marks on the rape victim's legs were his. DNA evidence showed that he was innocent and that a fellow sailor, Jerry Crotty, was responsible. Harward is one of at least 25 individuals 'to have been wrongfully convicted or indicted based at least in part on bite mark evidence.' He is now free, but he says to those who tell him he is a free man, 'I will never be free of this…I spent more than half my life in prison behind the opinions and expert egos of two odontologists.'

The 2016 Report to the President pointed out that a '2010 study of experimentally created bitemarks…found that skin deformation distorts bitemarks so substantially and so variably that current procedures for comparing bitemarks are unable to reliably exclude or include a suspect as a potential biter.' In fact, evidence 'showed a disturbing lack of consistency in the way that forensic odontologists go about analyzing bitemarks, including even on deciding whether there was sufficient evidence to determine whether a photographed bitemark was a human bitemark.' That bite mark evidence still finds its way into court cases is a sad commentary on the failure of our judicial system to come to grips with such forensic floaters.

Introducing facial recognition is an improvement over bitemarks, but not over fingerprints. Algorithms are used to fill in the gaps in prints, and their use misidentified Mayfield. Facial recognition also depends upon algorithms, and the same problem affects it and will lead to misidentification—and convictions of innocent individuals.

KEY WORDS: Facial recognition, rules of skill, identification of suspects, misidentification.

REFERENCES

Harward, Keigth Allen. The Innocence Project, available online at https://www.innocenceproject.org/cases/keith-allen-harward/; accessed April 4, 2019.

Office of Inspector General, Oversight and Review Division, A Review of the FBI's Handling of the Brandon Mayfield Case, January 2006, p. 1; available online at https://oig.justice.gov/special/s0601/exec.pdf; accessed April 4, 2019.

Oliver, John. 'Forensic Science: Last Week Tonight,' October 1, 2017. From 13:21 to 14:40 in the video; available online at https://www.youtube.com/watch?v=ScmJvmzDcG0; accessed April 4, 2019.

Otterman, Sharon. 'She Was Fired After Raising Questions About a DNA Test. Now She's Getting $1 Million,' *The New York Times*, April 23, 2019.

Report to the President, Forensic Science in Criminal Courts: Ensuring Scientific Validity in Feature-Comparison Methods, Executive Office of the President, President's Council of Advisors on Science and Technology, September 2016, p. 1; available online at https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_forensic_science_report_final.pdf; accessed April 1, 2019.

Sciolino, Elaine. 'BOMBINGS IN MADRID: THE ATTACK: 10 Bombs Shatter Trains in Madrid, Killing 192,' *The New York Times*, March 12, 2004.

The Innocence Project, Misapplication of Forensic Science; available online at https://www.innocenceproject.org/causes/misapplication-forensic-science/; accessed April 1, 2019.

Wang, Amy B. 'Video shows police trying to explain why they pulled over a Florida state attorney,' *The Washington Post*, July 13, 2017.

JUDICIAL PROHIBITION OF THE COLLECTION AND PROCESSING OF IMAGES AND BIOMETRIC DATA FOR THE DEFINITION OF ADVERTISING IN PUBLIC TRANSPORT

George Niaradi, Nilson Nascimento

Instituto Damásio de Direito (Brazil), CEAP – Centro Educacional Assistencial Profissionalizante (Brazil)

ganiaradi1@uol.com.br, nlsn.neves@gmail.com

EXTENDED ABSTRACT

The performance of the Judiciary is enshrined in the Brazilian legal system when it is determined that it is responsible for the inspection of the other Public Branchs, when provoked, on the protection of Fundamental Rights. Thus, it must ensure that its decisions have content that respects them, being directly linked, in the content, and in its way of acting (MARTINS ALVES NUNES JÚNIOR, 2010).

Fundamental rights have binding force over all constituted Powers. The Legislative Branch is not allowed to revoke later or limit fundamental rights; such situations would be the same as hurting the essential core of the Constitution. The Executive Branch has a duty to comply with fundamental rights in its actions. The Judiciary Branch is responsible for inspecting the other Branches in the application of fundamental rights, and for observing the guaranteeing constitutional rules in its decisions and in the conduct of proceedings.

One type of judicial proceeding to fulfill this purpose is the Public Civil Action; provided for in the Brazilian Federal Constitution, it serves for the protection of public and social assets, the environment and other diffuse and collective interests. This is an appropriate procedural instrument for the exercise of popular control over acts of public authorities (COSTA, 2011). It is, therefore, a relevant technique for the defense of individual and collective rights being used in various fields of activity (WALD, 2003).

The Public Prosecutor's Office, the Federal Union, the Member States, municipalities, local authorities, public undertakings, foundations, mixed-economy companies and also associations that have been established for at least one year and have among their institutional objectives the protection of fundamental rights shall be entitled to bring a public civil action.

In the Public Civil Action, Case No. 1090663-42.2018.8.26.0100, in the process of the 37th Civil Court of the Court of Justice of São Paulo - Brazil, the Consumer Defense Institute (IDEC), a Brazilian non-governmental organization without governmental or business ties, founded in 1987 that aims to promote education, the defense of consumer rights and ethics in consumer relations, required the prohibition of the collection and treatment of images and biometric data taken, without prior consent, from users of subway lines in the City of São Paulo - Brazil.

The subway concessionaire's objective in capturing the images was to evaluate the physiognomic expressions of the subway line users, along with the digital panel system, to adapt the commercial advertising of the products. Therefore, these were mechanisms of facial recognition, without the prior consent of individuals, receiving the revenue from this advertising activity performed.

Thus, this is a compulsory opinion survey for advertising purposes, generating a doubt about the legality of the collection and processing of images and biometric data of subway users. The situation

described is aggravated by the knowledge that the traffic of children and adolescents in subway stations and lines is commonplace.

In this judicial case, the judge decided to maintain the limitation to the capture of images, sounds and any other data through cameras or other devices, regardless of being involved in the advertising matter, setting a daily fine to the subway concessionaire in case it does not proceed to turn off the devices.

In technological terms, facial recognition softwares has evolved, thanks to the advent of artificial intelligence, to the point of detecting people's emotions such as happiness, sadness, angry, surprise, disgusted, calm, confusion and fear (FAZZINI, 2018). They can also identify the gender of each person and estimate an age range with some precision. With this information advertising companies can direct advertisements to passers-by more assertively, but in return end up collecting biometric data, often shared with the company that provided the software. In possession of this data, an unethical company could track some of its user's habits and resell this information (SMITH, 2018). All this without counting the inconvenience caused by an eventual data leak, a problem that has already reached even the companies that are considered giants of technology.

The clear intention taken by the court was to ensure the preservation of the presupposed ethical image capture of people translated into its transparent use. In the field of fundamental rights, the protection of the individual's privacy can never be left out of the debates; the extent of this protection is unrestricted to the human condition, in other words, it is part of the very situation of the person inserted in society, regardless of the role he or she is playing.

Because they are artificial intelligence technologies, based on a statistical model, facial recognition can be used in a perverse way, away from preservationist ideals of the human condition; they are situations such as those described in the aforementioned judicial case, which offend the privacy and freedom of choice of the individual, inferring wishes and submitting, without prior consent, choices to subway users in the City of São Paulo - Brazil, who may even be children and adolescents.

Despite the diversity of purposes that facial recognition can assume, the ethical dilemma brought in this case is related to the improper and intrusive use of images, performing the interpretation of physiognomies to stimulate the interest of the individual in acquiring a product in advertising exhibition within the public space of the subway. It is an undue control, generating a kind of selection (FOUCAULT, 2001) in which one passes from a configuration of sovereign power, incident on the possibility or not of life of vassals, to a disciplinary logic, in which the ways of life of the vassals' bodies will be the object of intervention.

In this sense, (FONSECA, 2008) life is not being preserved in society. The Universal Declaration of Human Rights, created in the 1940s, is a series of fundamental principles for the regulation of relations, based on a general idea of justice, translated into common sense and judgment about common sense. The decision to grant the injunction of the cited Public Civil Action is incorporated of innate values to the human condition, ruling out the intrusion in privacy and the imposition of choices to individuals using the subway in the City of São Paulo - Brazil.

KEYWORDS: Justice, Facial Recognition, Advertising, Freedom of Choice.

REFERENCES

Fazzini, K. (2018, December 6). Amazon's facial recognition service is being used to scan mugshots, but it's also used to track innocuous things like soccer balls. *CNBC*. Retrieved from http://www.cnbc.com/2018/12/06/how-amazon-rekognition-works-and-what-its-used-for.html

Fonseca, Tania Mara Galli et al. Microfascismos em nós: práticas de exceção no contemporâneo. Psicol. clin., Rio de Janeiro, v. 20, n. 2, p. 31-45, 2008. Retrieved from http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-56652008000200003&lng=en&nrm=iso

Foucault, Michel. Os anormais: curso no Collège de France. São Paulo: Martins Fontes, 2001.

Kalleo Castilho Costa (2011). Ação Popular e Ação Civil Pública. Retrieved from https://ambitojuridico.com.br/cadernos/direito-constitucional/acao-popular-e-acao-civil-publica/

Martins Alves Nunes Júnior, Flávio (2010). O Poder Judiciário e a sua vinculação aos direitos fundamentais. Retrieved from https://ambitojuridico.com.br/cadernos/direito-constitucional/o-poder-judiciario-e-a-sua-vinculacao-aos-direitos-fundamentais/

Smith, B. (2018, December 6). Facial recognition: It's time for action. *Microsoft On the Issues*. Retrieved from https://blogs.microsoft.com/on-the-issues/2018/12/06/facial-recognition-its-time-for-action/

Wald, Arnoldo. Aspectos Polêmicos da Ação Civil Pública. São Paulo: Saraiva. 2003.

THE USE OF FACIAL RECOGNITION IN CHINA'S SOCIAL CREDIT SYSTEM: AN ANTICIPATORY ETHICAL ANALYSIS

Ion A. Iftimie, Richard L. Wilson, Michele C.A. Iftimie, Francis Lukban

NATO Defense College (Italy), University of Baltimore (U.S.A.), European Union Research Center (U.S.A.), Mount St. Mary's University (U.S.A.)

iftimie@gwu.edu, wilson@towson.edu, falconqu@gmail.com, fjlukban@email.msmary.edu

EXTENDED ABSTRACT:

In 2014, China has implemented the Social Credit System (SCS), which aims to assess and predict (Liang et al. 2018) "the trustworthiness of Chinese citizens [to comply] with legal rules, moral norms, and professional and ethical standards" (Chen and Cheung 2017). China has also partnered with the private sector (Creemers 2018) to integrate facial recognition and machine learning technologies (Wright 2018) with the SCS, which "is currently on track for full deployment on 1.4 billion citizens by 2020" (Liang et al. 2018). This paper reviews existing legislation and regulations in China governing biometric data collection, aggregation (storage) and analytics, and conducts an anticipatory ethical analysis of current facial biometrics technologies adopted by the SCS. This study is needed to guarantee the ethical development and implementation of facial biometrics technologies by nation states—and particularly, by law enforcement, investigative, and national security agencies—in the absence of clear regulations governing emerging biometrics and machine learning technologies.

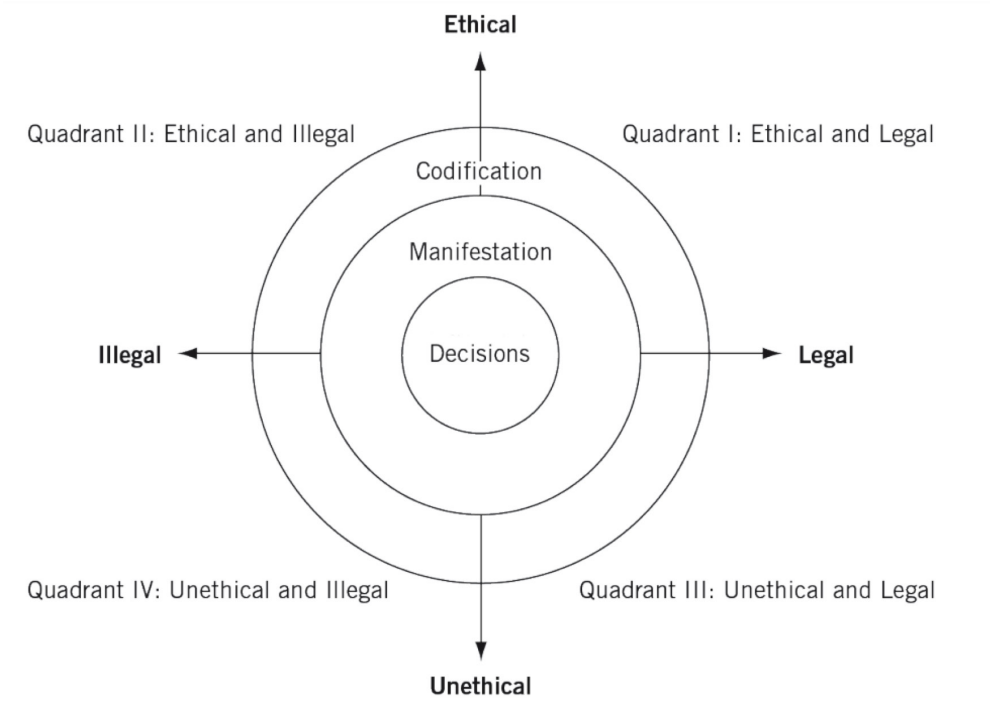
Given the fast pace of technological advancements in the field of biometric identification, laws sometimes fail to clearly delineate the confines of use by both the private sector and government institutions worldwide. We pose that ethical considerations must thus supplement the legal ones when it comes to the integration of biometric identification technologies by both private and government institutions. The dilemma of legal versus ethical behavior has often been discussed in the field of ethics, dealing with corporate decisions being separated into i) ethical and legal; ii) ethical and illegal; iii) unethical and legal; and iv) unethical and illegal (see Figure 1). In evaluating the integration of biometric identity technologies within SCS, we recognize that Eastern perceptions of moral norms (that is, of what is "right" and "wrong") differ from Western ones. This of ethical pluralism is important to consider, particularly when assessing stakeholder perceptions on the right balance between privacy and security. For example, the SCS has a high degree of approval among stakeholders, who largely interpret the program "through frames of benefit-generation and promoting honest dealings in society and the economy instead of privacy-violation" (Kostka 2019). We pose that ethical dilemmas dealing with the integration of biometric identity technologies must be analyzed primarily from the perspective of the stakeholders inhabiting the geographic area where the technology is used (and who are, thus, governed by the same laws).

In order to better understand the totality of ethical dilemmas behind the integration of biometric identity technologies within SCS, we need to have a clear understanding of [anticipatory] ethics. Ethics in a basic definition relates to agents who perform actions and how these actions affect other agents. Dwight Furrow identifies the focus of ethical analysis as involving a series of factors, where ethics is related to evaluating actions and actions are performed by those capable of being moral agents. Per Furrow, "when we evaluate an action, we can focus on various dimensions of the action. We can evaluate the person who is acting, the intention or motive of the person acting, the nature of the act itself, or the consequences" (Furrow 2005). In order to evaluate the integration of biometric identity

technologies within SCS two conditions arise: 1) that ethical issues related to the integration of biometric identity technologies within SCS are based upon the idea that someone performs such action; and 2) the said action is only capable of being evaluated based upon the actions of the person(s) engaging and controlling them. If this is true and we endorse Furrows distinctions identified in the preceding passage, applying them to the integration of biometric identity technologies within SCS, there are three possible levels of ethical evaluation that need to be considered: i) we can evaluate the actions of the person(s) controlling the integration of biometric identity technologies within SCS; ii) we can evaluate the intentions of the person(s) controlling the integration of biometric identity technologies within SCS; and finally, iii) we can evaluate the consequences of the actions intended by the person(s) controlling the integration of biometric identity technologies within SCS. This must be realized from the perceptions of the stakeholders impacted most by the integration of biometric identity technologies within SCS (the Chinese citizens).

At the center of anticipatory ethics is the study of a technology, the technology behind the development of artifact, or a study of a specific technological artifact. Anticipatory ethics studies how each of these will work and projects trajectories of future technological developments, how they may work, and potential ethical issues related to how these future developments may work. A clear statement of the stages of technological development has been presented by Phillip Brey. According to Brey, as technology develops, there are a series of stages through which it passes: 1) the R & D stage; 2) the introduction stage; 3) the permeation stage; and 4) the power stage (Brey 2012). The proposed anticipatory ethical framework will, first, address the codification of salient SCS features (benefits) and consequences (punitive effects) on stakeholders, and then discuss the ethical issues related to how integrated biometric identity technology at different stages of developments are perceived (from both Eastern and Western perspectives).

Figure 1. Ethical and legal dilemmas of decisions



Source: Modified from Verne Henderson in Sloan Management Review (Henderson 1982, 42)

KEYWORDS: Facial Biometrics, Anticipatory Ethics, China's Social Credit System, Morality in Eastern and Western Culture.

REFERENCES

Brey, P. A. E. (2012). "Anticipating Ethical Issues in Emerging IT." *Ethics and Information Technology* 14 (4): 305–17.

Chen, Y., & Cheung, A. S. Y. (2017). "The Transparent Self Under Big Data Profiling: Privacy and Chinese Legislation on the Social Credit System." *Available at SSRN* 2992537. https://doi.org/10.2139/ssrn.2992537.

Creemers, R. (2018). "China's Social Credit System: An Evolving Practice of Control." *Available at SSRN 3175792*. https://doi.org/10.2139/ssrn.3175792.

Furrow, D. (2005). *Ethics: Key Concepts in Philosophy*. New York, NY: Bloomsbury Publishing.

Henderson, V. E. (1982). "The Ethical Side of Enterprise." *Sloan Management Review* 23 (3): 37.

Kostka, G. (2019). "China's Social Credit Systems and Public Opinion: Explaining High Levels of Approval." *New Media & Society* 21 (7): 1565–93.

Liang, F., Das, V., Kostyuk, N., & Hussain, M. M. (2018). "Constructing a Data-Driven Society: China's Social Credit System as a State Surveillance Infrastructure." *Policy & Internet* 10 (4): 415–53.

Wright, P. (2018). "Facial Recognition Database: Civil Liberties Nightmare." *Green Left Weekly*, no. 1201: 20.

7. Management of Cybercrime: Where To From Here?

Track chair: Shalini Kesar, Southern Utah University, USA

HOW TO BE ON TIME WITH SECURITY PROTOCOL?

Sabina Szymoniak

Czestochowa University of Technology (Poland)

sabina.szymoniak@icis.pcz.pl

EXTENDED ABSTRACT

The paper discusses a very important problem which is verification of security protocols. We propose a new method for security protocols verification including timed parameters and their influence on security. Our method includes analysis of time parameters using the specially implemented tool. We can calculate the correct time protocol execution, indicate time dependencies and check the possibility of Intruder's attack. We present on well-known Needham Schroeder protocol example.

1. INTRODUCTION

Our Internet communication and living in smart cities must be properly secured to avoid unauthorized interception of confidential information or fall victim to cybercrime. Users in the cyber world could be exposed to dishonest users' actions, called Intruders. For this purpose, security protocols (SP) are used. SP are designed to ensure that the transmitted data remain resistant to various attacks. From the ethical and social point of view the security protocols are a key element of sensitive data exchange. People should feel safe both walking on the streets and also walking on virtual paths. Security protocol guarantees that no unauthorized person should come into possession of sensitive data and use it for unethical purposes. In the case of SP, time plays an integral role. The passing of seconds can often allow an attacker to acquire knowledge to launch an attack. They can also allow decrypting intercepted messages.

SP should provide a proper level of security and secure society's existence. Due to continuous technological development, the security of protocols should be regularly verified to confirm their correctness. Over the past years, several methods to check security of the protocols have been introduced ((Paulson, 1999), (Burrows et al., 1989), (Lowe, 1996), (Steingartner et al., 2017), (Dolev et al., 1983), (Chadha et al., 2017), (Nigam et al., 2016), (Basin et al., 2018), (Siedlecka-Lamch O., et. Al, 2019)). However, these methods did not take into account the influence of time parameters on the SP' security.

Time properties were taken into account only in the Jakubowska and Penczek research (Jakubowska et al., 2006), (Jakubowska et al., 2007). Their research considered calculating the correct duration of the session. Unfortunately, these studies were not continued. A very interesting model of SP' executions was presented in (Kurkowski, 2013). Thanks to this model, it was possible to specify different in time protocol executions. We expanded this model by time parameters. We calculate the duration of the session and check how the included time parameters affect the security of protocol's users. Our considerations include both constant and random values of these parameters.

2. METHODS AND MATERIALS

According to Kurkowski's model we can generate a set of different in time protocol's executions, including four Intruder's models (Dolev-Yao, restricted Dolev-Yao, lazy Intruder and restricted lazy

Intruder). The model takes into account changes in participants' knowledge during the protocol. The model defines a set of time dependencies that allow to calculate the duration of a session and prepare appropriate time conditions.

In our approach, we consider the minimum, current for step and maximum value of delays in the network (D_{min},D_s,D_{max}) to determine the range of tested values of this parameter. A similar situation occurs for the step time (T_s^{min},T_s,T_s^{max}) and the session time (T_{ses}^{min},T_{ses},T_{ses}^{max}). The step time consists of message composition's time (T_c), encryption time (T_e), delays in the network and decryption time (T_d). The session time consists of all steps' times. The current value of each parameter refers to its value in the current step or session in progress. The step and session times depend on used delays in the network values. Lifetime is a value that cannot be exceeded in any of the executed steps. Exceeding its value will suggest protocol users that they are communicating with the Intruder and such connection should be immediately terminated. Lifetime value in one step is a sum of maximal step times of this step and next steps.

For our research, we created a tool to verify the timed security protocols. We can conduct two types of research on the loaded protocol. The first of them was timed analysis which enables the determination of limits for delays in the network and lifetime for which the protocol remains secure. The second type of research was simulations. We simulated delays in the network values to provide a real representation of the network.

3. EXPERIMENTAL RESULTS

We will use the Needham Schroeder (NSPK) protocol to present the results (Needham et al., 1978). This protocol consists of three steps, during which two honest users (*A, B*) exchange with each other messages with timestamps (T_A,T_B), IDs (I_A) encrypted by theirs public keys (K_A,K_B). The syntax of NSPK timed version in Common Language is presented in the first part of Figure 1.

Figure 1. Scheme of NSPK protocol and attack on it

$$\begin{array}{llll} \alpha_1 & A \to B & : & \{T_A, I_A\}_{K_B} \\ \alpha_2 & B \to A & : & \{T_A, T_B\}_{K_A} \\ \alpha_3 & A \to B & : & \{T_B\}_{K_B} \end{array} \qquad \begin{array}{llll} \alpha_1 & B \to I & : & \{I_B, T_B\}_{K_I}, \\ \beta_1 & I \to A & : & \{I_I, T_I\}_{K_A}, \\ \beta_2 & A \to I & : & \{T_I, T_A\}_{K_I}, \\ \alpha_2 & I \to B & : & \{T_B, T_A\}_{K_B}, \\ \alpha_3 & B \to I & : & \{T_A\}_{K_I}, \\ \beta_3 & I \to A & : & \{T_A\}_{K_A}. \end{array}$$

Source: self-elaboration based on (Needham et al., 1978)

In the second part of Figure 1, there is a scheme of the attacker's intrusion on this protocol. The Intruder (*I*) must execute additional steps (*β*), according to NSPK protocol, to acquire knowledge to complete *α*-execution.

In Table 1 we present our assumptions of time parameters, created according to NSPK protocol structure.

Timed analysis of NSPK protocol is presented in Table 2. Please note that even if Intruder will not end *β*-execution, α-execution will be correctly ended including additional steps' times. So the attack on this protocol is possible for assumed values of the time parameters.

In Figure 2 we show how changes in the delay in the network range would affect protocol security.

Table 1 Timed analysis assumptions

Parameter	Value in [tu]
T_e	5
T_d	5
T_c	2 (first and second step) 1 (third step)
$<D_{min},D_{max}>$	1-10
D	1
L_1	65
L_2	43
L_3	21
T_s^{max}	65

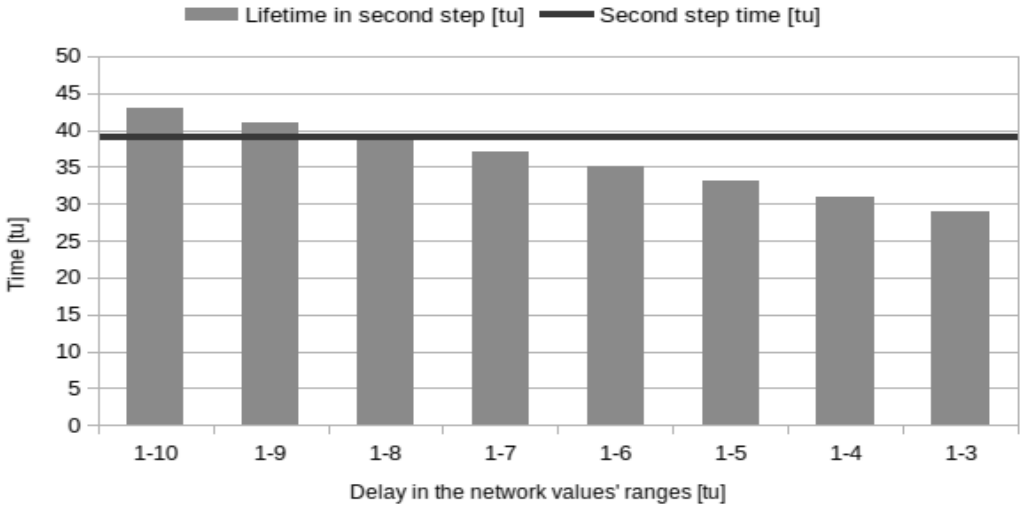
Source: self-elaboration

Table 2 Timed analysis of attacking execution in[tu]

α step	β step	T_e	T_c	D	T_d	T_s	T_{ses}	Result
α_1		5	2	1	5	13	13	ok
	β_1	5	2	1	5	13	26	ok
	β_2	5	2	1	5	13	39	ok
α_2		5	2	1	5	(13) **39**	52	ok
α_3		5	1	1	5	12	64	ok
	β_3	5	1	1	5	(12) **37**	76	$T_{ses}>T_{ses}^{max}$ && $T_3>L_3$

Source: self-elaboration

Figure 2. Changes in the delay in the network range



Source: self-elaboration

The setting the upper limit of delay in the network values to 7[tu] protects the protocol. In such a situation, a lifetime set in the second step will end the communication and the attack on NSPK protocol is not possible.

Next, we perform simulations NSPK protocol's executions using randomly generated values of delay in the network (according to normal, uniform, Cauchy's, Poisson's, exponential probability distributions). The values of the rest time parameters remained unchanged.

We carried out 18,000 test series for each probability distribution. In Table 3, we present time values for several test series classified with generated executions of NSPK protocol. The obtained results confirmed the considerations of time analysis. For such selected time parameter values, the protocol remained safe if the delay in the network value was equal to or less than 7[tu].

Table 3 Timed values for simulations according to an exponential probability distribution

Test series no.	Session time [tu]			The verage delay in the network [tu]
	Minimal	Average	Maximal	
1	17.03	17.31	18.38	1.1
2	17.02	17.31	18.09	1.1
3	17.07	17.2	17.33	1.4
4	32.13	32.61	33.76	1.1
5	17.01	17.1	17.31	1.37
6	32.13	32.59	33.75	1.1

Source: self-elaboration

4. CONCLUSION

In this paper, we discussed the problem of security protocols' verification. SP are widely used in the cyber world, so it is important to verify if they provide appropriate security level.

We presented a new approach to this issue. We take into account time parameters to calculate correct protocol's execution time and designate time dependencies. Such dependencies should protect prevent loss of confidential information. We researched by timed analysis and simulation of delays in the network.

Obtained results showed that time has a huge impact on protocols' security. Badly selected time dependencies could allow Intruder to perform additional actions to steal the data. During our research, we analyzed how delays in the network range affect Intruder's capabilities. We observed that if delays in the network range will be to extensive, it will not be secure for honest users because Intruder could have enough time to compromise the protocol.

KEYWORDS: timed analysis, security protocols, cybersecurity, verification.

REFERENCES

Basin D., Cremers C., Meadows C. (2018). Model Checking Security Protocols, in Handbook of Model Checking, Springer International Publishing

Burrows M., Abadi M., Needham R. (1989). A Logic of Authentication, In: Proceedings of the Royal Society of London A, vol. 426

Chadha R., Sistla P, Viswanathan M. (2017). Verification of randomized security protocols, Logic in Computer Science

Dolev D., Yao A. (1983). On the security of public key protocols. In: IEEE Transactions on Information Theory, 29(2)

Jakubowska G., Penczek W. (2006). Modeling and Checking Timed Authentication Security Protocols, Proc. of the Int. Workshop on Concurrency, Specification and Programming (CS&P'06), Informatik-Berichte 206(2)

Jakubowska G., Penczek W. (2007). Is your security protocol on time?, In Proc. Of FSEN'07, volume 4767 of LNCS, Springer-Verlag

Kurkowski M. (2013). Formalne metody weryfikacji wlasności protokolow zabezpieczajacych w sieciach komputerowych, in polish, Exit, Warsaw

Needham R. M., Schroeder M. D. (1978). Using encryption for authentication in large networks of computers. Commun. ACM, 21(12)

Nigam V., et. al (2016). Towards the Automated Verification of Cyber-Physical Security Protocols: Bounding the Number of Timed Intruders, Computer Security – ESORICS 2016", Springer International Publishing

Paulson L. (1999). Inductive Analysis of the Internet Protocol TLS, ACM Transactions on Information and System Security (TISSEC), vol 2 (3)

Siedlecka-Lamch O., et. al (2019) A fast method for security protocols verification, Computer Information Systems and Industrial Management, Springer

Steingartner W., Novitzka V. (2017). Coalgebras for modelling observable behaviour of programs, In: Journal of applied mathematics and computational mechanics. 16(2)

LEGAL AND ETHICAL CHALLENGES FOR CYBERSECURITY OF MEDICAL IOT DEVICES

Jaroslaw Greser

Poland Adam Mickiewicz University

jarek.greser@gmail.com

EXTENDED ABSTRACT

INTRODUCTION

Development of communication technologies, particularly Internet of Things, creates new ways of providing health care services. Thanks to it, persons requiring permanent medical supervision, e.g. suffering from chronic diseases, or the elderly can be granted better medical care. Inventions such as support schemes for people with Parkinson's disease[14], devices for 24-hour heart rhythm monitoring which transmit the information on an ongoing basis to the doctor or CTG devices for home use are results of this development. Another instance of such devices are straps monitoring bodily functions, which task is to inform certain people about heart rate, body temperature and fall of person wearing it. Research indicates that such solutions contribute to reduction of number of primary health care patients of 20%[15]. They are also improving the standard of care for patients with chronic diseases[16].

Data pertaining to the health status can be also obtained from other loT devices. Example are so called *wearables* devices type, like bands that provide information about your physical activity or smartphones gathering such pieces of information[17]. On this basis, not only some information about user present health status may be collected but also user's future condition can be predicted.

Health information, besides its primary purpose – diagnosis of health condition – may be relevant for representatives of different industries, particularly for insurance sector and marketing of goods and services. At the same time, they can be used to perform actions that breach either the ethics – e.g. to profil the partner in negotiations – or the law – for instance to blackmail a given individual. This is also the reason why medical IoT devices constitute significant goal for cybercriminals. The attack on database maintained by Singapore Health Services Private Limited, resulting in the leakage of data of approximately 1.5 million of patients, including sensitive data such as race, is an excellent illustration of this issue. In about 10% of cases also the medical records were stolen, e.g., the list of prescribed

[14] https://www-03.ibm.com/press/us/en/pressrelease/49475.wss

[15] *S. Fielding, T. Porteous, J. Ferguson, V. Maskrey, A. Blyth, V. Paudyal, G. Barton, R. Holland, C.M. Bond, M.C. Watson*, Estimating the burden of minor ailment consultations in general practices and emergency departments through retrospective review of routine data in North East Scotland, Family Practice, 2015 nr 2.

[16] D. *Su, J. Zhou, M. S. Kelley, T.L. Michuad, M. Siahpush, J. Kim, F. Wilson, J.P. Stimpson, J.A. Pagánde*, Does telemedicine improve treatment outcomes for diabetes? A meta-analysis of results from 55 randomized controlled trials, Diabetes Research and Clinical Practice 2016, Nr 116.

[17] Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions on enabling the digital transformation of health and care in the Digital Single Market; empowering citizens and building a healthier society; https://ec.europa.eu/digital-single-market/en/news/communication-enabling-digital-transformation-health-and-care-digital-single-market-empowering

medications together with dosage. Prime minister of Singapore was one of the victims of these attack[18].

Simultaneously, Internet of Things technology is considered to be the most prone to cyberattacks. It is assessed, that 70% of IoT devices is incorrectly secured[19]. Several reasons of this phenomenon are being indicated[20]. Among them we can mention the amounts and variety of types of such equipment – its estimated number in 2020 shall be between 25 and 30 billions[21]. This leads to use of different solutions, often proprietary, which are not updated after completion of the project. Simultaneously due to licence conditions users are unable to change the software stand- alone[22]. Another reason is the common practice to protect equipment by standard passwords without forcing the users to change them[23]. This can result in unauthorized access to data, as well as device damage[24].

Despite the existence of these risks, the problematic related to the Internet of Things remains unregulated by the law. The only exception is the law of the State of California, which imposes on manufacturers offering devices for sale in this State, an obligation to ensure appropriate measures in the field of cybersecurity and to assign to each device a unique password[25].

MAIN AIM OF THE PAPER

In the view of the above remarks a series of questions regarding the role of companies producing medical IoT devices in providing their security arises. It is essential to find standards designating obligations of the companies in this respect. In my paper I would like to focus on two kinds of standards.

The first of these are legal standards. In this respect I will discuss legal requirements for cybersecurity of medical devices imposed by EU law. Specifically, I will examine provisions of regulation 2017/745, the GDPR, Directive 2016/1148 and ENISA guidelines related to cybersecurity of technology IoT. In addition, I will discuss main problems remaining outside the regulation area.

The second area I would like to discuss is the ethical question concerning scope of liability of traders for their products in the case when regulatory framework is unclear or does not even exist. In this respect, I will attempt to answer to the question if a company has an ethical duty to maintain a high level of cybersecurity of their appliances. I will consider this issue in two cases. First of them is ceasing the support of a device and withdrawing it from the company's offer. The second regards equipment

[18] Public Report of the Committee of Inquiry into the Cyber Attack on Singapore Health Services Private Limited's Patient Database on or Around 27 June 2018 10 January 2019, https://iapp.org/media/pdf/publications/Report_of_the_COI_into_the_Cyber_Attack_on_SingHealth_10_Jan_2019.pdf

[19] Hewlett Packard Internet of Things Research Study, 2015, s. 3, http://www8.hp.com/h20195/V2/GetPDF.aspx/4AA5-4759ENW.pdf;

[20] OWASP Internet of Things (IoT) Project https://www.owasp.org/index.php/OWASP_Internet_of_Things_Project

[21] https://www.gartner.com/en/newsroom/press-releases/2017-02-07-gartner-says-8-billion-connected-things-will- be-in-use-in-2017-up-31-percent-from-2016

[22] Baseline Security Recommendations for IoT in the context of Critical Information Infrastructures, European Union Agency For Network And Information Security 2017, s. 44. https://www.enisa.europa.eu/publications/baseline-security-recommendations-for-iot/at_download/fullReport

[23] Code of Practice for Consumer IoT Security, Departament for Digital, Culture Media&Sport 2018, s. 6-9. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/747413/Code_of_Practice_for_Consumer_IoT_Security_October_2018.pdf

[24] *Z. Gwarzo*, Security and Privacy Issues in Internet of Things, Jusletter IT 2016, s.1

[25] Senate Bill Nr 327 z 28.9.2018 https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201720180SB327

manufacturer informing about gap in the security setup in case it is caused by the subcontractor's mistake.

KEYWORDS: cybersecurity, IoT, medical devices, GDPR, sensitive data.

REFERENCES

Fielding S., Porteous T., Ferguson J., Maskrey V., Blyth A., Paudyal V., Barton G., Holland R., Bond C.M., Watson M.C., Estimating the burden of minor ailment consultations in general practices and emergency departments through retrospective review of routine data in North East Scotland, Family Practice, 2015 nr 2.

Su D., Zhou J., Kelley M. S., Michuad T.L., Siahpush M., Kim J., Wilson F., Stimpson J.P., Pagánde J.A., Does telemedicine improve treatment outcomes for diabetes? A meta-analysis of results from 55 randomized controlled trials, Diabetes Research and Clinical Practice 2016, Nr 116.

Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions on enabling the digital transformation of health and care in the Digital Single Market; empowering citizens and building a healthier society. Retrieved from https://ec.europa.eu/digital-single-market/en/news/communication-enabling-digital-transformation-health-and-care-digital-single-market-empowering

Public Report of the Committee of Inquiry into the Cyber Attack on Singapore Health Services Private Limited's Patient Database on or Around 27 June 2018 10 January 2019, Retrieved from https://iapp.org/media/pdf/publications/Report_of_the_COI_into_the_Cyber_Attack_on_SingHealth_10_Jan_2019.pdf

Hewlett Packard Internet of Things Research Study, 2015. Retrieved from http://www8.hp.com/h20195/V2/GetPDF.aspx/4AA5-4759ENW.pdf;

Baseline Security Recommendations for IoT in the context of Critical Information Infrastructures, European Union Agency For Network And Information Security 2017. Retrieved from https://www.enisa.europa.eu/publications/baseline-security-recommendations-for-iot/at_download/fullReport

Code of Practice for Consumer IoT Security, Departament for Digital, Culture Media&Sport 2018 . Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/747413/Code_of_Practice_for_Consumer_IoT_Security_October_2018.pdf

Z. Gwarzo, Security and Privacy Issues in Internet of Things, Jusletter IT 2016.

Senate Bill Nr 327 z 28.9.2018. Retrieved from https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201720180SB327[1]

OWASP Internet of Things (IoT) Project. Retrieved from https://www.owasp.org/index.php/OWASP_Internet_of_Things_Project

SMART CITIES BRING NEW CHALLENGES IN MANAGING CYBERSECURITY BREACHES

Shalini Kesar

Southern Utah University (USA)

kesar@suu.edu

EXTENDED ABSTRACT

This paper outlines some challenges and suggestion to manage and minimize cybersecurity breach within smart cities. According to definition, a smart city is designation given to a city that incorporates Information and Communication Technologies (ICT) to enhance the quality and performance of urban services such as energy, transportation and utilities in order to reduce resource consumption, wastage and overall costs. A smart city is a complex ecosystem where city infrastructure that are constantly interacting with each other with technology. For example, public and private entities, people, processes, devices, and city infrastructure. In general, smart cities use connected technology and data to (1) improve the efficiency of city service delivery (2) enhance quality of life for all (3) increase equity and prosperity for residents and businesses. Reports echo that underlying technology infrastructure of the ecosystem comprises three layers: the edge, the core, and the communication. The edge layer comprises devices such as sensors, actuators, other IoT devices, and smartphones. The core is the technology platform that processes and makes sense of the data flowing from the edge. The communication channel establishes a constant, two-way data exchange between the core and the edge to seamlessly integrate the various components of the ecosystem. The growth of Smart Cities is projected to increase fourfold by 2025 and it will continue.

Cybersecurity, by definition refers to a set of techniques used to protect the integrity of an organization's security architecture and safeguard its data against attack, damage or unauthorized access. In a report (2019) commented that smart cities are increasingly under attack by a variety of threats. Further, "these include sophisticated cyberattacks on critical infrastructure, bringing Industrial Control Systems (ICS) to a grinding halt, abusing Low-Power Wide Area Networks (LPWAN) and device communication hijacking, system lockdown threats caused by ransomware, manipulation of sensor data to cause widespread panic (e.g., disaster detection systems) and siphoning citizen, healthcare, consumer data, and personally identifiable information (PII), among many others," explains Dimitrios Pavlakis, Industry Analyst at ABI Research. "In this increasingly connected technological landscape, every smart city service is as secure as its weakest link." (Help Net Security, 2019). Across the globe, cybersecurity breaches are increasing. Recent report conducted by Symantec Internet Security Threat Report (2019) highlighted some alarming figures: average of 4,800 websites compromised each month; Ransomware shifted targets from consumers to enterprises, where infections rose 12 percent; More than 70 million records stolen from poorly configured S3 buckets, a casualty of rapid cloud adoption; Internet of Things (IoT) was a key entry point for targeted attacks.

Other statistics indicate the worldwide cybersecurity forecast: in 2022, the cost of cyber breaches will reach $133.7 billion; 62% of businesses experienced phishing and social engineering attacks in 2018; 68% of business leaders feel their cybersecurity risks are increasing; only 5% of companies' folders are properly protected, on average; 52% of breaches featured hacking, 28% involved malware and 32–33% included phishing or social engineering, respectively (Verizon, 2019).

Dependency on technology has a lot of beneficial to people living in smart cities. At the same time technology brings risks and vulnerabilities to cybersecurity. There are many examples of cyber breaches in smart cities that have warrant us to think about consequences and solutions. For example,

Atlanta, capital of Georgia State in the United States, faced SamSam, a ruthless "ransomware" bug in March 2019. This lasted to approximately two weeks and a cost $55,000 worth of bitcoin in payment was demanded. The aftermath of denying the demand left Atlanta City processing reports and legal documents, which cost of this attack in millions. Baltimore, another smart city in the United States faced cyber breach. The attack involved a ransomware attack that led to accessibility issues to their Computer Aided Dispatch (CAD) system of Emergency services for 17 hours. The city's Emergency services rely on this system to automatically divert calls to emergency responders who are closest in location so that emergency assistance is directed as efficiently as possible. While the system was down responders operated by taking phone calls manually, a far slower process and one which could have had a more sinister outcome if the cyber-attack had been prolonged. There are many other examples of cyber breaches in smart cities. This will concern will keep increasing. It has been pointed that by 2050, about 66% of the world's population is expected to live in cities. Methods to minimize, manage, and mitigate cyber breaches should be one of the first priority while using advanced technologies within smart cities. While developing solutions, it is important to keep in mind some of the realities of cybersecurity in smart cities: 1) The introduction of new web and mobile apps, IoT, connected homes, connected cars and even connected logistics. The increase use of such new technologies will make more data and gadget accessible to criminals. 2) Cybercriminal motivations increase to get access to IoT. The sophistication of technology also means increases in skills and tactics of cyber criminals. 3) Lack of skill of cybersecurity experts. Statistics indicate there is more demand than supply for cybersecurity experts. This reality can become a problem as cities become more dependent on technology for everyday activities. In addition to the realities of cybersecurity and smart cities, ethical implications also become a concern.

This paper reflects on the suggested three layers of ecosystem in context of smart cities: the edge, the core, and the communication. Further it reviews cybersecurity challenges in the ecosystem of smart cities. This is significant since it is estimated that the world's urban population will rise by 72 per cent between 2011 and 2050. To combat this growing demand, it is important to keep a check on use and misuse of technology. This will help smart city service providers such as networking Internet of Things (IoT) technology with existing infrastructure are balanced and reshape supply chains and manage assets and resources more efficiently.

KEYWORDS: Smart cities, cybersecurity, Cyber breaches.

REFERENCES

Help Net Security. (2019). Cybersecurity challenges for smart cities: Key issues and top threats. Retrived from https://www.helpnetsecurity.com/2019/08/21/cybersecurity-smart-cities/

Symantec Security. (2019). Internet Security Report. Retrieved from https://www.frontiersin.org/articles/438810

Rowling, J.K. (2001). *Harry Potter and the socerer's stone*. London: Bloomsburg Children's.

Sanders, S. R. (2007). [Introduction]. In L. Williford & M. Martone (Eds.), *Touchstone anthology of contemporary creative nonfiction: Work from 1970 to present* (pp. 148-151). New York, NY: Simon & Schuster.

Verizon. (2019), *2019 Data Breach Investigations*. https://enterprise.verizon.com/en-gb/resources/reports/dbir/

8. Marketing Ethics in Digital Environments

Track chairs: Crisitina Olarte Pascual, University of La Rioja, Spain – Eva Reinares, Universidad Rey Juan Carlos, Spain – Jesús García de Madariaga, Complutense University of Madrid, Spain – Teresa Pintado, Complutense University of Madrid, Spain

BRAND, ETHICS AND COMPETITIVE ADVANTAGE

Orlando Lima Rua, António Oliveira

Centro de Estudos Interculturais (CEI), P.PORTO/ISCAP (Portugal)

orua@iscap.ipp.pt; ajmo@iscap.ipp.pt

EXTENDED ABSTRACT

Media refer that customers buy brands because of ethical or moral concerns and that the majority of the population confirms this by saying that they take moral and ethical concerns in account while making a purchasing decision. Devinney (2012) proves this statement wrong with his research. He sustains that customers indeed care about ethical concerns of products, brands and companies but that these are not the most important ones in customers' purchasing decisions. On the contrary of what the media call "ethical consumers", the majority of the consumers chose price, taste, positioning, the context in which they buy, and convenience over a company's ethical concerns. To summarise, there is an attitude gap, which means that consumers say they make their purchasing decisions based on ethical reasons, while in reality, the company's social and ethical responsibility isn't the most vital criterion.

On the other hand, marketing strategies should be aware of online social networking tends in the digital domain, where consumers are able to communicate more proactively (Tiago & Veríssimo, 2014). Knowledge and relationship are boosted by the digital dimension, allowing individuals to share different cultures (Budden, Anthony, Budden, & Jones, 2011; Kumar, Novak, & Tomkins, 2010).

The brand is one of the most fundamental intangible resources (Kayo, 2002). Fakhrutdinova, Kolesnikova and Yurieva (2014) state that this should be the cornerstone of a sustained and differentiated international strategy that can ensure the organization's competitive advantage and clearly communicate its positioning to its audiences (Morgan and Pritchard, 2004). Holt, Quelch and Taylor (2004) argue that brand value is even more relevant in an international context, where competitiveness levels are also higher. Thus, it should convey a unified and coherent idea, but it must also be adapted to local specificities, ie, it should be oriented to the markets in which it operates (Kirca Jayachandran and Bearden et al., 2005), ensuring an effective response to consumers' needs and demands (Kohli & Jaworsk, 1990).

Market orientation is the concern of an organization to understand and respond to the characteristics of the market in which it operates (Kohli et al., 1990), shifting its focus from an internal perspective to an external perspective (Kirca, Jayachandran & Bearden, 2005). Popoli (2015) argues that the organization's direct link to the needs of its markets is central to the effective and full realization of a strategy through differentiation, so the present study aimed to test the following hypothesis:

There is a progression where internationalization is defined by the "specific priorities of a country, institution or a specific group of stakeholders" (Knight, 2015, p. 2), and where the success of its implementation depends not only on the costs of but of the ability of organizations to understand differences in the home and international markets, as well as to develop competitive advantage and respond to the difficulties that arise from this heterogeneity (Brouthers, Brouthers & Werner, 2008; Hitt, Ireland & Hoskisson, 2007; He, 2012), through a recognized and differentiated brand (Popoli, 2015).

We intend to conduct an online survey to marketing managers from Portuguese footwear firms analysing their (1) Facebook, (2) Twitter, (3) Orkut and (4) blogs. Based on the preceding discussion the following hypotheses were developed for this study:

H1: Brand has a positive effect on market orientation on digital consumers.

H2: Brand has a positive effect on competitive advantage through differentiation on digital consumers.

H3: Market orientation has a positive effect on competitive advantage through differentiation on digital consumers.

H4: Market orientation mediates the relationship between brand and competitive advantage through differentiation on digital consumers.

KEYWORDS: Brand, Ethics, Market orientation, Competitive advantage, Digital consumers.

REFERENCES

Brouthers, K. D., Brouthers, L. E., & Werner, S. (2008). Resource-based advantages in an international context. *Journal of Management*, 34(2), 189-217.

Budden, C. B., Anthony, J. F., Budden, M. C., & Jones, M. A. (2011). Managing the evolution of a revolution: Marketing implications of Internet media usage among college students. *College Teaching Methods and Styles Journal*, 3(3), 5-10.

Devinney, T. (2012). *Do consumers really care? The myth of the ethical consumer*. Retrieved from http://www.modern-cynic.org/2012/02/27/do-consumers-really-care-the-myth-of-the-ethical-consumer/

Fakhrutdinova, E., Kolesnikova, J., & Yurieva, O. K. (2014). Current trends of realization of the intellectual capital and problems of intellectual migration. *Procedia Economics and Finance*, 14(1), 326-332.

Hitt, M. A., Ireland, R. D., & Hoskisson, R. E. (2007). *Strategic management: competitiveness and globalization*. Mason: Thomson SouthWestern.

Holt, D. B., Quelch, J. A., & Taylor, E. L. (2004). How global brands compete. *Harvard Business Review*, Setembro, 1-9.

Kayo, E. K. (2002). *A estrutura de capital e o risco das empresas tangível e intangível-intensivas: uma contribuição ao estudo da valoração de empresas*. Tese Doutorado em Administração. FEA/USP.

Kirca, A. H., Jayachandran, S., & Bearden, W. O. (2005). Market orientation: A meta-analytic review and assessment of its antecedents and impact on performance. *Journal of Marketing*, 69(1), 24–41.

Knight, J. (2015). Updated Definition of Internationalization. *International Higher Education*, 33. doi.org/10.6017/ihe.2003.33.7391

Kohli, A. K., & Jaworski, B. J. (1990). Market orientation: the construct, research propositions and managerial implications. *Journal of Marketing*, 54(2), 1-18.

Kumar, R., Novak, J., & Tomkins, A. (2010). Structure and evolution of online social networks. In P. S. Yu, J. Han, & C. Faloutsos (Eds.), *Link mining: Models, algorithms, and applications* (pp. 337-357). New York: Springer.

Morgan, N., &. Pritchard, A. (2004). *Meeting the destination branding challenge*. In N. Morgan, A. Pritchard & R. Pride (2004). Destination branding: Creating the unique destination proposition (Vol. 2, pp. 59-78). Burlington: Elsevier Butterworth-Heinemann.

Popoli, P. (2015). Reinforcing Intangible Assets through CSR in a Globalized World. *Journal of Management Policies and Practices*, 3(1), 23-30.

Tiago, M.T.P.M., & Veríssimo, J.M.C. (2014). Digital marketing and social media: Why bother? *Business Horizons*, 57, 703-708.

ETHICAL CHALLENGES OF ONLINE PANELS BASED ON PASSIVE DATA COLLECTION TECHNOLOGY

Jesús García-Madariaga, Ingrit Moya Burgos, María-Francisca Blasco López, Pamela Simón Sandoval

Complutense University of Madrid (Spain)

jesgarci@ucm.es; ingritvm@ucm.es; fblasco@ucm.es; pamsimon@ucm.es

EXTENDED ABSTRACT

The first record of a formal survey application is dated in 1879 when the agency NW Ayer & Son asked people about their expectations of grain production in the United States (Maclaran, 2009). Since then the market research industry has drastically changed. This industry has been compelled to adapt to the new digital environment and move forward to align itself with advances, both technological and societal.

It is true that the main goal of market research is still enhance the understanding of consumers. However, in this context of technological advances, the methodologies used to find out how people feel, think and perceive their environment are not the same.

As aforehand mentioned, at the beginning of market research, techniques were based on person to person interactions but with the advent of the internet, on-line communications started taking place and online panels become a prominent way to collect survey data (Callegaro et al., 2014). Therefore, the online panels became the natural evolution of the traditional consumer panels, used for decades on the basis of mail, phone, and face-to-face surveys (Vaygelt, 2006).

The online panels are defined in the international standard, ISO 20252 as "a sample database of potential respondents who declare that they will cooperate for future data collection if selected" (ISO, 2012, p. 1) and its huge attraction is based on threefold advantages: (1) fast data collection; (2) lower cost per data; and (3) sampling efficiency due to extensive profiling (Callegaro et al., 2014). Those advantages have been potentiated with the emergence of mobile technology that not only are increasingly replacing computers as the primary way people access the web (Stern, et al., 2014) but also provide the consumer access to an omnipresent environment (Pelet and Papadopoulou, 2016).

In addition, mobile technology has opened new opportunities to online panels through the possibility to monitor consumers in real-time by using passive mobile data collection (Link et al., 2014). This technology allows obtaining data directly from mobile devices including geolocation, physical movements, browser history, app usage, and call and text message logs. These data enable researchers to study, among other things, users' mobility patterns, physical activity and health, consumer behaviour, and social interactions (Keusch et al., 2019) avoiding the bias of self-reporting techniques and the measurement errors.

The rapid adoption of this approach of market research has supported a growing body of literature oriented to assess the use of smartphones for data collection (Revilla, Ochoa and Loewe, 2016; Elevelt, Lugtig and Toepoel 2017). Among recent studies highlights those regarding the variables that influence the willingness to participate in the passive mobile measurement (Revilla, Couper, and Ochoa 2018; Wenz, Jäckle, and Couper 2017). Those studies point out that participation in passive data collection is influenced by three variables: (1) Perceived benefits and interest, (2) Privacy and risk of disclosure and (3) Comfort and experience with the data collection process. In particular, privacy aspects are

often mentioned as a barrier for being willing to participate (Revilla, Couper and Ochoa 2018; Revilla, Ochoa and Loewe, 2016).

Regarding this issue, Maher, et al (2019) found that current privacy procedures for passive data collection are insufficient and explain that most people do not read or understand consent forms but instead, they tend to agree to terms and services reflexively. Moreover, due to the passive nature of the data, consumers are not aware of the type, amount, or implications of the collected data (Weinberg et al., 2015). Therefore, although passive data collection has the potential to provide a vast amount of data on consumer behaviour, its velocity of data generation and sharing, which will continue increasing, will have a cost on consumer data integrity if the industry does not provide an ethical framework able to boost passive data-driven innovation while protecting the consumer.

Based on aforehand mentioned, the aim of the present research is to review the current ethical literature to characterize the ethical challenges and suggested solutions related to the successful implementation of passive data collection in market research. To accomplish that goal, it will be identified and analyse all relevant papers published in the last 10 years, by using as inclusion criteria all papers that discuss normative standpoints of ethical issues regarding the use of passive data in market research.

The obtained results will provide a systematic review that will contribute to identifying the main ethical concerns, normative standpoints, and underlying arguments related to the use of mobile passive data in market research. Besides, the present study will lay a foundation for the construction of an ethical framework related to the use of data collected by using passive metering technology no matter the type of device.

KEYWORDS: ethics, privacy, market research, online panels, passive data collection, mobile technology.

REFERENCES

Callegaro, M., Baker, R., Bethlehem, J., Goritz, A. S., Krosnick, J. A., & Lavrakas, P. J. (2014). Online panel research: History, concepts, applications and a look at the future.

Comley, P. and Beaumont, J. (2011) Online market research: Methods, benefits and issues. Journal of Direct, Data and Digital Marketing Practice. Volume 12, Issue 4, pp 315–327.

Elevelt, Anne, Peter Lugtig, and Vera Toepoel. 2017. “Doing a Time Use Survey on Smartphones Only: What Factors Predict Nonresponse at Different Stages of the Survey Process?” 7th Conference of the European Survey Research Association (ESRA), July 17–21, Lisbon, Portugal.

Keusch, F., Struminskaya, B., Antoun, C., Couper, M. P., & Kreuter, F. (2019). Willingness to participate in passive mobile data collection. Public opinion quarterly, 83(S1), 210-235.

International Organization for Standardization (ISO). (2009). ISO 26362 Access panels in market, opinion, and social research: Vocabulary and service requirements. Geneva: ISO.

International Organization for Standardization (ISO). (2012). ISO 20252 Market, opinion and social research: Vocabulary and service requirements (2nd ed.). Geneva: ISO.

Keusch, F., Struminskaya, B., Antoun, C., Couper, M. P., & Kreuter, F. (2019). Willingness to participate in passive mobile data collection. Public opinion quarterly, 83(S1), 210-235.

Link, Michael W., Joe Murphy, Michael F. Schober, Trent D. Buskirk, Jennifer Hunter Childs, and Casey Langer Tesfaye. 2014. “Mobile Technologies for Conducting, Augmenting and Potentially Replacing Surveys: Executive Summary of the AAPOR Task Force on Emerging Technologies in Public Opinion Research.” Public Opinion Quarterly78:779–87.

Maclaran, P. (Ed.). (2009). The SAGE handbook of marketing theory. Sage Publications. USA.

Maher, N. A., Senders, J. T., Hulsbergen, A. F., Lamba, N., Parker, M., Onnela, J. P., ... & Broekman, M. L. (2019). Passive data collection and use in healthcare: A systematic review of ethical issues. International journal of medical informatics.

Mutepfa, M. M., & Tapera, R. (2019). Traditional Survey and Questionnaire Platforms. Handbook of Research Methods in Health Social Sciences, 541-558.

Nunan, Daniel and Di Domenico, M. (2013) Market research & the ethics of big data. International Journal of Market Research 55 (4), pp. 505-520.

Pelet, J. E., and Papadopoulou, P. (2016). Consumer behavior in the mobile environment: An exploratory study of m-commerce and social media. In Geospatial Research: Concepts, Methodologies, Tools, and Applications (pp. 1168-1182). IGI Global.

Revilla, Melanie, Mick P. Couper, and Carlos Ochoa. 2018. “Willingness of Online Panelists to Perform Additional Tasks.” Methods, Data, Analyses. Advance Access publication June 2018

Revilla, M., Ochoa, C., & Loewe, G. (2016). Using passive data from a meter to complement survey data in order to study online behavior. Social Science Computer Review, 35(4), 521-536.

Stern, M. J., Bilgen, I., & Dillman, D. A. (2014). The state of survey methodology: Challenges, dilemmas, and new frontiers in the era of the tailored design. Field Methods, 26(3), 284-301.

Vaygelt, M. (2006). Emerging from the shadow of consumer panels: B2B challenges and best practices. Panel Research 2006. ESOMAR.

Weinberg, B. Milne, G. Andonova, Y. Hajjat, F. (2015) Internet of Things: convenience vs. Privacy and secrecy. Bus. Horiz., 58 (November (6)) (2015), pp. 615-624

Wenz, Alexander, Annette Jäckle, and Mick P. Couper. (2017). “Willingness to Use Mobile Technologies for Data Collection in a Probability Household Panel” ISER 2017-10.

ETHICAL DILEMMAS IN NON PROFIT ORGANIZATIONS CAMPAIGNS

María Francisca Blasco López, Jesús García de Madariaga,
Ingrit Moya Burgos, Pamela Simón Sandoval

Complutense University of Madrid (Spain)

fblasco@ucm.es; jesgarci@ucm.es; ingritvm@ucm.es; pamsimon@ucm.es

EXTENDED ABSTRACT

Nonprofit Organizations (NPO) help coordinate humanitarian activities for the less fortunate and attempt to build a humane society (Chang & Lee, 2009). For which, NPO advertisements aim to motivate people to donate either money or time (Reed, Aquino, & Levy, 2007). Many Non Profit Organizations have created a presence via websites, to generate awareness of their causes, for fundraising and for managing their brands. But now they are trying to employ the potential of online social network sites (Quinton & Fennemore, 2013) to achieve their goals. There is no doubt that online social networks is now integrated into the daily lives of millions of people worldwide, in Spain 85% of the internet users from 16 to 65 years old are using social networks (ELOGIA, 2019). So, this is a very essential tool for NPO to communicate their advertisement campaigns.

It is well known that Non Profit Organizations (NPO) use emotional images to persuade people to donate to their cause. Since NPO advertising tries to touch the desire of people of helping those less fortunate, this type of sector tend to use affective effect in their advertisings. It is widely recognized that emotional appeals are very effective tools for persuasion (Burt & Strongman, 2005; Bagozzi & Moore, 2006; Poels & Dewitte, 2009). Emotional appeals are instrumental in providing the creative punch to enhance persuasion (Bebko, Sciulli, & Bhagat, 2014). Emotions have the ability to capture attention, influence attitudes, and affect consumer behavior.

But the problem is that some NPO have been collecting a huge amounts of donations but maybe doing a lot of harm at the same time. Using images of people in developing countries that manifest suffering by starving people, begging eyes, and distended bellies, to gather donations. This type of biased images are known as the “Starving Baby Appeal” (Fine, 1990, p. 154). This kind of campaigns involve inherent ethical concerns and dilemmas.

One of the ethical dilemmas is that this type of images makes that people have a view of the Third World as a place of misery. The extensive sense of hopelessness, which is strengthen by the news and NPO campaigns, understandably incite responses that range from indifference to aversion (Nathanson, 2013).

According to McQuillin and Sargeant (2017) fundraising ethics has received little scholarly attention. There is sparse work of ethical theorizing by scholars in philanthropy and fundraising which haven´t proposed a coherent normative theory that might inform the ethics applied in these profession (MacQuillin & Sargeant, 2018). But the problem is not new, there have been critics that while images of suffering and desperate people may capture the attention, move emotions and promote donations, they also illustrate people from developing countries as hopeless and helpless, without the support of these organizations (Nathanson, 2013). The term 'degree zero images' refers to those images that want to illustrate a given portion of reality in an precise way (Grancea, 2015).

As stated in the Association of Fundraising Professionals’ International Statement on Ethical Principles in Fundraising (2018) “Funds will be collected carefully and with respect of donor’s free choice, without

the use of pressure, harassment, intimidation or coercion." This statement presents another ethical problem since a person might consider felt pressured because they saw an advertisement were they made him feel guilty by the use of explicit images and threatening messages, and felt they must do something but couldn´t afford it, making him feel guilty about it. So, the general ethical question of whether it is appropriate for donors to feel guilty if they decide not to donate to a cause, might be considered a form of pressure (MacQuillin & Sargeant, 2018).

So, negatively-valence images that present victims of different social problems may suggest ethical problems. The accusations most frequently appealed include, increased anxiety among vulnerable viewers, increased satisfaction from those not affected by the problem, lack of respect for the dignity of the people presented in the advertisement (Grancea, 2015).

Several ethical dilemmas in fundraising emerge because of a pressure between what the NPO needs to do for their beneficiaries (requesting in the most efficient and effective ways to guarantee enough money to help them) and what the donor wants (been asked less, asked in different ways, or simply not asked at all).

Thus, it is possible that advertising may have harmful effects for individuals and society that deserve ethical analysis. Furthermore, social marketers should notice that target viewers can have reservations about the ethicality of social advertising, even when they know their intentions are good (Hastings, Stead, & Webb, 2004).

So the purpose of this paper is to examine how consumers react to print advertisements of NPO that use negative-valence images. We will conduct an experiment using Neuromarketing techniques (EEG, GSR and Eye-tracking) and declarative methodologies to analyse the reactions (conscious and unconscious) of a group of 30 subjects (W-M) aged between 18-34 years.

All participants will be tested individually, and will be asked to observe a series of images including advertisements of NPO using negative-valence images. Afterwards participants will fill out a questionnaire related to their experience observing the advertisements, were they will be asked to check 12 adjectives that they found appropriate to describe the advertisement, six negative (threatening, sad, violent, offensive, uncomfortable and disgusting) and six positives (happy, friendly, interesting, convincing, creative and informative). Then they will be asked the level of guilt aroused by the advertisement to measure it, it will be used the guilt scale of (Coulter & Pinto, 1995), using a 7-point Likert-scale ranging from "not at all" to "very strongly" to measure the intensity of each feeling. And that includes the following items: guilty, bad, ashamed, upset, irresponsible, accountable, and uneasy. Also it will be measured the manipulative intent using the Inferences of Manipulative Intent (IMI) scale (Campbell, 1995). They will indicate the extent to which they agree or disagree with each of the statements relating to the advertiser´s manipulative intent (ex. The way this ad tries to persuade people seems acceptable to me, the advertiser tried to manipulate the audience in ways I do not like, I was annoyed by this ad because the advertiser seemed to be trying to inappropriately manage, I didn't mind this ad; the advertiser tried to be persuasive without being excessively manipulative, The ad was fair in what was said and shown. I think that this advertisement is unfair/fair.) (Hibbert, Smith, Davies, & Ireland, 2007).

The results obtained will help us examine the impact of negative valence images on consumers experience and their neurophysiological and behavioural reactions. This will provide evidence of weather this type of advertisements has a negative effect on consumers. Finally, this research implies direction for future research which can also analyse the effect of these negatively-valence images on government institutions since these advertisements may show the economic and social problems of a country.

KEYWORDS: Ethics, Non Profit Organizations, Advertising, Negative valence image, Neuromarketing.

ACKNOWLEDGMENTS: This work is supported by the Spanish Ministry of Economy under grant RTC2016-4718-7. The authors also gratefully acknowledge the support of BitBrain Technologies.

REFERENCES

Association of Fundraising Professionals. (2017). International statement on ethical principles infundraising. http://www.afpnet.org/Ethic s/IntlA rticl eDeta il.cfm?ItemN umber =3681.

Bagozzi, R. P., & Moore, D. J. (2006). Public Service Advertisements: Emotions and Empathy Guide Prosocial Behavior. *Journal of Marketing*, *58*(1), 56. https://doi.org/10.2307/1252251

Bebko, C., Sciulli, L. M., & Bhagat, P. (2014). Using Eye Tracking to Assess the Impact of Advertising Appeals on Donor Behavior. *Journal of Nonprofit and Public Sector Marketing*, *26*(4), 354–371. https://doi.org/10.1080/10495142.2014.965073

Burt, C., & Strongman, K. (2005). Use of images in charity advertising: Improving donations and compliance rates. *International Journal of Organisational Behaviour*, *8*(8), 571–580. Retrieved from http://www.usq.edu.au/extrafiles/business/journals/HRMJournal/InternationalArticles/Volume 8/Burt Vol 8 no 8.pdf?origin=publication_detail

Chang, C.-T., & Lee, Y.-K. (2009). *Framing Charity Advertising: Influences of Message Framing, Image Valence, and Temporal Framing on a Charitable Appeal 1*.

Coulter, R. H., & Pinto, M. B. (1995). Guilt Appeals in Advertising: What Are Their Effects? *Journal of Applied Psychology*, *80*(6), 697–705. https://doi.org/10.1037/0021-9010.80.6.697

ELOGIA. (2019). Estudio anual de Redes Sociales IAB 2019. *IAB Spain*, *2019*, 52. Retrieved from https://iabspain.es/wp-content/uploads/estudio-redes-sociales-2019_vreducida.pdf

Grancea, I. (2015). Visual Arguments and Moral Causes in Charity Advertising: Ethical Considerations. *Nursing Ethics*, *2*(2), 167–185. https://doi.org/10.1191/0969733005ne828oa

Hastings, G., Stead, M., & Webb, J. (2004). Fear appeals in social marketing: Strategic and ethical reasons for concern. *Psychology and Marketing*, *21*(11), 961–986. https://doi.org/10.1002/mar.20043

Hibbert, S., Smith, A., Davies, A., & Ireland, F. (2007). Guilt Appeals: Persuasion Knowledge and Charitable Giving. *Psychology & Marketing*, *24*(August 2007), 723–742. https://doi.org/10.1002/mar

MacQuillin, I., & Sargeant, A. (2018). Fundraising Ethics: A Rights-Balancing Approach. *Journal of Business Ethics*, (0123456789), 1–12. https://doi.org/10.1007/s10551-018-3872-8

Nathanson, J. (2013). The Pornography of Poverty: Reframing the Discourse of International Aid's Representations of Starving Children. *Canadian Journal of Communication*, *38*(1), 103–120. https://doi.org/10.22230/cjc.2013v38n1a2587

Poels, K., & Dewitte, S. (2009). Getting a Line on Print Ads: Pleasure and Arousal Reactions Reveal an Implicit Advertising Mechanism. *Journal of Advertising*, *37*(4), 63–74. https://doi.org/10.2753/joa0091-3367370405

Quinton, S., & Fennemore, P. (2013). Missing a strategic marketing trick? The use of online social networks by UK charities. *International Journal of Nonprofit and Voluntary Sector Marketing*, *18*, 36–51. https://doi.org/10.1002/nvsm

Reed, A., Aquino, K., & Levy, E. (2007). Moral identity and judgments of charitable behaviors. *Journal of Marketing*, *71*(1), 178–193. https://doi.org/10.1509/jmkg.71.1.178

Spaulding, T. J. (2010). How can virtual communities create value for business? *Electronic Commerce Research and Applications*, *9*(1), 38–49. https://doi.org/10.1016/j.elerap.2009.07.004

ETHICAL IMPLICATIONS OF LIFE SECONDARY MARKETS

Jorge de Andrés-Sánchez, Teresa Pintado-Blanco, Sonia Carcelén-García, Mario Arias-Oliva

Universitat Rovira i Virgili (Spain), Universidad Politécnica de Madrid (Spain),
Universidad Politécnica de Madrid (Spain), Universitat Rovira i Virgili (Spain)

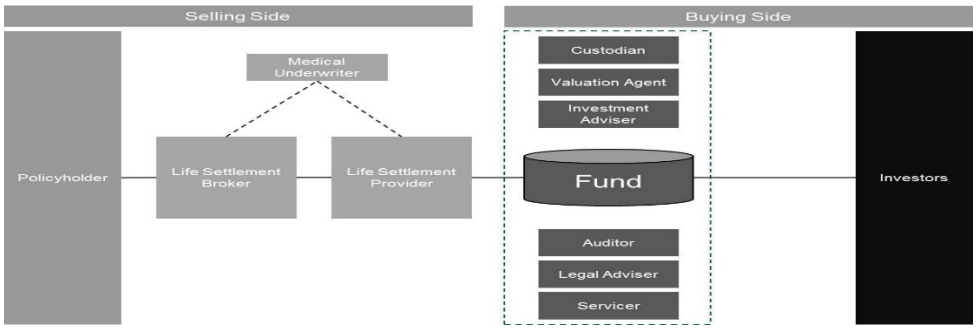
jorge.deandres@urv.cat; tpintado@ucm.es, slcarcelen@ccinf.ucm.es; mario.arias@urv.cat

EXTENDED ABSTRACT

A life settlement (LS) is an agreement in which a life insurance policyholder obtains an amount by transferring its ownership to an investor. The buyer acquires the right to obtain the benefits and the obligation to pay the outstanding premiums. The development of life settlement market begins at 80s of the last century with the so-called viatical settlements (VSs) (Giacolone, 2001). The AIDS epidemy that began in that decade, caused to people affected to make their assets liquid to cope with the costly and ineffective existing treatments. The liquidation of the life policies was carried out through the VSs which is the denomination that the LSs have when the insured is a terminal patient. McMinn and Zhu (2017) indicate that in 1990s, due to medical advances, AIDS ceases to be a terminal illness and the VS on this disease decline, starting then the growth of LSs. In the operations of LSs, the insured is not terminally ill, but usually people over 65 years who have a life expectancy below the standard according to their age and sex (Giacalone, 2001). The stimulus to the policyholder for LSs agreements instead of claiming surrender value is given by the higher price that can be obtained with the first option, since they are valued according to the actual life expectancy, which is below the average. On the other hand, the surrender value is calculated with a standard life expectancy.

Figure 1 represents the process of a LS or VS agreement. The policyholder will search for the best price for his policy through a broker. Broker must have informed policyholder on the existence of LSs agreements as an alternative to obtain the surrender value. The intermediaries of the buying party are the suppliers of LSs that acquire insurance on behalf of institutional investors. Latter insurance policies are negotiated in tertiary markets. The complexity of executing LSs implies the intervention of several agents that facilitate the development with maximum guarantees for all parties. A fundamental service is done by the medical underwriters (MU), that quantifies the life expectancy (LE) of the insured.

Figure 1. Life settlement negotiation process



Source: Braun et al. (2016)

Empirical literature outlines the importance of ethical and emotional variables (which often may have been induced by moral considerations) in economic decisions as consumption or the adoption of new technologies (Pelegrin-Borondo et al., 2018; Pelegrín-Borondo et al., 2017). Our paper conducts a conceptual analysis of several ethical and moral problems arising from LSs transactions that will allow evaluating empirically their impact on the possible development of secondary and tertiary markets in less developed insurance markets such as Europeans. In this way Kohli (2006), Blake and Harrison (2008) and Bayston (2015) advice a strict regulation of professional and practice codes for all agents that act in insurance secondary markets. Even some authors like Glac at el. (2012) point out that property right does not imply automatically the right of selling as is the case of second kidneys. So, in their opinion, life insurance policies must fall in this category.

A first ethical concert is due to LSs and VSs modifies the economic purpose of life insurance. The objective of giving an economic protection to policyholder's beloved persons turns into a bet on a death date. This issue may produce several consequences:

1. As it is pointed out by Nurnberg and Lackey (2010), a clear implication was the raising of fraudulent contracts known as STOLIS. STOLIs are new policies that were initially taken out with the financial encouragement of and loans provided by investors, with the intention of being purchased by these investors in satisfaction of the loans at the end of the contestability period, usually 2 years. Unfortunately, in practice it is not easy to demonstrate in a Court tan an STOLI is actually an STOLI.
2. Likewise, despite in 99% times the contracts in the primary market may have an insurable interest, Leimberg (2005) points out that in the secondary and tertiary market insured persons are reduced to financial assets like stocks or bonds.
3. Another consequence, as it is indicated by Gene et al. (2012), is that the definitive beneficiary of the insurance is not designed by the insured person. So the emotional and love link between policyholder and beneficiary implied in primary insurance markets is lost.
4. Braun et al. (2019) point out the possibility that the expansion of LS and VS transactions might trigger widespread increases of insurance premiums and so young families and small business with a fragile financial situations are expelled from primary insurance markets. That is why insurance companies although sometimes denies it, included lapse rates in their policy pricing. LSs market lowered policy lapse rates and reduce insurance companies' margins and so, insurers may raise premiums to balance profit levels.

The second reason to ague is that conventional financial investment generates "good wishes" in the investor. The investor in bonds or stocks is benefited by issuer's business success. On the contrary, the buyer of life insurance policies is benefited by an early death of policyholder. Of course, from the perspective of a large portfolio of LSs it can be argued that the investment is not on a single life death date but on a statistical average (Nolan and Knott, 2013). However, it is also undoubtable that good news as medical advances are always bad news for LSs buyers.

Thirdly Nurmberg and Lackey (2010) and Glac et al. (2012) point out that LS and more specifically VSs transactions in secondary market are done under a policyholder's emotional high stress situation. The originator usually has a terminal ill and needs urgently cash to pay expensive medical cares. So, it is reasonable to suppose that seller's emotional situation is not good to negotiate a fair insurance price. However, LSs and VSs transactions eliminate life insurance monopsony over secondary life insurance markets and so, it is easier for the policyholder to obtain a fair price (Doherty and Singer, 2004). Likewise, due to in the secondary transactions do not act uniquely policyholders and investors but also

other agents (see figure 1), whose income usually are a proportion of transaction prices, it is ensured the price agreed is close to the "fair price" (NordShip Association, 2016). In any case, market agents are under an interest conflict (Kohli, 2006) given that they have incentives to facilitating life insurance transactions and so sometimes might hide information to policyholders on possible better alternative sources to obtain cash as surrounding the policy by the facial value (in some exceptional situations) or borrowing founds against the policy.

The fourth consequence is that policyholder losses a great amount of privacy (Glac et al., 2012). On one hand, a great amount of medical information must be provided by the policyholder to price the policy in secondary market. Subsequently, seller's life is scrutinized until the death date. That is because the policy can be resold (and so, repriced) in the tertiary market and also, of course, because the investor wants to get insurance facial value as soon as possible. For an empirical example on this matter see Dolan (2013).

Another concern is the high moral hazard level existing in LS and VS transactions. It exists in both buyer and seller sides of the market. Agents that are interested in LSs transactions have a greater amount of knowledge and information in insurance issues than policyholders. On the other hand, policyholders have a better knowledge of their health and habits. In this way Glac et al. (2012) expose cases where the seller acted consciously against his health to obtain a better price for the policy.

Blake and Harrison (2008) expose moral objections from non-professional investors (savers) point of view. On one hand, the profit of these assets depends on mortality trends and is not correlated with conventional financial asset prices. Thus small investors may not have information and knowledge enough on LS and VS risks. Likewise, often the decision on investing in these assets is taken by collective found managers and not directly by found owners. So, the possible moral objections against this kind of investment by actual investors are avoided because of a lack of transparency. A further critical factor for investors is the reliability of the LE report since MU are also under a conflict of interests (Kohli, 2006). MUs are under the pressure to make easier LS and VS transactions by underestimating LEs. In this way, notice that several authors as Bauer et al. (2017) or Xu (2019) show that life expectancies provided by medical underwriters have been traditionally possibly underrated.

KEYWORDS: life insurance, insurance secondary markets, life settlements, viatical agreements.

REFERENCES

Bauer, D., Fasano, M. V., Russ, J., & Zhu, N. (2018). Evaluating Life Expectancy Evaluations. *North American Actuarial Journal*, *22*(2), 198-209.

Bayston, D. (2015). "Beyond the headlines: life settlements industry is ethical and regulated". In https://www.lisa.org/life-policy-owners/consumer-blog/blog

Blake, D. P., & Harrison, D. (2009). And Death Shall Have No Dominion: Life Settlements and the Ethic of Profiting from Mortality. *Available at SSRN 1344332*.

Braun, A., Affolter, S., & Schmeiser, H. (2016). Life Settlement Funds: Current Valuation Practices and Areas for Improvement. *Risk Management and Insurance Review*, *19*(2), 173-195.

Braun, A.; Cohen, L.; Malloy, C.J.; Xu, J. (2019) "Introduction to life settlement". En Xu, J. Essays on the US Life Settlement Market. University of St. Gallen.

Doherty, N. A., & Singer, H. J. (2003). The benefits of a secondary market for life insurance policies. *Real Prop. Prob. & Tr. J.*, *38*, 449.

Dolan, V. F. (2013). Advantages of a Life Expectancy Using Life Insurance Underwriting and Life Settlement Methods in the Legal Setting, https://www.experts.com/Expert-Witnesses/Epidemiologist-Expert-Vera-Dolan

Giacalone, J. A. (2001). Analyzing an emerging industry: Viatical transactions and the secondary market for life insurance policies. *Southern Business Review*, *27*(1), 1-7.

Glac, K., Skirry, J. D., & Vang, D. (2012). What Is So Morbid about Viaticals? An Examination of the Ethics of Economic Ideas and Economic Reality. *Business and Professional Ethics Journal*, *31*(3/4), 453-473.

Kohli, S. (2006). Pricing death: Analyzing the secondary market for life insurance policies and its regulatory environment. *Buff. L. Rev.*, *54*, 279.

Knott, S.; Nolan, S. (2013). Investing in Life settlements: Ethical Investment for Investors. https://lifesettlementsfund.com/blog-view/

Leimberg, S. R.: 2005, 'Stranger-Owned Life Insurance: Killing the Goose That Lays Golden Eggs!', The Insurance Tax Review 811(May), 811ff.

MacMinn, R. D., & Zhu, N. (2017). Hedging Longevity Risk in Life Settlements Using Biomedical Research-Backed Obligations. *Journal of Risk and Insurance*, *84*(S1), 439-458.

Nordship Association (2016). "When Returns are a question of Life and Death – Ethics of Life Insurance". https://nordsip.com/2016/12/08/

Nurnberg, H., & Lackey, D. P. (2010). The ethics of life insurance settlements: Investing in the lives of unrelated individuals. *Journal of business ethics*, *96*(4), 513-534.

Pelegrín-Borondo, J., Arias-Oliva, M., & Olarte-Pascual, C. (2017). Emotions, price and quality expectations in hotel services. *Journal of Vacation Marketing*, *23*(4), 322-338.

Pelegrín-Borondo, J., Arias-Oliva, M., Murata, K., & Souto-Romero, M. (2018). Does Ethical Judgment Determine the Decision to Become a Cyborg?. *Journal of Business Ethics*, 1-13.

Xu, J. (2019). "Dating Death: An Empirical Comparison of Medical Underwriters in the US Life Settlements Market." En Xu, J. Essays on the US Life Settlement Market. University of St. Gallen.

ETHICS IN ADVERTISING. THE FINE LINE BETWEEN THE ACCEPTABLE AND THE CONTROVERSIAL

Jesús García-Madariaga, Ingrit Moya Burgos, María-Francisca Blasco López, Pamela Simón Sandoval

Complutense University of Madrid (Spain)

jesgarci@ucm.es; ingritvm@ucm.es; fblasco@ucm.es; pamsimon@ucm.es

EXTENDED ABSTRACT

Mass media is known as being one of the most significant forces in modern culture. They are a major source of information for the majority of the population in most countries and have the power to shape public opinion and change people's ideas, beliefs and values (Enikolopov and Petrova, 2017). The influence of mass media on people's behaviour is embodied in two main ways. Firstly, mass media can change people's beliefs by providing relevant information, and secondly, it can have a direct effect on behaviour, through persuasion (DellaVigna and Gentzkow, 2010).

Among the different types of mass-media content, advertising plays an important role in creating individual knowledge, experiences, and values (Sheehan, 2013). From a simple point of view, advertising has been defined as a tool to give notice, to inform, to notify or to make known (Nicosia, 1974). However, in recent times the role of advertising has evolved to become an important element with probed influence on economic, political, social and cultural spheres (Adena, et al., 2015: Bond, et al., 2012; Engelberg and Parsons, 2012; Gerber, Karlan and Bergan, 2008; Gentzkow and Shapiro, 2008).

Thanks to the digital economy the advertising landscape has changed dramatically. The nature, diversity and volume of advertising are not the same as before the arrival of the internet (Tellis and Ambler, 2007). Although most of the budget of advertising is still invested on TV, a rising proportion is spent now on digital media, mobile apps, social media, games and other innovative formats able to pitch the right the consumer at the right time and/or location (IAB, 2019). These new formats clearly constitute a move away from information-based advertising, which presents facts about the product, to an evaluative conditioning format in which the product or brand is linked with rewarding stimuli (Nairn and Fine, 2008).

In this context and thanks to its versatility, advertising can perform many different functions and accomplish different objectives. However, its most important role is oriented to shape the consumer knowledge of the brands (Tellis and Ambler, 2007). Advertising works by creating favorable, strong and unique brand associations in consumers' memory that elicit positive brand judgments and feelings (Plassmann, 2007; Keller and Lehmann, 2003). However, to achieve these results, advertising needs a suitable design and execution. In particular, one of the main concerns devising an advertising strategy relates to the creative strategy since by using original, creative and different advertising strategies, companies can capture consumers' attention, develop higher brand awareness and create positive perceptions of their brands (Buil, Chernatony and Martínez, 2010).

Unfortunately, content saturation makes each time harder to think about unique advertising messages. Consequently, the search for innovative concepts and ideas has brought the use of resources such as exaggeration, deceptive statements, sexual content or stereotypical representations. All of them, considered not only unethical practices but also offensive or outrageous to consumers.

Unethical advertising is defined as those advertisements having potentially harmful effects for society (Tai, 1999) and has been widely studied from different perspectives (Drumwright and Murphy, 2009). In contrast, offensive advertising provides a more comprehensive definition since it includes messages that transgress laws and customs (e.g. human rights), breach moral or social codes (e.g. vulgarity) or outrages the moral or physical senses (e.g. gratuitous use of violence or explicit images) (Dahl et al., 2003).

In order to avoid unethical advertising, local and global institutions control the contents of the advertisements and penalty the practices against the conduct codes. However, brands not always can be penalized because their violations do not transcend the legal frame although they constitute a flagrant offense for consumers.

Those offensive advertisings become ethical controversies that are perceived differently by consumers since the offensive content is subjective, context sensible, and culture-specific (Okazaki et al., 2007). Anyway, offensive advertisements can be harmful to brands since consumers' feelings generated by controversial advertisements are transferred to their evaluation of the brand. Therefore, the disparity in the evaluation of offensive content creates a fine line between acceptable and controversial advertising that has been poorly studied until now (Okazaki et al., 2007).

Based on the aforementioned, the present research is focused on understanding how consumers perceive ethical controversies in advertising. To accomplish this purpose, we will conduct an experiment using neuromarketing techniques and declarative methodologies to analyze the reactions (conscious and unconscious) toward eight controversial ads of a sample of thirty people (15W-15M), aged between 18 and 30 years.

Three topic areas will be considered: advertising with stereotypical portrayals of men and women, advertising with stereotypical portrayals of housewives and advertising with sexually explicit. In order to find out how perceptions vary as a function of each consumer's ethical position it will be applied the Ethics Perception Questionnaire (EPQ) and to measure the unconscious reactions to ads it will be used three neuromarketing techniques: eye-tracking (ET), electroencephalogram (EEG), and galvanic skin response (GSR). Also, subjects will be asked about their attitude toward the ad and toward the brand to understand the impact of ethical controversies in advertising on those two variables.

The results obtained will provide manifold conclusions. First, they will afford evidence regarding the conscious and unconscious reactions of consumers toward ethical controversies in advertising. Second, they will enhance the understanding of the potential origins of ethical perceptions and consequently, they will shed light on why people's perceptions on controversial advertising are different, and third, they will show how ethical controversies in advertising can influence the consumers' attitudes toward the advertising and the brand.

Additionally, since unethical advertising practice can lead to a number of unwanted outcomes, ranging from consumer indifference toward the advertising, the product and/or the brand to more serious actions such as boycotts or demands for government regulation (Debbie et al., 1994), from the perspective of practitioners, the present research will help to understand where are the boundaries that define the fine line between unacceptable and controversial advertising in order to avoid surpassing it.

KEYWORDS: advertising, ethics, advertising ethics, consumer attitudes, controversial advertising, neuromarketing.

ACKNOWLEDGMENTS: This work is supported by the Spanish Ministry of Economy under grant RTC2016-4718-7. The authors also gratefully acknowledge the support of BitBrain Technologies.

REFERENCES

Adena, M., R. Enikolopov, M. Petrova, V. Santarosa and E. Zhuravskaya (2015), "Radio and the rise of Nazis in prewar Germany", Quarterly Journal of Economics, 130(4), 1885- 1939.

Bond, R. M., C. J. Fariss, J. J. Jones, A. D. Kramer, C. Marlow, J. E. Settle and J. H. Fowler (2012), "A 61-million-person experiment in social influence and political mobilization", Nature, 489, 295-298.

Buil, I., De Chernatony, L., and Martínez, E. (2013). Examining the role of advertising and sales promotions in brand equity creation. Journal of Business Research, 66(1), 115-122.

Dahl, D.W., Frankenberger, K.D. and Manchandra, R.V. (2003), "Does it pay to shock? Reactions to shocking and non-shocking advertising content among university students", Journal of Advertising Research, Vol. 43 No. 3, pp. 268-280.

DellaVigna, S. and M. Gentzkow (2010), "Persuasion: Empirical evidence", Annual Review of Economics, 2(1), 643-669.

Drumwright, M. E., and Murphy, P. E. (2009). The current state of advertising ethics: Industry and academic perspectives. Journal of Advertising, 38(1), 83-108.

Engelberg, J. E. and C. A. Parsons (2011), "The causal impact of media in financial markets", The Journal of Finance, 66(1), 67-97

Enikolopov, R. and Petrova, M. (2017). Mass media and its influence on behavior. Els Opuscles del CREI, 44, 1-45.

Gentzkow, M. and J. M. Shapiro (2008), "Preschool television viewing and adolescent test scores historical evidence from the coleman study", Quarterly Journal of Economics, 123(1), 279-323.

Gerber, A. S., D. Karlan and D. Bergan (2009), "Does the media matter? A field experiment measuring the effect of newspapers on voting behavior and political opinions", American Economic Journal: Applied Economics, 1(2), 35-52.

Keller, K.L. (1993) Conceptualizing, measuring, and managing customer-based brand equity. Journal of Marketing, 57(1), pp. 1–22.

Keller, K.L. & Lehmann, D. (2003) How do brands create value? Marketing Management, July/August, pp. 15–19.

Murphy, P. E. (1998). Ethics in advertising: review, analysis, and suggestions. Journal of Public Policy & Marketing, 17(2), 316-319.

Nicosia, F. M. (1974). Advertising, Management, and Society: a Business Point of View. Ed. McGraw-Hill. New York. USA.

Okazaki, S., Mueller, B., Chan, K., Li, L., Diehl, S., & Terlutter, R. (2007). Consumers' response to offensive advertising: a cross-cultural study. International marketing review.

Plassmann, H., Ambler, T., Braeutigam, S., & Kenning, P. (2007). What can advertisers learn from neuroscience?. International Journal of Advertising, 26(2), 151-175.

Sheehan, K. B. (2013). Controversies in contemporary advertising. Second Edition. Sage Publications. USA.

Tai, H. S. (1999). Advertising ethics: The use of sexual appeal in Chinese advertising. Teaching Business Ethics, 3(1), 87-100.

Tellis, G. Ambler, T. (2007). Handbook of Advertising. SAGE Publications. USA.

Treise, D., Weigold, M. F., Conna, J., & Garrison, H. (1994). Ethics in advertising: Ideological correlates of consumer perceptions. Journal of Advertising, 23(3), 59-69.

Waller, D. S. (1999). Attitudes towards offensive advertising: an Australian study. Journal of consumer marketing, 16(3), 288-295.

NATIVE ADVERTISING: ETHICAL ASPECTS OF KID INFLUENCERS ON YOUTUBE

Natalia Gorshkova, Lorena Robaina-Calderín, Josefa D. Martín-Santana

Universidad de Las Palmas de Gran Canaria (Spain)

natalie.go.alex@gmail.com; lorena.robaina@ulpgc.es; josefa.martin@ulpgc.es

EXTENDED ABSTRACT

Technological advances and cheaper prices in technology have made it easier for different people to have access to the digital world. This situation has provided advertisers with great opportunities, in addition to the creation of new online channels and formats. However, the quick growth of promotional content on the internet has significantly increased advertising saturation, diminishing the effectiveness of advertisements. This problem is further aggravated by the use of invasive formats (e.g. pop-ups). As a response to this issue, new advertising formats have been created, such as native advertising, which seek to improve users' browsing experience. Broadly speaking, native advertising encompasses advertising contents adapted to the environment of media and their themes (Fuente, 2018; Harms, Bijmolt and Hoekstra, 2017).

Unlike traditional advertising, native advertising has a long-term goal focused on improving branding to promote engagement and, subsequently, customer loyalty (Carlson, 2015). This new format is becoming more and more present in companies' branding communication plans. In fact, according to the study made by Oath and IAB Europe on native advertising, it is estimated that native advertising, in its different formats, will grow by 20% in 2019 (IAB, 2018). Although there are several native advertising formats, both off and online, its future is mainly associated with social media content (IAB Spain, 2018). Nowadays, social media incorporate different formats of native advertising, with branded content being one of the most popular. This modality consists in contents sponsored by brands, but edited and posted by influencers (IAB Spain, 2017; Tomas, 2018).

Branded content is mainly aimed at capturing users' interest (Costa-Sánchez, 2014). Although Instagram is the quintessential social network to develop this format, it can also often be found on YouTube (Tomas, 2018). The importance of YouTube for the study of branded content is justified by the fact that it accounts for 28% of users who are loyal to influencers (IAB Spain, 2018).

Currently, there are different business sectors that, due to their affinity to their target audience, collaborate with child influencers through their relatives or other adult companions. They provide influencers with free products aimed at a child audience in exchange for mentioning the products or demonstrating them on their channels (Martínez Pastor, 2019). Using child influencers is an extremely controversial issue from an ethical point of view because both the influencer and the target audience are highly impressionable minors. For this reason, it is necessary to pay special attention to the regulation governing this sort of activities.

The Spanish Constitution (1978, art. 12) states that every person aged below 18 years is a minor. In some cases, however, the law provides for the possibility that children over 14 years of age can carry out certain activities (Martínez Pastor, 2019). In spite of this, the great majority of influencer campaigns feature children younger than 14 who, with the consent of their family, can create and star in advertising content (Martínez Pastor, 2019). Additionally, although the content aimed at a child audience is regulated by the same national and European legislation as the rest of digital advertising contents, one of the most significant regulatory clauses requires advertising messages to be clearly

labelled as such, given the fact that minors are less able to distinguish reality from advertisements (Martínez Pastor, 2019; King Juan Carlos University and IAB, 2018). To complement the regulation, the ICC Advertising and Marketing Communications Code (2011) provides recommendations on contents targeting minors. In particular, it is recommended that (1) the message be appropriate for children and teenagers' age; (2) the advertisements do not exploit minors' "inexperience and gullibility"; and (3) the contents do not directly encourage minors to buy the advertised products. Additionally, the legal guide on child influencers written by the Universidad Rey Juan Carlos and IAB (2018) incorporates more recommendations, for instance: (1) influencers should clearly introduce themselves as promoter of a brand and (2) the characteristics of the advertised products should be shown as accurately as possible.

At present, the toy industry, which is in the middle of a transition from traditional to technological toys, believe that social media networks play a very significant role in promotional campaigns of new products aimed at a child audience (Irastorza, 2018). From a regulatory point of view, the toy industry is governed by initiatives of the International Council of Toy Industries (ICTY) and the International Chamber of Commerce (ICC). These organisations promote improving standards for toys and protect the rights of child consumers, establishing restrictions for advertising messages (Martínez Pastor, 2019). In Spain, the toy industry is also regulated by the Code of Self-regulation of Toy Advertising (Código de Autorregulación de la Publicidad de Juguetes).

One of the main focuses in the toy industry, both at global and national levels, is on the great importance of activities carried out by young influencers on YouTube (Martínez Pastor, 2019). With their parents' or legal tutors' support and consent, children actively participate in creating and presenting advertising contents for the toy industry, especially by starring in YouTube videos (Martínez Pastor, 2019).

Based on the above, this study is aimed at assessing the presence and notification (required labeling) of native advertising in the toy industry, in the shape of branded content, in YouTube channels starring child influencers. To this end, an observational study has been carried out over two months using a number of indicators to study the existence of this form of advertising in videos featured on four YouTube channels. The channels studied were chosen because (1) they are currently active and dedicated to the toy industry, (2) minors are the protagonists, (3) they have a large number of followers, and (4) new videos are posted on them almost weekly. The data provided by the indicators for each of the videos included in the study (41) have been collected by playing and watching their contents before bringing it together in an Excel sheet for later analysis and drawing of conclusions. Thirty indicators were used in the study.

Among the results of the study, it should be noted that the 41 videos show more than 70 different brands with their respective products, but in most cases the brands are not explicitly mentioned. After watching and analysing the contents of the videos, it can be concluded that the channels do use native advertising because they create content perfectly tailored to YouTube that provides subscribers with relevant information, without disrupting user experience. To be clear, the videos use the branded content format, although they are created by the channel managers (parents or tutors) and presented by the children who star in them (young minors, either alone or accompanied). It should also be highlighted that no verbal messages promoting purchase were detected, but links to the YouTube channels of the advertised brands were included in the videos. Finally, it should be mentioned that none of the 41 videos incorporates the "advertising content" notification required by the national and European regulation. This, in addition to being illegal, betrays such influencers' lack of ethics with respect to the target audience i.e. children.

Our study concludes by highlighting the need of digital platforms such as YouTube to apply stricter measures to branded content for industries targeted at children such as the toy industry.

KEYWORDS: advertising, influencers, youtube.

REFERENCES

Carlson, Matt. (2015). When news sites go native: Redefining the advertising-editorial divide in response to native advertising. *Journalism,* 16(7), 84.

Advertising and Marketing Communications Code of ICC (2011). Retrieved from https://iccwbo.org/content/uploads/sites/3/2011/09/ICC-Consolidated-Code-of-Advertising-and-Marketing-2011-Spanish.pdf.

Spanish Constitution. Boletín Oficial del Estado, Madrid, 27 de diciembre de 1987.

Costa-Sánchez, C. (2014). El cambio que viene. Branded content audiovisual*. Telos: Cuadernos de comunicación e innovación*, 99, 84-93.

Fuente, O. (2018, February 14). Native advertising: la nueva tendencia en la publicidad online. Retrieved from https://www.iebschool.com/blog/que-es-native-advertising-nativa-publicidad-online/.

Harms, B, Bijmolt, T.H.A., and Hoekstra, J.C. (2017). Digital Native Advertising: Practitioner Perspectives and a Research Agenda. *Journal of Interactive Advertising,* 17 (2), 80-91.

IAB (2018, November 15). Los publishers españoles prevén un crecimiento del 20% en formatos nativos, de vídeo y mobile en 2019. Retrieved from https://Interactive Advertising Bureauspain.es/los-publishers-espanoles-preven-un-crecimiento-del-20-en-formatos-nativos-de-video-y-mobile-en-2019/

IAB Spain (2018). Estudio Anual de Redes Sociales 2018. Retrieved from https://iabspain.es/wp-content/uploads/estudio-redes-sociales-2018_vreducida.pdf

IAB Spain (2017). Primer Estudio de Branded Content & Native Advertising. Retrieved from https://iabspain.es/wp-content/uploads/estudio-content-native-advertising-2017-vcorta.pdf

Irastorza, E. (2018). Las reglas del ʻjuegoʼ. Claves de la redefinición del sector de los juguetes. Retrieved from http://marketing.eae.es/prensa/SRC_ToysGames.pdf

Martínez Pastor, E. (2019). Menores youtubers en el ecosistema publicitario de los juguetes: límites normativos. *Revista Espacios,* 40 (7), 5-17.

Tomas, D. (2018). ¿Qué es la publicidad nativa? Ventajas y casos de éxito. Retrieved from https://www.cyberclick.es/que-es-la-publicidad-nativa-ventajas-y-casos-de-exito

Universidad Rey Juan Carlos and IAB. (2018). Guía legal sobre niños influencers. Retrieved from https://iabspain.es/wp-content/uploads/gua_legal_nios_influencers_iab_spain_2018.pdf

PREGNANCY LOSS AND UNETHICAL ALGORITHMS: ETHICAL ISSUES IN TARGETED ADVERTISING

Fatemeh Golpayegani

University College Dublin (Ireland)

Fatemeh.golpayegani@ucd.ie

EXTENDED ABSTRACT

Pregnancy loss which happens for one in four pregnancies can be a traumatic experience that impacts women's personal and social life in several ways. Studies have shown that women who have experienced pregnancy loss are at a great risk of depression and anxiety and they are more likely to have major social difficulties. This can be followed by social life distortion and isolation leaving social media and disclosure to anonymous online support groups a way to grief as well as seeking some help. However, so-called intelligent targeted advertising will follow them on every on-line media. Targeted advertisement on baby products or pregnancy clothes can target pregnant women as soon as they do a few searches on the web, or when the first pregnancy tracking app is installed. Although targeted advertising algorithms are intelligent enough to identify pregnant women in very early stages through their search history, the algorithms fail to adapt their mechanism when a pregnancy loss occurs. Despite all the search history on pregnancy loss symptom and reasons, and activities such as joining support groups with distinctive names, the algorithms continue to target these women . It will be a false claim that recognising such situation is not possible despite all the advancement in the field of social media behaviour tracking, emotion analysis, and history of the on-line activities. Extensive effort and research has been conducted on ethics in advertisement to regulate the content and reach of advertising to protect the public over the past years. This includes ethical issues in controversial advertising, matters of deception, representation, targeting of vulnerable population. However, there is no particular research on how these issues will be addressed when pregnant women are targeted and how algorithms can be designed to behave differently when they figure out that a loss has happened.

1. ETHICAL ISSUES OF TARGETED ADVERTISING ON PREGNANT WOMEN

It is common amongst pregnant women to use on-line resources, pregnancy-tracking Apps[26] and social media to access pregnancy-related information [27] [28]. Such resources remain as primary resources that women use even after a pregnancy loss or miscarriage [29]. Pregnancy is not a similar experience for all

[26] Rodger, D., A. Skuse, M. Wilmore, S. Humphreys, J. Dalton, M. Flabouris, and V. L. Clifton. "Pregnant women's use of information and communications technologies to access pregnancy-related health information in South Australia." *Australian journal of primary health* 19, no. 4 (2013): 308-312.

[27] Zhu, Chengyan, Wei Zhang, Runxi Zeng, Richard Evans, and Rongrong He. "Pregnancy-related Information Seeking and Sharing in the Social Media Era: A Qualitative Study of Expectant Mothers in China (Preprint)." (2019).

[28] Robinson, F., and C. Jones. "Women's engagement with mobile device applications in pregnancy and childbirth." *The practising midwife* 17, no. 1 (2014): 23-25.

[29] Andalibi, Nazanin, Gabriela Marcu, Tim Moesgen, Rebecca Mullin, and Andrea Forte. "Not Alone: Designing for Self-Disclosure and Social Support Exchange After Pregnancy Loss." In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, p. LBW047. ACM, 2018.

women, which can depend on many factors such as the family support, women economic and social support, or the burden of their jobs. Both physical changes and the whole unknown that a new child will bring to a woman's life sometimes makes pregnancy an overwhelming experience. The targeted advertising can make this experience even worse by featuring glowing in shape pregnant women, or pregnant moms with their kids in clean and organised houses. Such ads will start as soon as the first pregnancy tracking app is installed or after a few searches on the internet, and will follow the pregnant women recommending them Yoga classes, anti-stretchmarks body oils and towards the end of the pregnancy showing off all the beautiful nurseries and baby products. This might seem intelligent as the advertisement algorithm will keep the product recommendation updated as the pregnancy progresses, however this will be a torturous experience for a woman who has lost a pregnancy but keeps receiving all of these updated targeted ads. Many of the general ethical issues such as gender stereotypes[30], and female role portrayals[31] [32], of targeted advertising on women apply to pregnant women as well. Although the literatures studies many of such ethical issues raised by advertisement targeted at women, they lack the analysis of such advertisements targeted at pregnant women. This includes both general products and specific maternity or baby products targeted at pregnant women. The very important question to be asked is why these ads keep following women experiencing a loss, despite the fact that it is easy to identify individuals experiencing pregnancy loss or miscarriage [33], either through their social media activity or their search history. This question would not be raised in the first place if the algorithms deciding who to target and what ad to show were designed ethically.

Algorithms are designed to learn about us and structure our lives in several ways such as targeted advertisement, and determining our search results which might follow a decision that we will make[34]. Such algorithms are continuously learning about us through the available Big Data. For example, they can know people's ethnicity, gender and religion, if somebody is getting a loan, has got fired[35], the places that has already visited or going to visit, or even more personal information such as if getting married or pregnant. Although algorithms can shape our lives in many beneficial ways, their ethical implications must be discussed[36], especially for more delicate cases such as advertisements targeted at pregnant women.

Advertising algorithms are getting more accurate having access to more and more data and computational resources to process those data [37] [38]. They aim to target the best advertisement to the best audience to maximize businesses' profit. However, they raise multiple ethical issues that has to

[30] Plakoyiannaki, Emmanuella, Kalliopi Mathioudaki, Pavlos Dimitratos, and Yorgos Zotos. "Images of women in online advertisements of global products: does sexism exist?." *Journal of business ethics* 83, no. 1 (2008): 101.

[31] Tuncay Zayer, Linda, and Catherine A. Coleman. "Advertising professionals' perceptions of the impact of gender portrayals on men and women: A question of ethics?." *Journal of Advertising* 44, no. 3 (2015): 1-12.

[32] Grau, Stacy Landreth, and Yorgos C. Zotos. "Gender stereotypes in advertising: a review of current research." *International Journal of Advertising* 35, no. 5 (2016): 761-770.

[33] Pang, P. C., Meredith Temple-Smith, Clare Bellhouse, Van-Hau Trieu, Litza Kiropoulos, Helen Williams, Arri Coomarasamy, Jane Brewin, Amanda Bowles, and Jade Bilardi. "Online health seeking behaviours: what information is sought by women experiencing miscarriage." *Stud Health Technol Inform* 252 (2018): 118-125.

[34] Martin, Kirsten. "Ethical implications and accountability of algorithms." *Journal of Business Ethics* (2018): 1-16.

[35] O'neil, Cathy. *Weapons of math destruction: How big data increases inequality and threatens democracy*. Broadway Books, 2016.

[36] Mittelstadt, Brent Daniel, Patrick Allo, Mariarosaria Taddeo, Sandra Wachter, and Luciano Floridi. "The ethics of algorithms: Mapping the debate." *Big Data & Society* 3, no. 2 (2016): 2053951716679679.

[37] Thompson, Kerri A. "Commercial Clicks: Advertising Algorithms as Commercial Speech." *Vand. J. Ent. & Tech. L.* 21 (2018): 1019.

[38] Naor, Joseph Seffi, and David Wajc. "Near-optimum online ad allocation for targeted advertising." *ACM Transactions on Economics and Computation (TEAC)* 6, no. 3-4 (2018): 16.

be considered. As the literature suggests there are several types of ethical concerns that can be raised by algorithms: algorithms:

- traceability, how a result of an algorithm can be traced to its source
- transformative effects, how the original data is transformed to derive to a particular result
- unfair outcomes, an algorithm can produce discriminative or biased results.
- misguided evidence, as the produced evidence will depend on the quality of the input data.
- Inscrutable evidence as often it is not possible to understand the relation between the data point being used and the drawn conclusion when machine learning algorithms and big data is used.
- Inconclusive evidence as they are probable but uncertain knowledge that are derived from machine learning or inferential statistics.

In the following section, a number of new ethical concerns and some recommendation is offered:

- Ads' timeliness, how relevant is the outcome of the algorithm during the time, for example a pregnant woman might experience a very early miscarriage, should the ads go on for months without taking into account any further data or user activity?
- Ads' reach, who is seeing the ads? For example, Should the ads featuring active glowing pregnant women be targeted at the ones experiencing complications during their pregnancy? This also concerns about the reach of the inappropriate ads to single moms, or ones with weaker social or economical situations.

Recommendation: The targeted ads can also become a helpful tool if they can recognize cases of loss, by recommending relevant mind or body care for women experiencing a loss. This requires the algorithms to always adapt and learn from the changes and the feedback provided by the user either directly or indirectly through search histories or online activities.

KEYWORDS: Pregnancy loss, targeted advertising, ethical algorithms, miscarriage.

THE IMPACT OF ETHICS ON LOYALTY IN SOCIAL MEDIA CONSUMERS

Orlando Lima Rua, António Oliveira

Centro de Estudos Interculturais (CEI), P.PORTO/ISCAP (Portugal)

orua@iscap.ipp.pt; ajmo@iscap.ipp.pt

EXTENDED ABSTRACT

We live in the Era of connectivity, characterized by the integration and transparency of information technologies. Especially the social media that influence and enhance a market fostered by customers with demanding choices and strong competition between the traditional market and the online market. The market for selling products or services to end customers need to understand that their consumer has free access to information when connecting to the internet. For example, when the customer is chatting with the store attendant and then checking the feedbacks of real consumers, either online or in conversation with friends, or when he is looking for an item online and many ad suggestions appear in the browser with more affordable prices. Thus, for a company to build the bases for sustained competitiveness, it is necessary to take into account innovation and the dynamic learning capacity (Rua & Melo, 2015), and ethics usually are not considered in this process.

Due to a large number of options present in the market, if a company does not know how to formulate strategies that differentiate it from its competitors, win the loyalty of its customers and have reputational and relational values in its brand, it can involuntarily pass unnoticed before the choice of its consumers. Barney (1991) explains that what makes companies different from each other is the management and development of existing resources and capacities within the organization. According to Kotler, Kartajaya and Setiawan (2017, p. 87) "the number of brands that people recommend is less than the number of brands that people buy, which, in turn, is less than the number of brands that people know". Thus, while a company may be disregarded by its potential customers because its products do not have an intangible value compared to the options available in the market, it can also be "discarded" due to its reputation, built through reports of shopping experience shared by social circles and digital media.

The search for a prominent position, among the variety of offers on the market, requires companies to have an authentic personality and the development of differentiated and/or innovative business strategies within the value chain, and ethics should be a keystone of loyalty strategies. Augusto and Almeida Júnior (2015), refer that the investment in the relationship with the customers can be a possibility of differentiation before the competition. Kotler et al. (2017), state that social media is a powerful relationship channel for the connection between client and company, as it breaks barriers and allows the parties to interact as friends, to develop a relationship between them. Kotler and Keller (2012, p. 19), argue that "attracting a new customer can cost five times more than maintaining an existing one and relationship marketing emphasizes customer retention". It is important to remember the importance of retaining this client, as it is useless to stand out and not be able to have a stable client base, Reichheld and Sasser (1990), carried out a study to prove that, by reducing the desertion of clients by 5%, the profit potential of companies can increase up to 80%.

Freire, Lima and Leite (2009), refer that relationship marketing is a tool that can be used to increase the perception of the brand value and the profitability of the company over time, as well as the understanding and relationship management between this and its customers, current and potential. On the other hand, Shani and Chalasani (1992, p. 44) define relationship marketing as "an integrated

effort to identify, maintain and build a network with individual consumers and to continually strengthen the network for the mutual benefit of both sides, through interactive, individualized and value-added contacts over a long period". Scholars such as Shani and Chalasani (1992), Hennig-Thurau and Hansen (2000), Freire, Lima and Leite (2009), Kotler and Keller (2012) and Kotler et al. (2017) argue that relationship marketing powers the creation of company value, specifically in intangible resources. The aforementioned studies present characteristics and constructs of relationship marketing and intangible resources, however, the purpose of this research is to add value to the studies by developing a theoretical model in which it is possible to analyze the causal relationship between the theoretical perspective and the practice of entrepreneurs on the themes of relationship marketing, loyalty and intangible resources of the company.

Thus, this research aims to analyze the influence between relationship marketing and ethics as an intangible resource. To make this possible, a proposed model will be developed to analyze Relationship Marketing and its indicators (relationship, satisfaction, commitment, trust and ethics between the parties involved) with the dependent latent variable Intangible Resources and its indicators (relational and reputational resources), taking into account the mediator variable of Loyalty.

We intend to conduct an online survey to marketing managers from Portuguese footwear firms analyzing their social media communication instruments.

KEYWORDS: Relationship marketing, Ethics, Loyalty, Intangible resources, Social media consumers.

REFERENCES

Augusto, M., & Almeida Júnior, O. (2015). Marketing de relacionamento: a gestão do relacionamento e suas ferramentas para fidelização de clientes. *Revista de Educação, Gestão e Sociedade: Revista da Faculdade Eça de Queirós*, 5, 1–17.

Barney, J. (1991). Firm Resources and Sustained Competitive Advantage. *Journal of Management*, 17, 99- 120.

Freire, C. P., Lima, M. V. S., & Leite, B. C. (2009). Marketing de relacionamento e sua influência na conquista e mantenção de clientes. *Revista Eletrônica de Administração*, 8(15). Retrived from http://periodicos.unifacef.com.br/index.php/rea/article/view/369/355

Hennig-Thurau, T., & Hansen, U. (2000). *Relationship Marketing:* Gaining Competitive Advantage Through Custumer Satisfaction and Custumer Retencion. Berlin: Springer-Verlag.

Kotler, P., & Keller, K. L. (2012). *Administração de marketing;* tradução Sônia Midori Yamamoto; revisão técnica Edson Crescitelli (14a ed). São Paulo: Pearson Education do Brasil.

Kotler, P., Kartajaya, H., & Setiawan, I. (2017*). Marketing 4.0:* do tradicional ao digital; tradução tradução de Ivo Korytowski. Rio de Janeiro: Sextante.

Reichheld, F. F., & Sasser, J. W. (1990). Zero defections: Quality comes to services. *Harvard Business Review*, 68(5), 105-111.

Rua, O. L., & de Melo, L. F. (2016). *Estratégia, competitividade e internacionalização*. Porto: Vida Económica Editorial.

Shani, D., & S. Chalasani (1992). Exploiting niches using relationship marketing. *Journal of Service Marketing*, 6(4), 43-52.

TOURIST SHOPPING TRACKED. ETHIC REFLEXION

Alba García-Milon, Emma Juaneda-Ayensa, Jorge Pelegrín-Borondo, Cristina Olarte-Pascual, Eva Reinares-Lara

Universidad de La Rioja (Spain), Universidad de La Rioja (Spain), Universidad de La Rioja (Spain), Universidad de La Rioja (Spain), Universidad Rey Juan Carlos (Spain)

algarcm@unirioja.es; emma.juaneda@unirioja.es; jorge.pelegrin@unirioja.es; cristina.olarte@unirioja.es; eva.reinares@urjc.es

EXTENDED ABSTRACT

Shopping when in a destination is an attractive activity done by tourists (Choi, Heo, and Law, 2016; Law and Au, 2000). Indeed, one third of the expenses is dedicated to purchase goods to take home (Yu and Littrell, 2003; Yüksel and Yüksel, 2007). Tourists are a great source of income for stores and businesses, therefore, destinations and companies are more and more concerned about understanding this specific type of consumers (Jin, Moscardo, and Murphy, 2017).

Nowadays, achieving this aim is easier thanks to the increasing Information and Communication Technologies in today's digital era. Follow tourist's journey is possible from its beginning to its end, getting to know each developed activity. Specially in touristic purchases, consumers find great benefits from using technologies (García-Milon, Juaneda-Ayensa, Olarte-Pascual, and Pelegrín-Borondo, 2019), this fact could be caused fundamentally by:

- Lack of awareness. Tourists are especially fond of using technologies to overcome their lack of awareness about their shopping destinations. They want to avoid the costs of time and money of not choosing the best option available. In this context, smartphones have a privileged role as the main accessible guide.
- Necessity to share. There is a growing need to share and comment shopping experiences (Ariely and Holzwarth, 2017), above all when the store or item purchased is unusual to the tourist.
- Effort in payments. Tourists look for ways to reduce energy when handling foreign currency, in this way, cashless transactions are the most convenient strategy (Yuvaraj and Sheila Eveline, 2018).
- Tax refund. Under certain circumstances, tourists are allowed to recover the taxes from the items bought in destination (Global Blue, n.d.). This process can be easier with the use of new technologies.

New technologies have ameliorated tourists shopping experiences but, simultaneously, they force them to leave behind an electronic trail of where, what and when they buy; how much money they spend; and their perceived shopping experiences. This sort of information is collected by companies to "serve us better" but indeed, they use it to bill consumers (Quinn, 2014) in order to have the resources to make them shop more and with less consciousness.

Logically, shopping tourists' privacy is at stake. The combination of two activities where technology is widely employed, pave the way for a triple surveillance: for being a tourist, for being a shopper and for being a shopping tourist.

On the one hand, visitors' movements are followed thanks to geolocation technologies such as wearable GPS (Global Positioning Systems), smartphone applications, Bluetooth technology or social media geotagging and hashtag analytics (Hardy et al., 2017). They provide more spatial and temporal accuracy, longer tracking periods and easier data processing (Raun, Ahas, and Tiru, 2016). Knowing tourists' movements in a destination has expanded the concept of "traceability" from products to individuals (Chantre-Astaiza, Fuentes-Moraleda, Muñoz-Mazón, and Ramirez-Gonzalez, 2019). And, as a result, tourists are treated as numbers.

On the other hand, shoppers' privacy in their transactions is now in great risk. Most of the mobile payment systems collect personal information controlling all purchases. Moreover, when a shopper posts reviews on social media (Xu, Wang, Li, and Haghighi, 2017) generating UGC (user-generated content), they surpass an invisible threshold of privacy that reaches their beliefs, experiences and opinions getting inside the shopper's mind (Kumar, Kumar, and Bhasker, 2018).

Concretely for shopping tourists, if they purchase in airports, they are asked to show their boarding card. This measure is connecting the flight and personal information with the items purchased. Also, residents in a country outside the European Union can have the tax refunded from the items purchased (Global Blue, n.d.). This process, which involves a great amount of personal and shopping information, is systematized thanks to new technologies. Some countries force tourists to use digital terminals, for instance, Spain uses a system called DIVA which is compulsory since 2019 (AEAT, 2019).

Notwithstanding the foregoing, and considering that tourism involves different cultures, when a tourist visits a country, he or she has to accept destinations privacy stipulations, independent of the existing terms in his or her country of origin. Regarding tourists personal and cultural characteristics, the ethical judgement concerning privacy differs in its dimensions and in its perceived thresholds.

All in all, it is a fact that technology is changing our reality, it is impossible to escape from a global network that connects the physical and the virtual (Popescul and Georgescu, 2013). This creates a new ethical landscape that needs to have appropriate regulations in order to be able to control a proper use of Technologies in society (Tzafestas, 2018). When shopping in a destination, tourists weight the pros and cons of using technology in a utilitarian function and the vast majority prefer to lose privacy instead of losing the great advantages they provide. Recent surveys revealed that a 91% of consumers agreed that they have lost control about personal information and data (Hong, Chan, and Thong, 2019).

It could be believed that this reality is a win-win relation as both parts get benefits. Nevertheless, shopping tourists must value which is the price of their privacy. Are they a product to be traced?

KEYWORDS: Shopping tourist, technology, privacy, control, traceability, ethics.

REFERENCES

AEAT. (2019). DIVA. VAT refund in Spain.

Ariely, D., and Holzwarth, A. (2017). The choice architecture of privacy decision-making. *Health Technology*, *7*(4), 415–422. http://doi.org/10.1007/s12553-017-0193-3

Chantre-Astaiza, A., Fuentes-Moraleda, L., Muñoz-Mazón, A., and Ramirez-Gonzalez, G. (2019). Science Mapping of Tourist Mobility 1980–2019. Technological Advancements in the Collection of the Data for Tourist Traceability. *Sustainability*, *11*(17). http://doi.org/10.3390/su11174738

Choi, M. J., Heo, C. Y., and Law, R. (2016). Progress in Shopping Tourism. *Journal of Travel and Tourism Marketing*, *33*(April), S1–S24. http://doi.org/10.1080/10548408.2014.969393

García-Milon, A., Juaneda-Ayensa, E., Olarte-Pascual, C., and Pelegrín-Borondo, J. (2019). Tourist Shopping and Omnichanneling. In S. Teixeira & J. Ferreira (Eds.), *Multilevel Approach to Competitiveness in the Global Tourism Industry* (pp. 87–97). IGI Global. http://doi.org/10.4018/978-1-7998-0365-2.ch006

Global Blue. (n.d.). How to Shop Tax Free – Global Blue | Global Blue. Retrieved October 28, 2019, from https://www.globalblue.com/tax-free-shopping/how-to-shop-tax-free

Hardy, A., Hyslop, S., Booth, K., Robards, B., Aryal, J., Gretzel, U., and Eccleston, R. (2017). Tracking tourists' travel with smartphone-based GPS technology: a methodological discussion. *Information Technology and Tourism*, *17*(3), 255–274. http://doi.org/10.1007/s40558-017-0086-3

Hong, W., Chan, F. K. Y., and Thong, J. Y. L. (2019). Drivers and Inhibitors of Internet Privacy Concern: A Multidimensional Development Theory Perspective. *Journal of Business Ethics*. http://doi.org/10.1007/s10551-019-04237-1

Jin, H., Moscardo, G., and Murphy, L. (2017). Making sense of tourist shopping research: A critical review. *Tourism Management*, *62*, 120–134. http://doi.org/10.1016/j.tourman.2017.03.027

Kumar, S., Kumar, P., and Bhasker, B. (2018). Interplay between trust, information privacy concerns and behavioural intention of users on online social networks. *Behaviour and Information Technology*, *37*(6), 622–633. http://doi.org/10.1080/0144929X.2018.1470671

Law, R., and Au, N. (2000). Relationship modeling in tourism shopping: A decision rules induction approach. *Tourism Management*, *21*(3), 241–249. http://doi.org/10.1016/S0261-5177(99)00056-4

Popescul, D., and Georgescu, M. (2013). Internet of Things – Some Ethical Issues. *The USV Annals of Economics and Public Administration*, *13*(2(18)), 208–214.

Quinn, M. (2014). *Ethics for the information age*. *Pearson* (6th Editio).

Raun, J., Ahas, R., and Tiru, M. (2016). Measuring tourism destinations using mobile tracking data. *Tourism Management*, *57*, 202–212. http://doi.org/10.1016/j.tourman.2016.06.006

Tzafestas, S. (2018). Ethics and Law in the Internet of Things World. *Smart Cities*, *1*(1), 98–120. http://doi.org/10.3390/smartcities1010006

Xu, X., Wang, X., Li, Y., and Haghighi, M. (2017). Business intelligence in online customer textual reviews: Understanding consumer perceptions and influential factors. *International Journal of Information Management*, *37*(6), 673–683. http://doi.org/10.1016/j.ijinfomgt.2017.06.004

Yu, H., and Littrell, M. A. (2003). Product and process orientations to tourism shopping. *Journal of Travel Research*, *42*(2), 140–150. http://doi.org/10.1177/0047287503257493

Yüksel, A., and Yüksel, F. (2007). Shopping risk perceptions: Effects on tourists' emotions, satisfaction and expressed loyalty intentions. *Tourism Management*, *28*(3), 703–713. http://doi.org/10.1016/j.tourman.2006.04.025

Yuvaraj, S., and Sheila Eveline, N. (2018). Consumers' perception towards cashless transactions and information security in the digital economy. *International Journal of Mechanical Engineering and Technology*, *9*(7), 89–96.

9. Meeting Societal Challenges in Intelligent Communities Through Value Sensitive Design

Track chair: Oliver Burmeister, Charles Sturt University, Australia

A HOLISTIC APPLICATION OF VALUE SENSITIVE DESIGN IN BIG DATA APPLICATIONS: A CASE STUDY OF TELECOM NAMIBIA

Emilia Shikeenga, Rosa Gil, Roberto García

Universitat de Lleida (Spain)

es16@alumnes.udl.cat; rgil@diei.udl.cat; roberto.garcia@udl.cat

EXTENDED ABSTRACT

In order to encourage ethical considerations and integrity in Big Data applications that incorporate Machine Learning techniques, this paper introduces a case as to how we intend to apply Value Sensitive Design (VSD) methodology in the design of a Telecom Customer Churn Prediction model. The VSD approach identifies stakeholders throughout the design process and this assists in steering clear of any biases in the design choices that might compromise any of the stakeholders' values. In this paper, we realize a VSD conceptual investigation of a churn prediction model, including stakeholder identification and the selection of human values to be included in the design.

1. INTRODUCTION

In recent years, big data technologies have been putting some pressure with regard to what is deemed acceptable or not acceptable from an ethical point of view. A great deal of the literature that focuses on ethical issues related to big data mostly concentrates on the following values: privacy, human dignity, justice or autonomy (La Fors et al., 2019).

Telecom Namibia is facing ever-increasing competition from new entrants such as MTN, Paratus Telecom, and Capricorn Mobile. With these new entrants, all chasing the same pool of customers and declining customer spend, Telecom Namibia needs to be able to retain its customer base in order to protect its revenues and ensure growth.

According to Harvard Business Review (Gallo, 2014), it costs between 5 times and 25 times as much to find a new customer than to retain an existing one. Thus, preventing customer churn is quickly becoming an important business function.

Telecom Namibia currently has a very basic churn model in place, which simply looks at churn on the basis of how many customers have discontinued the use of telecom services but that is too late to win back the customer. Consequently, this churn model is no longer practical nor efficient.

Telecom needs to have a more robust churn model built to predict customer churn with machine learning algorithms. Ideally, telecom can nip the problem of unsatisfied customers in the bud to keep the revenue flowing and ring-fencing its customer base.

During the development of the churn model, the team will also take the opportunity to explore the customer data by determining the different personality types of each customer through accessing their personal social media profiles. This will allow the team to provide proof that it is in fact possible for companies to “use” their customers’ personal data in various ways that might be unethical.

This will therefore require that the value aspect be taken into account because the model is going to utilize data that is sensitive which might have value implications. It is for this reason why the churn model design process will employ the Value Sensitive Design approach.

Value Sensitive Design is a theoretically grounded approach to the design of technology that accounts for human values in a principled and comprehensive manner throughout the design process (Friedman et al., 2013).

Our approach will be to apply Value Sensitive Design to design a churn prediction model for Telecom Namibia. To allow us to proactively make use of the values, we will be engaging the stakeholders throughout the design process including the prototype development. The study will consider the VSD values starting from those listed by Friedman et al. (2013) and focusing on the values for big data technologies listed by La Fors et al. (2019), as shown in Table 1.

Through the implementation of this case, this paper will provide support on incorporating ethics and human values in Big Data applications. The findings will outline and demonstrate how viable the VSD approach is for providing a more comprehensive view and balancing of human values and ethics in Big Data applications.

Table 1 Integrated view of human values from different domains. Source: La Fors et al. (2019)

Technomoral values (Vallor 2016)	**Values from value-sensitive design (VSD)** (Friedman et al. 2006)	**Values from Anticipatory emerging technology ethics** (Brey 2012)	**Values in biomedical ethics** (Beauchamp & Childress 2012)	**Integration: values for big data technologies**
Care	Human Welfare	Well-being and the common good	Beneficence	Human welfare
Autonomy	Autonomy	Autonomy	Autonomy	Autonomy
Humility, self-control	Calmness	Health, (no) bodily and psychological harm	Non-maleficence	Non-maleficence
Justice	Freedom from bias; Universal usability	Justice (distributive)	Justice	Justice (incl. equality, nondiscrimination, digital inclusion)
Perspective	Accountability	N/A	N/A	Accountability (incl. transparency)
Honesty, self-control	Trust	N/A	Veracity	Trustworthiness (including honesty and underpinning also security)
N/A	Privacy; informed consent; ownership and property	Rights and freedoms, including Property	N/A	Privacy
Empathy	Identity	Human dignity	Respect for dignity	Dignity
Empathy, flexibility, courage, civility	Courtesy	N/A	N/A	Solidarity
Courage, empathy Environmental	Sustainability	(No) environmental harm, Animal welfare	N/A	Environmental welfare

2. APPROACH

Our approach draws on the Value Sensitive Design theory and involves three types of investigations: conceptual, empirical, and technical (Friedman et al., 2013, La Fors et al., 2019). As the goal of our research is to design a churn prediction model for Telecom Namibia using VSD, our research consists of the following investigations:

1. Conduct conceptual investigations to find the indirect and direct stakeholders, plus the values that are implicated. To achieve this we have to identify the different stakeholders including discovering how they are affected and the values that are implicated with regard to the implementation of the application. Applying stakeholder analysis (Friedman & Hendry, 2019) we identify:

    - Policy Makers: This includes the government Republic of Namibia, as well the regulators- Communications Regulatory Authority of Namibia (CRAN) which is mandated to regulate the telecommunication services and networks in Namibia.It also includes the Ministry of ICT.
    - Contractors: Any person or firm that undertakes a contract to provide materials or labor to perform a service or do a job for Telecom Namibia.
    - Competitors: Other companies in Namibia that offer the same products and services offered by Telecom Namibia. The competitors include MTC, MTN, Paratus Telecom, and Capricorn Mobile.
    - Shareholders: The organizations and individuals that have a stake in Telecom Namibia.
    - Customers: The people, organizations, businesses, etc. who buy and apply for the products and services that Telecom Namibia offers and makes use of those services.

2. Choosing the ethical values to consider: the values are selected according to the Telecom Namibia company values and the value considerations for techno-social change in Big Data contexts presented by La Fors et al. (2019). Telecom Namibia's company values are (Telecom Namibia, 2017): Integrity, Care, Commitment, Accountability, Empowerment, Teamwork and Mutual Respect.

    The following VSD values are the values we will consider for the particular case of the design of the churn prediction model using Big Data techniques: human welfare, ownership, and property, autonomy, calmness, universal usability, accountability, trust, privacy, identity, courtesy and sustainability (Friedman et al., 2013).

3. CONCLUSIONS AND FUTURE WORK

The increase in providing Ethical considerations in Big Data has become a concern and the values are also indicated in the ACM Principles for Algorithmic Transparency and Accountability (ACM, 2017). This paper introduced the application of VSD in telecom customer churn models construction. We have identified the direct and indirect stakeholders of Telecom Namibia and identified the associated human values. VSD has proven to be a promising approach in promoting ethical considerations in Big Data applications.

Our future work for this study is to clearly outline in detail how we applied VSD through the design process of the Telecom Namibia Churn Model. We will be researching and analyzing any laws or norms around the chosen values and we will define the design requirements.Additionally we will conduct a survey, the study will seek to prove the following hypothesis based on user experience, which is: One

of the reasons why customers churn is because of poor customer service (Retention Science, 2019). Another step will be to consider how we will verify/evaluate whether the designed model embodies the chosen human values

KEYWORDS: big data, machine learning, telecommunications, human values, value-sensitive design.

REFERENCES

ACM. (2017). Statement on Algorithmic Transparency and Accountability. Available from: https://www.acm.org/binaries/content/assets/public-policy/2017_joint_statement_algorithms.pdf

Friedman, B. & Hendry, D. G. (2019). *Value Sensitive Design: Shaping Technology with Moral Imagination*. Cambridge, MA: MIT Press.

Friedman, B., Kahn, P. H., Borning, A., & Huldtgren, A. (2013). Value sensitive design and information systems. In: *Early engagement and new technologies: Opening up the laboratory* (pp. 55-95). Springer, Dordrecht.

Gallo, A. (2014). The Value of Keeping the Right Customers. *Harvard Business Review*. Available from: https://hbr.org/2014/10/the-value-of-keeping-the-right-customers

La Fors, K., Custers, B., & Keymolen, E. (2019). Reassessing values for emerging big data technologies: integrating design-based and application-based approaches. *Ethics and Information Technology*, 1-18.

Retention Science. (2019) The Data-Driven Marketer's Guide to Predicting Customer Churn. Available from: https://go.retentionscience.com/retention-marketing

Telecom Namibia. (2017). 2016/2017 Annual Report. Available from: https://www.telecom.na/downloads/reports/2016-17/Annual%20Report%202016-2017.pdf

AUTONOMOUS SHIPPING SYSTEMS: DESIGNING FOR SAFETY, CONTROL AND RESPONSIBILITY

Xuan Wang, Luciano Cavalcante Siebert, Jeroen van den Hoven

Delft University of Technology (the Netherlands)

X.Wang-5@tudelft.nl; L.CavalcanteSiebert@tudelft.nl; M.J.vandenHoven@tudelft.nl

EXTENDED ABSTRACT

Autonomous cars and autonomous aerial vehicles are now joined by autonomous ships. Autonomous Shipping (AS) is being considered as the future of the maritime industry, and is assumed to revolutionize ship design and operations as well as the architecture and organization of ports (Briefing, 2017; den Boogaard, 2016). This development is driven mainly by considerations of cost-effectiveness (Negenborn, 2018), leading to high benefits to the maritime companies, and even to the whole of society. The labour costs on board (30% - 50% of the total ship operation costs) can be reduced (Rødseth & Burmeister, 2012). Because of reducing the crew on board, the voyage expenses can be decreased by consuming around 15% fewer fuels according to Rolls-Royce research (Rolls-Royce, 2016). AS systems are therefore high on the R&D agenda for researchers and maritime companies, such as Rolls-Royce, ManDiesel (Porathe, Burmeister, & Rødseth, 2013; H. Tvete, 2015), Konsberg Maritime, and the MUNIN research project (Benson, Sumanth, & Colling, 2018).

In addition to operational and cost efficiency, AS is also advocated on the basis of increased safety. Human errors are regarded as one of the main cause of maritime disasters (Pruitt, 2018). 75% to 96% of maritime accidents involve collisions or groundings were caused by “human errors” (Allianz, 2018). Therefore, the reduction of the role of human operators is seen as an important safety improvement. Besides, Artificial Intelligence (AI) can rapidly facilitate the development of powerful AS systems (Statheros, Howells, & Maier, 2008). Autonomous navigation systems, integrated advanced sensor systems and AI algorithm can enhance the situational awareness (H. A. Tvete, 2014) to accomplish collision avoidance control and navigation (Statheros et al., 2008). The other technologies from AI domain can supplement current technologies to achieve automatic object detection, recognition, path planning and other complex operations (Batalden, Leikanger, & Wide, 2017; Rødseth & Burmeister, 2012).

Will AS with AI technique actually deliver on safety, or is it an empty promise, at least for the first couple of decades? Is the AS system really safer than, or at least as safe as, the non-autonomous shipping system (Porathe, Hoem, Rødseth, Fjørtoft, & Johnsen, 2018)? The arguments here are analogous to the rationale for self-driving cars. However, the deadly autopilot crash of Tesla Cars (MARSHALL, 2017) and many more Self-driving car accidents gave rise to more worries about the safety of autonomous systems instead of giving us reassurance in this respect.

The main ethical -and related- concern with an autonomous system is a loss of human control, a condition in which the autonomous system with embedded AI is no longer behaving in accordance with human operator intentions (Scharre, 2016). The AS debate here mirrors the debate about Meaningful Human Control (MHC) over lethal autonomous weapon systems (LAWS) (Horowitz & Scharre, 2015), autonomous vehicles and self-driving cars (Heikoop et al., 2018). Human control should stay part of the equation of the control system. And MHC is necessary for developing safe and responsible AS systems with AI. If AS systems are not designed with human beings in or related to the

loop and under appropriate form of MHC, human responsibility and accountability will eventually evaporate. The conditions of responsibility consist of knowledge, control, freedom and choice. These conditions need to be taken into account when designing for responsibility and safety in AS systems. One possible approach to design these conditions in the context of designing for safety and responsibility in AS is Value Sensitive Design (VSD).

The VSD approach has three parts (Friedman, Kahn, & Borning, 2002), that is technical, empirical and conceptual part. The technical part provides a deeper understanding of the technology of autonomous ships. The three-stage model of the AS system (see Figure 1.) is discussed in connection to a model for types and levels of human interaction with automation (Parasuraman, Sheridan, & Wickens, 2000). Secondly, an empirical investigation exposes the underlying hazards of AS with AI in the maritime industry and the accumulated wisdom of a century of maritime accident investigations. Thirdly, the conceptual part, including the task ontology (see Figure 2.), makes explicit the key concepts of safety, control and responsibility. It indicates how an AS system should implement the idea of shared control with relevant human beings (e.g. the captain or sailors) for corrective actions, especially in critical situations. The values hierarchies (see Figure 3.) depicts the process of designing for safety and responsibility in AS systems. Based on the VSD concepts, the safety and responsibility can be translated into requirements and design specifications, which is a guideline to design AS system with AI technologies.

Designing for these and other values in AS systems in a systematic, orderly and transparent way is a necessary condition for moral acceptability in society of these maritime innovations. Designing for safety and responsibility is also an approach to control and regulate the development and application of AI in maritime industry. The aim of the paper is to introduce the VSD concept for autonomous shipping development and illustrate the designing for safety and responsibility into the requirements and design specifications for AS systems.

Figure 1. 3-stage model of AS systems

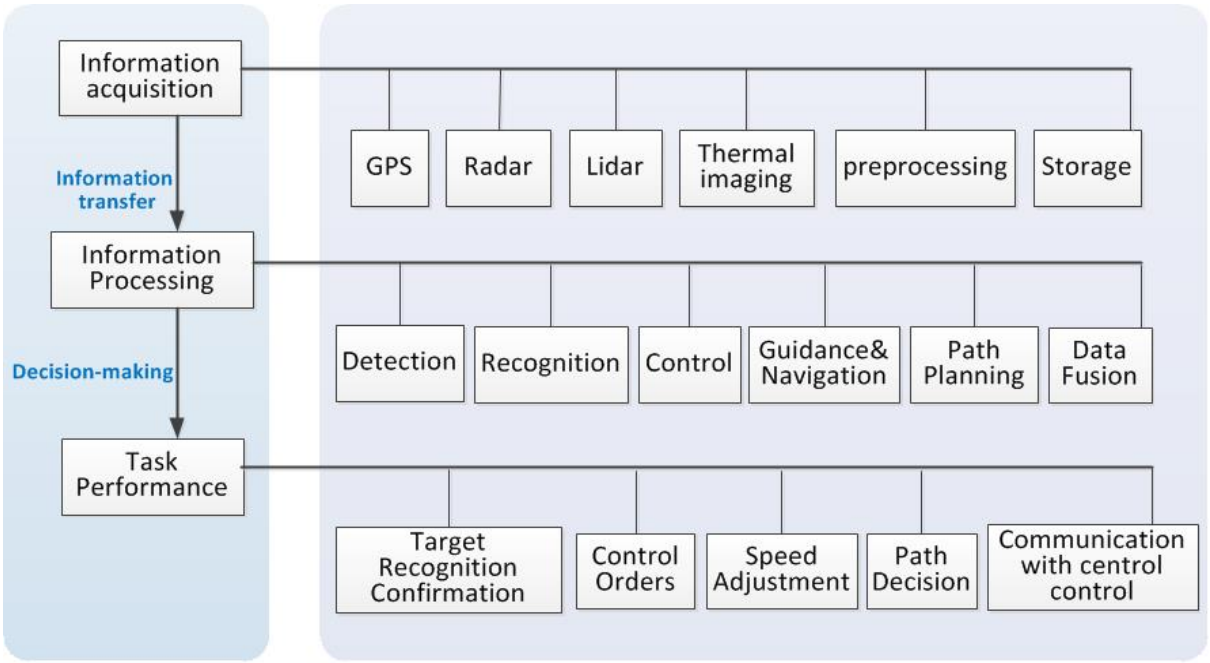


Figure 2. Task ontology of AS systems

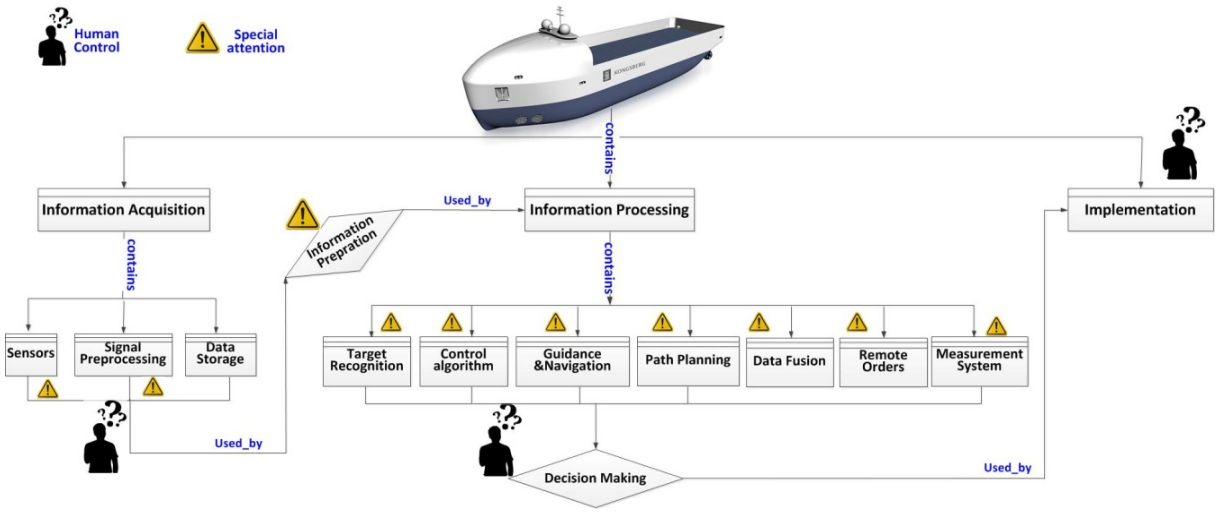
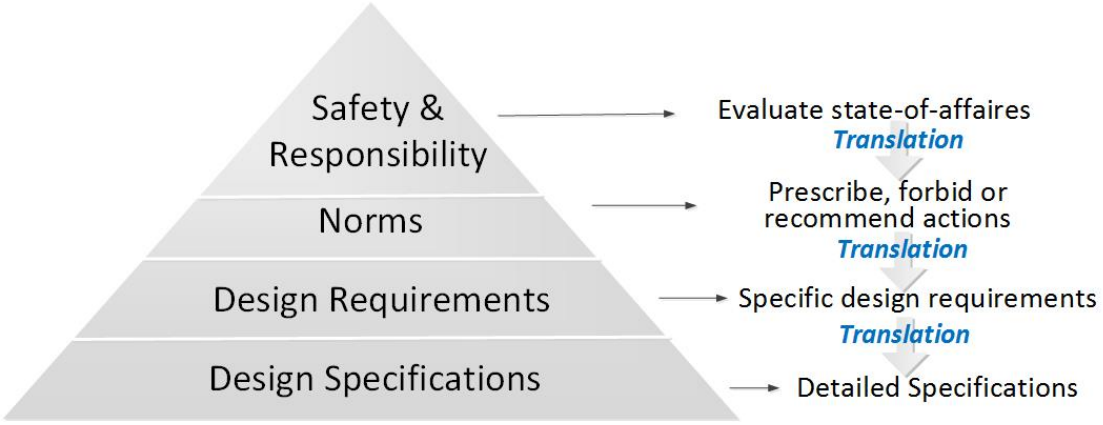


Figure 3. Extended Values Hierarchies



KEYWORDS: Autonomous shipping; Value Sensitive Design; Safety; Control; Responsibility; Task Ontology**.**

REFERENCES

Allianz. (2018). Safety considerations and regulation key to progress of autonomous vessels. Retrieved from http://www.agcs.allianz.com/insights/expert-risk-articles/safety-shipping-2017-autonomous-shipping/

Batalden, B.-M., Leikanger, P., & Wide, P. (2017). *Towards autonomous maritime operations.* Paper presented at the 2017 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA).

Benson, C. L., Sumanth, P. D., & Colling, A. P. (2018). A Quantitative Analysis of Possible Futures of Autonomous Transport. *arXiv preprint arXiv:1806.01696*.

Briefing, S. R. (2017). China's Autonomous Cargo Shipping Alliance – Digital Unmanned OBOR Maritime Deliveries. Retrieved from https://www.silkroadbriefing.com/news/2017/08/03/chinas-autonomous-cargo-shipping-alliance-digital-unmanned-obor-maritime-deliveries/

den Boogaard, M. F., Andreas & Overbeek, Mike & le Poole, Joan & Hekkenberg, Robert. (2016). *Control concepts for navigation of autonomous ships in ports*. Paper presented at the 10th symposium on high-performance marine vehicles - HIPER'16, Cortona, Italy.

Friedman, B., Kahn, P., & Borning, A. (2002). Value sensitive design: Theory and methods. *University of Washington technical report*, 02-12.

Heikoop, D. D., Hagenzieker, M., Mecacci, G., Santoni De Sio, F., Calvert, S., Heikoop, D., . . . Calvert, S. (2018). *Meaningful Human Control over Automated Driving Systems.* Paper presented at the 6th Humanist Conference.

Horowitz, M., & Scharre, P. (2015). *Meaningful Human Control in Weapon Systems: A Primer*: Center for a New American Security.

MARSHALL, A. (2017). *TESLA BEARS SOME BLAME FOR SELF-DRIVING CRASH DEATH, FEDS SAY*. Retrieved from https://www.wired.com/story/tesla-ntsb-autopilot-crash-death/

Negenborn, R. (2018). Autonomous shipping for smart logistics. Retrieved from https://www.tudelft.nl/technology-transfer/development-innovation/research-exhibition-projects/autonomous-shipping/

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on systems, man, and cybernetics-Part A: Systems and Humans, 30*(3), 286-297.

Porathe, T., Burmeister, H.-C., & Rødseth, Ø. J. (2013). *Maritime unmanned navigation through intelligence in networks: The MUNIN Project.* Paper presented at the 12th International Conference on Computer and IT Applications in the Maritime Industries, COMPIT'13, Cortona, 15-17 April 2013.

Porathe, T., Hoem, Å. S., Rødseth, Ø. J., Fjørtoft, K. E., & Johnsen, S. O. (2018). At least as safe as manned shipping? Autonomous shipping, safety and "human error". *Safety and Reliability–Safe Societies in a Changing World. Proceedings of ESREL 2018, June 17-21, 2018, Trondheim, Norway*.

Pruitt, S. (2018). Why Did the Titanic Sink? . Retrieved from https://www.history.com/news/why-did-the-titanic-sink

Rødseth, Ø. J., & Burmeister, H.-C. (2012). *Developments toward the unmanned ship.* Paper presented at the Proceedings of International Symposium Information on Ships–ISIS.

Rolls-Royce, M. (2016). Remote and Autonomous Ships - The Next Steps. Retrieved from www.rolls-royce.com/marine

Scharre, P. (2016). *Autonomous weapons and operational risk*: Center for a New American Security.

Statheros, T., Howells, G., & Maier, K. M. (2008). Autonomous ship collision avoidance navigation concepts, technologies and techniques. *The Journal of Navigation, 61*(1), 129-142.

Tvete, H. (2015). ReVolt: The Unmanned, Zero Emission, Short Sea Ship of the Future. *Dnv Gl Strategic Research & Innovation*.

Tvete, H. A. (2014). The Next Revolt. *Maritime Impact*, 18-23.

ETHICAL ENGINEERING AND ERGONOMIC STANDARDS: A PANEL ON STATUS AND IMPORTANCE FOR ACADEMIA

Sarah Spiekermann, Pieter E. Vermaas, Åke Walldius,

Vienna University of Economics and Business (Austria), Delft University of Technology (the Netherlands), Swedish Consumers' Association (Sweden)

sspieker@wu.ac.at; p.e.vermaas@tudelft.nl; aakew@kth.se

EXTENDED ABSTRACT

A background to this panel and its questions

Standards will have a large impact on how ethics and values will be handled in digital innovation projects. They are likely to be at the basis for future "ethics/privacy by design" certification programs. Therefore the coming battle for ethics is around the standards in the making.

While over 80 institutions have put forward value principles they consider important for the ethical design of IT systems (Nature, 2019), IEEE, ISO, IEC and other global standardization bodies have started concrete standardization efforts that aim to help companies to not only elicit and prioritized values, but also implement them in the design of a system. The goals in these endeavors are to a) identify the right values for a respective system context with the right priority b) ensure that these values are systematically respected in the design process and c) that positive and negative value potentials later unfolding during system deployment are monitored, iterated and controlled for.

The goal of this panel is to inform the Ethicomp participants about the state-of-the-art of the IEEE P7000 and ISO standards:

- How do these standards work?
- What is their current state of the art?
- How does their substance matter reflect the challenge of moving from ethical principles to practice?
- How do these standards reflect the achievements and works of all those who have been working in values in computing in the past two decades?
- What role did academia play in shaping these standards?
- What role should academia play once the standards are accepted?

The panel flow as envisioned by the organizers

To kick the panel off, a first short presentation will be given on the IEEE P7000 standard. IEEE has launched the so called "P7000 series". This standards series includes a baseline standard for value-based engineering, the IEEE P7000 core standard. In addition, the P700x standard series addresses grand stand-alone ethical concerns, such as, human wellbeing, system transparency, machine-readable privacy policies, avoidance of algorithmic bias, child protection, etc.

This first presentation will be followed by an overview of relevant ISO standards, including the ISO technical committee's work that is concerned with ergonomics of human-computer interaction in

Robotic, Intelligent and Autonomous systems (RIA systems) in a new technical report (ISO 9241-810); the Standards that are upcoming in the 81X series of standards; the ISO 9241-11 concerned with the definitions and concept of usability; -13 with user guidance; -110 with dialogue principles; -210 with design principles; and -220 with processes for enabling, executing and assessing human-centred design within organisations. The 9241-series is directed towards designers, developers and managers responsible for system quality within organisations. Furthermore the ISO26000 *Guidance on social responsibility* is relevant. It builds on the Brundtland report and states the benefits of social responsibility as an organization's competitive advantage. Finally, ISO 27500 *The human-centred organization — Rationale and general principles* is another standard with organizational orientation which has the potential to guide the development and practical implementation of more instrumental, compliance and sanction-oriented standards.

Both presentations will give an overview of the respective standards, the processes therein, some core terminology used and outcomes achieved by using them. In addition it is going to be shortly reported how the adoption of these standards are currently under way.

Perhaps the greatest challenge for the ISO ergonomics standards is to obtain concrete feedback from users of particular standards, especially from the standards mentioned above concerning design and organizational related human-system issues. A second challenge is to (re)introduce the concept of human values in the above-mentioned standards which would facilitate the alignment of HCI research with the standardization efforts. A third challenge is to reach out, and have an impact, to the newly formed public authorities that now have a central role in overseeing, and controlling, the digital transformation in the public sector. To introduce, and enforce, the use of ISO standards in local, regional, national and international policymaking.

After these two initial presentations on standards and their challenges, a podium discussion will ensue about the role of academia in the current and future phases in which the standards are developed and accepted. A first question is **how academia will or is already benefiting from these and other standards or is involved in their formation**. At KTH for instance ISO standards are included in the education of Human Computer Interaction. A limited number of research projects have made use of them. The same is true at Vienna University of Economics and Business. Is it desirable though that standards replace in part academic papers and textbooks? To what extent should academics get involved in their careers in standards works? Should national standards bodies actively reach out to academic institutions to incentivize researchers to take part in practical standards work?

As second question is **how these standards reflect the achievements and works of all those who have been working on values in computing in the past two decades?** Most experts involved in the ISO standards mentioned above began their work in the late 1990s and early 2000s. Here academia has played a foundational role in providing experts and research paradigms to the ISO ergonomic standards work. However, the pace of change of technology and, to a lesser degree, systems research (HCI, Information system studies etc.), has been higher than the modernization of the ISO institutions' work needed for it to sustain its role as de facto standard setter. This challenge is not only faced within ISO, but also at IEEE. There it seems that there are important benefits to be gained from a stronger focus on standards work within academia. The Value Sensitive Design research community has laid important grounds for the ethical standards that are now in the making. Their foresight has helped current ethical design standard initiatives, such as IEEE, to keep up with the pace of change and demand.

A third question is **how academia can help industry to apply the rules, and can monitor how the standards work**. Currently there are various ethical research centers in academia such as in Seattle, in Delft and at WU Vienna. Should these centers be transformed to take up larger volumes of consultancy

and executive training work? Or will industry in the future be helped in applying the standards by more focused and commercial consulting firms outside of academia? In both cases issues emerge about how to monitor the application of the standards, and about how to do scientific research on the results of these applications.

Panelists

The panelists represent both the IEEE and the ISO standards organizations and are all scholars with extensive experience in applying Value Sensitive Design in standard related research projects. The panel moderator is also a long-term member of the VSD community.

KEYWORDS: values in computing; ethics by design; ethical computing; standards.

REFERENCES

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape for AI Ethics guidelines. *Nature - Machine Intelligence*, 1, 389-399.

IEEE. (2019). P7000™ - Standard for Model Process for Addressing Ethical Concerns During System Design

G20 Ministerial Statement on Trade and Digital Economy, https://trade.ec.europa.eu/doclib/docs/2019/june/tradoc_157920.pdf

EXPLORING HOW VALUE TENSIONS COULD DEVELOP DATA ETHICS LITERACY SKILLS

Anisha TJ Fernando

Kathleen Lumley College, Adelaide (Australia) | South Australian Institute of Business & Technology, University of South Australia (Australia)

Anisha.Fernando@mymail.unisa.edu.au

EXTENDED ABSTRACT

Data ethics literacy skills aim to educate people about morally appropriate approaches concerning data, algorithms and its practices (Floridi & Taddeo 2016). Data ethics literacy skills are increasingly needed in the 21st century because potential harmful behavioural effects may be triggered by engaging with intelligent technologies. Values such as respect and trustworthiness underpin socio-technical interactions. Value tensions occur when focusing on one value triggers an impact on another value (Friedman & Hendry 2019). This extended abstract explores how value tensions (as embodied in value sensitive design) could support the development of data ethics literacy skills, and make ethical concepts and tools more accessible to people in addressing 21st century societal learning challenges.

Social interactions in the 21st century are immersed in technology. Advances in artificial intelligence, machine learning and big data analytics are changing the nature of interactions. At its core, these technologies provide communities of people, intelligence in the form of "statistical and economic notions of rationality — colloquially, the ability to make good decisions, plans, or inferences" (Russell, Dewey & Tegmark 2015 pp. 105).

Data powers these exciting new technological capabilities and industries, offering richer experiences, improved efficiencies and new sources of economic value. However, societal challenges associated with creation, collection, processing, dissemination and deletion of data remain (Solove 2009; Fernando 2017). Although old asymmetries of access and consumption of information are broken, newer information and power asymmetries are created, questioning the appropriateness of data flows (Nissenbaum 2010). Online interactions are mediated by sociotechnical affordances embedded in platforms. For example, the use of hooks in the design and development of apps as a way to monetise human attention and create potential markets for future prediction products pose significant risks to human agency and challenge healthy online social norms needed to sustain human flourishing (Eyal 2014; Centre for Humane Technology 2019; Zuboff 2019; Fernando et al. 2019).

Data ethics "studies and evaluates moral problems related to data, algorithms and corresponding practices, in order to formulate and support morally good solutions" (Floridi & Taddeo 2016 p.3). Ethical practices help people live a 'good life' and 21st century technologies are increasingly influencing the ways in which people explore how to live a 'good life' (Vallor 2018). Given the importance of healthy ethical practices for human flourishing, efforts to address these concerns from technology and policy perspectives are underway. However, if people create and use data, should they not also be engaged, educated and empowered with data ethics practices to manage these sociotechnical challenges?

While skills around how to meaningfully process and interpret data are becoming essential, the focus on ethics through data literacy initiatives is missing, and is important. Data literacy is generally defined as "the set of abilities around the use of data as part of everyday thinking and reasoning for solving

real-world problems" (Wolff et al. 2016 p. 2).. Borrowing a contemporary definition of ethics, "ethics is the process of questioning, discovering and defending our values, principles and purpose." (The Ethics Centre 2017). It is essential that an ethics lens be adopted in data literacy initiatives because data and the tools used for processing are a means to an end. It matters how people use data and the tools at hand for sense-making. For example, ethics is also about how we do things, as opposed to being about avoiding error. Visualising ethics as an impact model offers opportunities to understand contextual decision-making of data flows (Markham 2018).

From this perspective where ethics is viewed as how we do things, the values underpinning our sociotechnical interactions is an important focus of data ethics literacy efforts, because the values people expect to exercise should be embedded in the design of the technology. Given that "human values do not exist in isolation" and are central to value sensitive design (Friedman & Hendry 2019 p.44; Friedman, Kahn & Borning 2009), value tensions as a concept explicates the nuances of the different values at play in a given ethical decision-making context. One example is the disconnect between the social value of privacy and the market value around technology innovation, where the market value usually takes precedence over the social value of privacy. Another contextual example: standardised testing introduced mechanisms of measuring student and school performance in education, creating a de-facto measure of student learning, where the incentive is to perform better on tests as opposed to creating a healthy learning culture where good learning outcomes will naturally emerge (Ariely 2009).

Literacy initiatives have been historically used to address similar societal challenges. For example, children are taught about safety through school curriculums and family conversations. Information on occupational health and safety is communicated to employees. Digital security and privacy awareness is raised using data protection resources from national information privacy commissioners and via organisational training programs. Web literacy initiatives like the open web initiative by Mozilla and other privacy-by-design initiatives by community advocates create catalysts for change at the grassroots level. These literacy programs aim to make complex information accessible to people, given the contextual nature of ethical reasoning as input for decision-making.

The value tensions at stake between social and market values were conceptualised as a card game to offer people a way to unpack and understand the interactional nature of values influencing ethical decision-making regarding their data. A third category – healthy online social norms was introduced to understand scenarios which offered opportunities to exercise values in a healthy manner based on the concept of healthy data as a metaphor (Fernando et al. 2019). To commence this exercise, social values such as trustworthiness, respect and vulnerability and market values such as convenience, framing and power were surfaced from a prior research survey study sample of 441 people aged between 18 and 70 years from a diverse population of university students and staff.

Modelled on the concept of envisioning and metaphor cards which primarily cater to designers, these data ethics literacy conversation cards contribute to the VSD body of work, by offering people opportunities to explore the interactional relationship people have with their data and the values at play in specific contexts (see Figure 1 and 2 for examples) (Friedman & Hendry 2012; Logler et al. 2018). These surfaced values in the first iteration of the prototype is a work in progress, and importantly, offers participants in community workshops the opportunity to co-design the cards using their lived experiences. To date, the card game has been played with 3 different multicultural communities with people from diverse culture backgrounds and fields of interest, both in local communities and globally at Mozilla's 2019 internet health festival.

Figure 1 and 2. Examples of Value Tensions in a Data Ethics Literacy Context: Social (in Blue) and Market (in Green) Value Cards

Source: Author's own work with sources acknowledged where applicable

Future work will adapt the card game to address the needs of different community groups by co-designing cards with the participants, which could serve as input into the design of human-centred technologies. This data ethics literacy initiative offers people the ability to explore the values at play from their own perspective, and support them in understanding how they perceive the values at stake in terms of the contextual relationships they have with their data, to sustain healthy data ethics practices needed for human flourishing in the 21st century.

KEYWORDS: data ethics, data ethics literacy, value sensitive design, value tensions, value sensitive design tools, 21st century learning.

REFERENCES

Ariely, D. (2008). *Predictably Irrational: The Hidden Forces that Shape Our Decisions*, Kindle edn, HarperCollins Publishers Ltd., New York, NY.

Center for Humane Technology. (2019). *The Problem: The Extractive Attention Economy is Tearing Apart Our Shared Social Fabric*. Retrieved from https://humanetech.com/problem/

Eyal N. (2014). *Hooked: How to Build Habit-Forming Products,* Penguin, London, England.

Fernando, ATJ. (2017). *Designing Privacy Affordances for Data Release when Searching*, PhD thesis, University of South Australia, Adelaide.

Fernando, ATJ., Hall J & Scholl L. (2019 forthcoming). *Reframing the value of data: exploring healthy online social values, norms and practices*, In Proceedings of the 8th Australian Institute of Computer Ethics Conference, Melbourne.

Floridi, L., & Taddeo, M. (2016). What is data ethics? *Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences*, 374 (2083) 1-5. Retrieved from https://doi.org/10.1098/rsta.2016.0360

Friedman, B., Kahn, P. & Borning, A. (2009). [Value Sensitive Design and Information Systems]. In KE Himma & HT Tavani (Eds.), *The Handbook of Information and Computer Ethics* (pp. 69-101). Hoboken, New Jersey: John Wiley & Sons Inc.

Friedman, B., & Hendry, D. (2012). The envisioning cards: a toolkit for catalyzing humanistic and technical imaginations. In *Proceedings of the SIGCHI Conference on human factors in computing systems* (pp. 1145–1148). ACM. Retrieved from https://doi.org/10.1145/2207676.2208562

Friedman, B. & Hendry, D. G. (2019). *Value Sensitive Design: Shaping Technology with Moral Imagination*. Cambridge, MA: MIT Press.

Logler, N., Yoo, D., & Friedman, B. (2018). Metaphor Cards: A How-to-Guide for Making and Using a Generative Metaphorical Design Toolkit. In *Proceedings of the 2018 Designing Interactive Systems Conference* (pp. 1373–1386). ACM. Retrieved from

https://doi.org/10.1145/3196709.3196811

Markham, A.N. (2018). Afterword: Ethics as Impact—Moving from Error-Avoidance and Concept-Driven Models to a Future-Oriented Approach*, Social Media + Society*, 4 (3) 1-11.

Nissenbaum, H. (2010). *Privacy in Context.* 1st edn, Stanford, California: Stanford University Press.

Russell, S., Dewey, D. & Tegmark, M. (2015). Research Priorities for Robust and Beneficial Artificial Intelligence, *AI Magazine*, 36 (4) 105–114.

Solove, D. (2009). *Understanding Privacy.* Cambridge, Massachusetts: Harvard University Press.

The Ethics Centre. (2017). *Why We're Here: What is Ethics?* Retrieved from https://ethics.org.au/why-were-here/what-is-ethics/

Vallor, S. (2018). *An Introduction to Data Ethics: A Resource for Data Science Courses*. Retrieved from https://www.scu.edu/ethics/focus-areas/technology-ethics/resources/an-introduction-to-data-ethics/

Wolff, A., Gooch, D., Cavero Montaner, J., Rashid, U. & Kortuem, G. (2016). Creating an Understanding of Data Literacy for a Data-driven Society, *The Journal of Community Informatics*. 12 (3) 9–26.

Zuboff, S. (2019). *The Age of Surveillance Capitalism: The Fight for the Future at the New Frontier of Power*. New York: Public Affairs.

EXPLORING VALUE SENSITIVE DESIGN FOR BLOCKCHAIN DEVELOPMENT

Roberto García, Rosa Gil

Universitat de Lleida (Spain)

roberto.garcia@udl.cat; rgil@diei.udl.cat

EXTENDED ABSTRACT

The potential impact that blockchain technologies might have in our society makes it paramount to take into account human values during their design and development. Though the blockchain community has been moved from the beginning by a set of values that are favored by the underlying technologies, it is necessary to explore how these values play among the diverse set of stakeholders and the potential conflicts that might arise. The final aim is to motivate the establishment of a set of guidelines that make blockchains better support human values, despite the initial bias these technologies might impose.

1. INTRODUCTION

Blockchains have their roots in Bitcoin. After many attempts to create digital money, Nakamoto (2008) made a revolutionary proposal that resulted in the first cryptocurrency. The main breakthrough was that Bitcoin was completely decentralized, not requiring a central control responsible for keeping track of who owned every Bitcoin and, thus, putting too much power on it.

This is attained by implementing a distributed ledger, where all nodes participating in running the blockchain hold a copy of the ledger with all the Bitcoin transactions to date. This way, all blockchain nodes are responsible for controlling that no-one cheats, which is discouraged with an incentives system for those behaving properly, called mining rewards.

Second generation blockchains, like Ethereum (Buterin, 2014), move things one step further to create distributed ledgers that are not just capable of keeping track of currency payments, but also the transactions and current state of a shared computer. This shared computer is in fact replicated and run in every blockchain node to guarantee that it produces the same computations for everyone.

In this case, there are also application developers that can program this shared computer contributing pieces of code called smart contracts. They are contracts in the sense that it can be trusted that their code will execute as programmed. For instance, it is possible to develop an escrow payment application that does not require a trusted third party. There is guaranteed that the corresponding smart contract will make the payment if the escrow conditions are met.

Overall, the blockchain has the potential to change the ways that people and organizations trust each other, establishing a shared and tamper-proof registry of events that aims to be decentralized and neutral. This means a potential shift in money, law and government that those traditionally intermediating might perceive as a menace. Thinking even longer term, developing your own blockchain-based application you are not just making another application, it might evolve into a new form of society where humans and even machines can autonomously interact. For instance, self-driving cars that get paid and use the income to pay their energy consumption or repairs.

2. OBJECTIVES

To date, the core values that inspired blockchains design have been decentralization, transparency and neutrality. However, these values and intentions cannot be guaranteed just by the technical infrastructure alone and must be considered for each application built on top of existing blockchains. A decentralized computer network does not guarantee decentralized power, transparency does not guarantee legibility and finally, code and cryptography do not guarantee neutrality. Finally, it is important to assure that the use of blockchain technologies does not go against other values that, though no favored by the technology, should not be limited by them. For instance, transparency versus privacy.

Consequently, considering the big bias towards some specific human values, and against others conflicting, plus the enormous impact that blockchain technologies might have on our society (Tapscott & Tapscott, 2018), it is paramount to take into account human values throughout the design process of blockchains and blockchain applications.

The objective of this work is to start exploring the application of Value Sensitive Design (VSD) as a way to ensure that human values are taken into account in these cases (Friedman & Hendry, 2019; Spiekermann, 2015). VSD builds on an iterative methodology that integrates conceptual, empirical, and technical investigations, which can be aligned with the development processes of information systems.

3. BLOCKCHAIN STAKEHOLDERS

Following VSD, conceptual investigations first identify the direct and indirect stakeholders affected by the considered technology. In the case of blockchain, the identified stakeholders are:

- **Miners**: run nodes looking for rewards for those that do not try to cheat. The way of proving their commitment might involve a costly task (proof of work) or require the deposit of an economic amount as a guarantee (proof of stake), among other approaches.
- **Core Developers**: create and define the evolution of the blockchains they are involved in by contributing to its codebase. For instance, they can change the rewards that miners receive or the costs of transactions that users should satisfy.
- **Entrepreneurs**: create applications on top of blockchains that benefit from its features, especially the trust mechanisms. Trust makes it possible to develop smart contracts, pieces of code that, once deployed, guarantee their execution. These applications usually employ incentives like cryptocurrencies or tokens, which might also have economic value.
- **Investors**: buy cryptocurrencies and other tokens as an investment. They try to forecast the success of the associated blockchain or application, which might increase their demand and consequently their value.
- **Users**: employ blockchains to make cryptocurrency transactions or to use applications developed on top of blockchains, which might also include direct or indirect economic transactions but also other kinds of uses as registering agreements or voting.
- **Exchanges**: provide mechanisms to convert fiat currencies to the cryptocurrencies they have listed. Most of them are centralized and require that users move their holdings to accounts in the exchange. More recently and thanks to smart contracts, decentralized exchanges have also become available.

- **Key personalities and celebrities**: are people that have influence in a particular blockchain community, or its associate cryptocurrency. This includes outstanding developers like the creators of some blockchains or celebrities from media that advocate in favor of particular cryptocurrencies or blockchain applications (Business Insider, 2019).

4. BENEFITS, HARMS AND VALUES

Continuing with the VSD approach, we analyze the benefits and harms for each group and then map them to the corresponding values using a deductive approach: human welfare, ownership and property, privacy, universal usability, trust, autonomy, informed consent, accountability or environmental sustainability (Friedman et al., 2013). The output of stakeholders analysis plus the identified benefits, harms and values are listed in Table 1.

Table 1 Mapping Blockchain Stakeholders' Benefits and Harms to Values

Stakeholder	Benefits	Harms	Values
Miners	- Economic rewards. - Participating in the decentralization movement. - Enjoying additional privacy by interacting with the blockchain through an own node.	- Changes in rewards or costs (like electricity) might make mining not profitable. - Entry barriers, and risk of losing opportunities to earn rewards, due to the increasing investments required in computational resources or staked value because the chance of earning rewards is proportional to the commitment. - Environmental impact of mining when it is based on the intensive use of computational resources	- Human Welfare - Ownership and property - Privacy - Trust - Autonomy - Accountability - Environmental Sustainability
Core Developers	- Participating in the decentralization movement. - Public acknowledgement from the developer community, usually blockchains are open source projects to facilitate accountability and trust - Influencing the evolution of the blockchain or cryptocurrency ecosystem.	- Risk of losing the interest of miners or users that might abandon a blockchain and make it useless - Pressures from other stakeholders (including exchanges or key personalities and celebrities)	- Human Welfare - Ownership and property - Trust - Autonomy - Accountability
Entrepreneurs	- Participating in the decentralization movement. - Economic rewards from investors, including token offerings, or from users through utility tokens.	- High costs and risks of developing projects on top of a nascent technology with a lot of uncertainties - Complex technology imposes high entry barriers to potential users	- Human Welfare - Ownership and property - Universal usability - Trust - Autonomy - Accountability

Investors	- Participating in the decentralization movement, operating outside traditional and more restricted investment ecosystems - Investment returns are usually higher than other more mature markets.	- Higher risks than other more mature markets, including legal voids and potential scams	- Human Welfare - Ownership and property - Autonomy
Users	- Participating in the decentralization movement. - Economic incentives derived from cryptocurrencies and tokens earned as a reward for contributing to the application being used.	- Additional complexities introduced by an immature technology might produce economic harms - Risk of losing collected rewards if the economic volatility associated with the blockchain ecosystem makes them less valuable	- Human Welfare - Ownership and property - Privacy - Trust - Autonomy - Accountability - Informed Consent
Exchanges	- Economic profit from transaction fees. - Influencing the evolution of the cryptocurrency ecosystem, for instance choosing the currencies to be listed in the exchange. - Potential to reach a more diverse user base and reduced costs of operation	- Higher risks than other more mature markets, including legal voids and potential scams - Accumulation of value makes them very attractive to malicious hackers	- Ownership and property - Trust - Autonomy - Accountability
Key personalities and celebrities	- Participating in the decentralization movement. - Participation in investments and other economic rewards related to cryptocurrencies and tokens.	- Higher risks than other more mature markets - Potential popularity harms due to legal or other kinds of issues associated to the blockchain or application being supported	- Human Welfare - Ownership and property

5. CONCLUSIONS AND FUTURE WORK

The previous study of stakeholders and values following the VSD approach highlights potential conflicts like accountability vs. privacy or trust vs. environmental sustainability. These are trade-offs among competing values in the design, implementation, and use of blockchain-based systems. For instance, blockchain technologies due to their immutability imply serious risks for privacy. From a VSD perspective, this issue is addressed during the whole blockchain application development process so it implements measures than ensure user privacy. For instance, store personal data on chain once encrypted or just a hash of it.

Remains future work to conduce empirical investigations that help clarify the outcomes of different blockchains and applications regarding the identified values by exploring the corresponding white papers. The final target is to be able to characterize the properties and underlying mechanisms of blockchain technologies to generate a set of recommendations that make them and applications build on top of them better support human values, despite the initial bias the technology might impose.

KEYWORDS: blockchain, smart contract, human values, value sensitive design.

REFERENCES

Business Insider. (2019). 13 celebrities who back cryptocurrency and may own millions in bitcoin. Available from: https://www.businessinsider.com/13-celebrities-who-back-cryptocurrency-and-may-own-millions-in-bitcoin-2019-1

Buterin, V. (2014). A next-generation smart contract and decentralized application platform. *White Paper*, 3, 37.

Friedman, B., Kahn, P. H., Borning, A., & Huldtgren, A. (2013). Value sensitive design and information systems. In: *Early engagement and new technologies: Opening up the laboratory* (pp. 55-95). Springer, Dordrecht.

Friedman, B. & Hendry, D. G. (2019). *Value Sensitive Design: Shaping Technology with Moral Imagination*. Cambridge, MA: MIT Press.

Nakamoto, S. (2008). Bitcoin: A peer-to-peer electronic cash system.

Spiekermann, S. (2015). *Ethical IT innovation: A value-based system design approach*. Auerbach Publications.

Tapscott, D., & Tapscott, A. (2018). Blockchain revolution: how the technology behind bitcoin and other cryptocurrencies is changing the world. *Portfolio*.

ONTOLOGIES AND KNOWLEDGE GRAPHS: A NEW WAY TO REPRESENT AND COMMUNICATE VALUES IN TECHNOLOGY DESIGN

Kathrin Bednar, Till Winkler

Vienna University of Economics and Business (Austria), Vienna University of Economics and Business(Austria)

kathrin.bednar@wu.ac.at; till.winkler@wu.ac.at

EXTENDED ABSTRACT

Technology is a mediator for human values, it moulds its use context, changes the perceptions and actions of people and creates new practices (Verbeek, 2008). In this view, negative effects do not solely emerge as a result of technology use, but are triggered by the affordances inherent in technology itself (van den Hoven, 2017). A promising pathway towards an ethically aligned design of technology is to incorporate human values such as privacy or autonomy during the design process. The most prominent approach to do so is Value sensitive Design (VSD; Friedman et al. 2006).

VSD has inspired a number of related approaches and has developed 14 unique methods (Friedman, Hendry, & Borning, 2017). These methods produce a rich set of information through the identification of stakeholders, additional value sources, and the elicitation and analysis of values. However, not a single method focuses on the representation of values and the communication of value knowledge. We propose that ontology-building and the development of a knowledge base, approaches from the semantic web community, facilitate the representation and communication of values in an objective, valid, complete, and transparent way. From an engineering perspective, the translation of value knowledge into formally defined terms makes the fuzzy concept of values more tangible and can therefore help to bridge the gap between disciplines. In this paper, we develop a value ontology and a knowledge base for a product case study as a proof of concept.

The semantic web is an extension of the current web that aims at representing the semantic dimension of information in a formal, machine-readable way. Within the vision of the semantic web, ontologies play a key role in providing formally defined terms. An *ontology* is an explicit description of concepts (or classes), their properties, and restrictions, within an area of interest. In our case, the area of interest comprises human values for a specific technology context and their relations among each other as well as with the stakeholders. Ontologies can be used flexibly to define concepts and relations, but also allow the definition of constraints (e.g. a value being relevant only for the affected stakeholders is a constraint). Once an ontology is "filled" with individual instances (i.e. specific values for a specific technology context), a *knowledge base* is formed. The information contained in the knowledge base can be visually represented as graphs. A graph connects nodes (concepts) and thus represents their relation. An aggregated *knowledge graph* in turn allows the derivation of new knowledge, for example through querying the underlying information structure. Among others, building an ontology helps to share information among people, to analyze and reuse knowledge, and to make assumptions explicit (Noy & McGuinness, 2001). Building ontologies to represent values can not only improve the communication of value knowledge, but also allows building on existing knowledge and coming up with new knowledge through additional ways of value analysis.

Value knowledge should be represented and communicated in a transparent and traceable way throughout the whole design and development process. Table 1 summarizes requirements important

for a high quality method to represent and communicate value knowledge. We consider these requirements as currently underrepresented in VSD methods. In the extended version of this paper, we discuss to which degree the approach we present can live up to these requirements.

Table 1 Requirements for a high quality VSD method

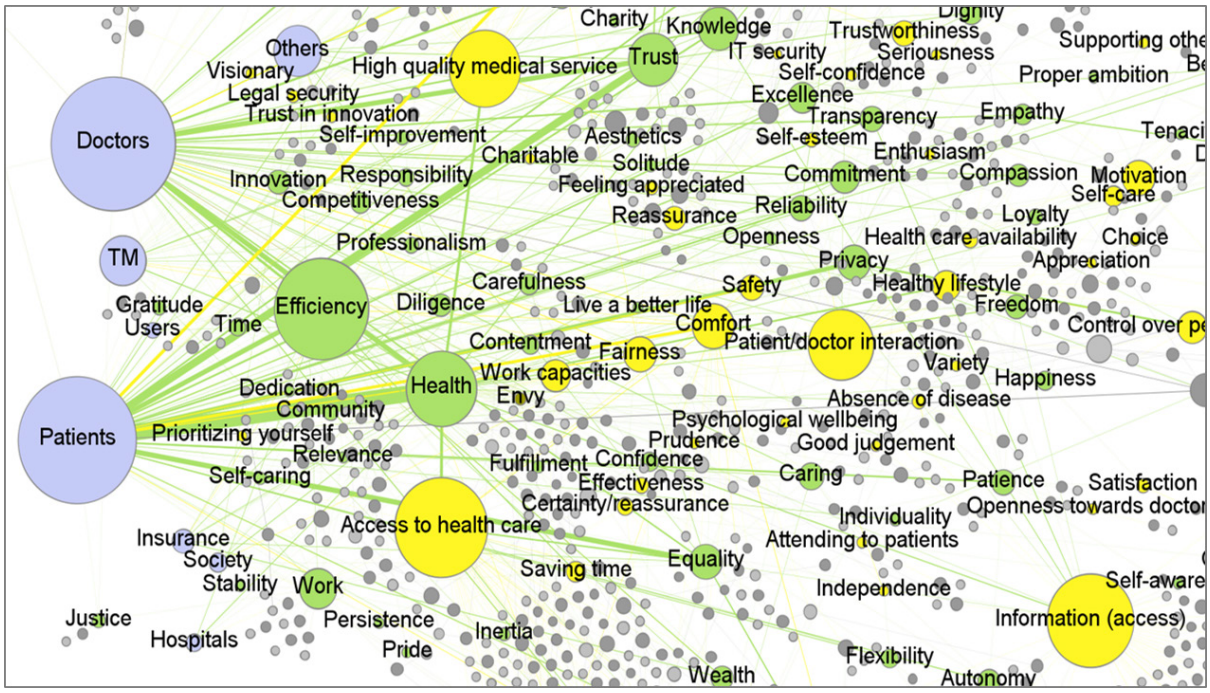
Requirement	Description
Validity	Design ideas reflect the collected value knowledge and the stakeholders' reality
Completeness	Nuances important for communicating, understanding and interpreting values are preserved (Steen & van de Poel, 2012)
Contextual integrity	Value knowledge stays connected to its relevant context
Building upon prior knowledge	Project teams explore and build on existing value knowledge
Reducing power discrepancy	The exploration of existing value knowledge by non-value-experts is supported (Borning & Muller, 2012)
Freedom from bias	The validity, reliability and value consciousness of the project is strengthened by reducing the potential for researcher bias (e.g. in the selection of data)
Transparency	The application of the method is transparent and reproducible; implicit value assumptions are made explicit; it is clear which values were selected, how, and why
Compatibility	The method is compatible with the full range of possible quantitative and qualitative methods as well as unique VSD elicitation methods
Scalability	The method can be applied for large scale VSD projects

For our case study, we chose a telemedicine system for doctors and patients and applied a value elicitation method to collect data. We want to emphasize that ontology-building and knowledge graphs can be developed for any value-oriented design approach, including VSD methods. The value elicitation was conducted by 35 students (age: M = 24.56, SD = 2.61, 38.2% female, 14 different nationalities) that formed teams of two. Within their teams, they were asked to identify values and specific value aspects related to the use context of the telemedicine system. Furthermore, they were asked to come up with design ideas based on the identified values. Participants entered their ideas in an online tool. For building an ontology based on the resulting dataset, we followed guidance provided by Noy and McGuinness (2001). We used „Gephi“ as software for the visual representation of the knowledge base. To ensure reproducibility, transparency and the completeness of information, we defined rules for data preparation, ontology building, and visualization.

Figure 1 presents a preliminary visual overview of the knowledge base for the telemedicine system. Participants identified 93 values, of which "health" was most often mentioned. Other important values were "trust", "privacy", "equality", "accuracy", "efficiency" and "truthfulness". Figure 1 shows a section of the created value knowledge base, where the color of the nodes stands for the different concepts (green: identified values, yellow: specific value aspects, grey: design ideas, blue: affected stakeholders). The size of the nodes directly represents the prominence of each (value) concept, that is, concepts that are more connected are represented by bigger nodes.

Already from this preliminary visual representation, new insights can be gained. For example, "efficiency" is more prominent than "health" because it is related to more design ideas. Also, specific value aspects such as "high quality medical service", "patient-doctor interaction" and "information access" emerge as prominent concepts connected to several values and design ideas. This shows that new knowledge can be easily be gained from aggregating value ideas of several stakeholders without having to go through a possibly biased selection and analysis process.

Figure 1. Preliminary visual representation of knowledge base



The approach we present preserves all original data, recognizes the context-specificity of values and allows tracing it back to the source. Developing a value ontology enables researchers to connect their results with other technology contexts and to explore the common value knowledge base. Overall, this can be considered as a novelty for the field of VSD, especially as it allows the representation and communication of value knowledge independent from the method used for data collection. The presented method enables additional approaches to value analysis, including visual and empirical approaches as well as methods never used before in VSD, for instance, developing knowledge patterns or counting relations (Presutti et al., 2011). Finally, the web-based and interactive nature of already available visualization tools allows the collaboration between multiple stakeholders over great distance. We hope that our preliminary insights inspire more VSD researchers to work on formally representing and sharing the value knowledge gained in their research projects.

KEYWORDS: value sensitive design, values, ontology, knowledge graphs, value communication.

REFERENCES

Borning, A., & Muller, M. (2012). Next steps for value sensitive design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'12)* (pp. 1125–1134). ACM.

Friedman, B., Hendry, D. G., & Borning, A. (2017). A survey of Value Sensitive Design methods. *Foundations and Trends® in Human–Computer Interaction*, *11*(2), 63–125.

Friedman, B., Kahn Jr., P. H., & Borning, A. (2006). Value sensitive design and information systems. In P. Zhang & D. Galletta (Eds.), *Human-computer interaction and management information systems: Foundations* (pp. 348–372). Armonk, NY: M.E.Sharpe.

Noy, N. F., & McGuinness, D. L. (2001). Ontology development 101: A guide to creating your first ontology. Retrieved from http://www.ksl.stanford.edu/people/dlm/papers/ontology101/ontology101-noy-mcguinness.html

Presutti, V., Aroyo, L., Adamou, A., Schopman, B., Gangemi, A., & Schreiber, G. (2011). Extracting core knowledge from Linked Data. In *Proceedings of the Second International Conference on Consuming Linked Data* (Vol. 782, pp. 37–48). CEUR-WS.

Steen, M., & van de Poel, I. (2012). Making values explicit during the design process. *IEEE Technology and Society Magazine*, *31*(4), 63–72.

van den Hoven, J. (2017). Ethics for the digital age: Where are the moral specs? In *Informatics in the future: Proceedings of the 11th European Computer Science Summit (ECSS 2015), Vienna, October 2015* (pp. 65–67). Cham: Springer Nature.

Verbeek, P.-P. (2008). Morality in design: Design ethics and the morality of technological artifacts. In S. A. M. Pieter E. Vermaas, Peter Kroes, Andrew Light (Ed.), *Philosophy and design: From engineering to architecture* (pp. 91–103). Springer Science + Business Media.

PHD STUDENT PERSPECTIVES ON VALUE SENSITIVE DESIGN

**Nick Logler, Naomi Jacobs, Anna Melnyk, Adam Poulsen,
Molly Balcom Raleigh, Till Winkler, Alan Borning**

University of Washington, Information School (USA), Eindhoven University of Technology (The Netherlands), Delft University of Technology, Technology, Policy and Management (The Netherlands), Charles Sturt University (Australia), Aalto University (Finland), Vienna University of Economics and Business (Austria), University of Washington (USA)

nlogler@uw.edu; n.jacobs@tue.nl; a.melnyk@tudelft.nl; apoulsen@csu.edu.au; molly.balcomraleigh@aalto.fi; Till.Winkler@wu.ac.at; borning@uw.edu

EXTENDED ABSTRACT

This panel will explore doctoral student perspectives on Value Sensitive Design (VSD). VSD has already played a central role in a number of completed dissertations (e.g., Davis 2006, Deibel 2011, Nathan 2009, Van Wynsberghe 2012, Yoo 2018), and we expect to see additional such dissertations in the future. Further, PhD students and early career scholars have played a significant role in developing value sensitive design (Davis & Nathan 2015). In this panel, the participants will reflect on their own experiences, views, and problems in doing research (dissertation or otherwise) in which VSD plays a critical role, as well as encouraging questions and comments from the audience (in particular other PhD students so engaged).

Panel Focus. The goal of this panel is to encourage conversation about the past, present, and future of how PhD students engage with VSD. The panel will help better understand the evolution and development of value sensitive design, specifically, regarding how students encounter and explore, accept and resist, and appropriate or ignore VSD. The panel will also open a conversation on how junior scholars envision the future of VSD, while informing how senior scholars might continue to catalyze an international community of VSD practitioners. Further, we seek to put forward topics and questions that ask the panelists and audience to consider tensions, limitations, and controversies in value sensitive design, and in contrast to the other approaches that address designing for human values (e.g., Value Based Design, Values in Design, Values at Play, Worth-Focused Design, etc.).

Here are examples of the kinds of questions that we plan to ask to start the discussion.

- How are you using or extending VSD in your research? This might involve simply using VSD as an established method; or alternatively, extending, critiquing, and developing VSD, for example by developing new methods, investigating the use of a different ethical frameworks for VSD investigations, or exploring the integration of VSD with other methodologies.
- There are a range of long-standing controversies in philosophy and ethics, for example, regarding whether there are (or should be) universal human values, what is an appropriate moral theory (and who decides), and so forth. Have you engaged with any of these controversies in your research? If so, how are you addressing them, and what do you see as the implications for the further development of VSD?
- There are a set of other approaches besides VSD to designing with human values at the fore, such as Value Based Design, Values in Design, Values at Play, and Worth-Focused Design. Are

you using some other approaches instead of, or in addition to, VSD? If so, how have the differences in approach impacted your work?
- What kinds of coursework have been be helpful (or perhaps, what kinds of coursework do you wish that you had available to you)? To what extent should the coursework focus on the theoretical constructs? On methods? On research or design practice?
- What kind of mentoring has been helpful? What other kinds of mentoring would you like? What, if anything, do you think is unique about being mentored with respect to value sensitive design?
- Has using VSD required a more interdisciplinary approach than is standard in your department or school? If so, how did you navigate the resulting issues?
- If applicable, give an example from your own work of framing or conducting research from a value sensitive design perspective in which you felt the approach really helped you to address the research and/or design challenge. What were the insights you gained?
- Conversely, if applicable, give an example from your own work where you felt the approach fell short. In what ways? What happened? Are there lessons for evolving VSD, or recommending other approaches in such situations?
- Value sensitive design projects often engage underrepresented, marginalized, or vulnerable stakeholder groups. Did you work with such groups in your research, or are you planning to? In what ways can working with these groups be challenging for researchers, especially doctoral students? What (if any) special training is needed?
- How do you account for your own views and values when conducting research with a VSD approach? What strategies and techniques did you use, and how well did they work?
- Regarding the theme of ETHICOMP 2020, "Paradigm Shifts in ICT Ethics: Societal Challenges in the Smart Society," what potential dissertation questions from a VSD approach would be applicable here? Are any of these ones you plan to take up?
- What advantages and opportunities arise from conducting a dissertation with a VSD approach? What risks, if any, are incurred?
- How, if at all, does doing graduate work in value sensitive design position you or other doctoral students in the job market? What are the benefits? What are the risks?
- If you were to give prospective graduate students advice about studying and/or conducting their dissertation work with VSD, what would it be and why?
- As a PhD student working with VSD, what is your position on values? Are they pluralistic? Are they entirely normative, or entirely descriptive, or somewhere in between, or both at the same time? How does your position influence how to approach and apply VSD in your research?
- Does employing a VSD approach highlight any conflicts in your PhD research that perhaps might have gone unnoticed otherwise? How do you resolve or navigate these conflict or tensions?
- Based on your experience of VSD what do you see as the most pressing open questions in VSD? How have these impacted your work? Or how have you situated your work in relation to these open questions?

Panelists. The panelists are a set of current PhD students working in centers, labs, and research groups that support a value sensitive design approach, or engaging VSD directly in their work. Panelists come from institutions in North America, Australia, and Europe, and represent disciplines including design, computer science, information science, and ethics. We have tried to assemble a diverse panel, but are nevertheless missing voices from young scholars working in South America, Asia, and Africa.

Our panelists are as follows:

Naomi Jacobs is a PhD candidate at Eindhoven University of Technology in the Netherlands. Her research focuses on values in design of technologies for health-related behavior change. She focuses specifically on the ethical problems that might arise with the design and use of these types of technologies for vulnerable people. She has written on why VSD should be complemented by an ethical theory together with Alina Huldtgren. Currently she is exploring how the Capability Approach could complement VSD. Her research is conducted at the Philosophy & Ethics group and the Human-Technology Interaction group of Eindhoven University of Technology. Previously she obtained a BA and MA in philosophy from the University of Amsterdam and Utrecht University, the Netherlands.

Nick Logler is a PhD candidate in the Information School at the University of Washington, USA, working in the Value Sensitive Design lab. His research interests are located at the intersection between design, theory, and making and building. His work explores how we understand and interact with materials in technical systems. Lately, Nick has been asking children and families to disassemble common consumer electronics, such as desktop printers, computer mice, and keyboards, in hopes of understanding how we might rethink our relationship with the materials of technical systems. Nick has also worked on generative design toolkits, longer-term thinking in information system design, and incorporating VSD into technical education.

Anna Melnyk is a PhD candidate in the Ethics & Philosophy of Technology section at TU Delft, the Netherlands, as well as a PhD representative at the 4TU Center for Ethics and Technologies. Her PhD research is a part of the ERC Advanced Grant research project "Design for Changing Values in Socio-Technical Systems." Before joining TU Delft, she obtained MSc in Philosophy of Science, Technology, and Society (University of Twente) specializing in technologies and values. In her current research, she is exploring the theoretical foundations of VSD and developing the dynamic account of values to target the potential implications of value change for low-carbon energy transition through institutional and technological design. By scrutinizing the interplay between values, stakeholders, technologies, and institutions, Anna's research aims at the consolidation of design strategies that will better deal with the value change in the energy sector.

Adam Poulsen is a computer scientist and PhD candidate at Charles Sturt University, Australia. His research interacts with several fields including human-robot interaction, eldercare robots, social robots, robot and machine ethics, VSD, care ethics, and LGBTIQ+ aged care. Broadly, Adam's primary focus is on creating adaptive, value sensitive, person-centered care robots to assist in, or enhance, the promotion of good care. Adam's research aims to further develop VSD in the care robot space, grounding design decisions care values which are 'goods' that are both normative and descriptive. In his current research, Adam is using VSD to model socially connective healthcare robots for LGBTIQ+ elders experiencing loneliness. It is his hope that such robots can be helpful in the self-care of this community, assisting in creating human-to-human connection for this group and others in the future.

Molly Balcom Raleigh is a master's student in Collaborative and Industrial Design at Aalto University, with an emphasis on strategic and service design for the public sector. Her thesis project is with Finland's Criminal Sanctions Agency (RISE), researching how values are understood and used in design processes for family visitation services in a new women's prison. She is interested in the possibility that attending to values in complex systems can help them become leverage points for transformative change, and is using VSD methods in this exploration. Before turning to design, Molly worked as an artist making participatory performance and installation, and as a communications and development consultant for non-profit arts organizations in the US and Finland.

Till Winkler is a PhD student at the Institute for Information Systems and Society, Vienna University of Economics and Business, Austria. His interests lay between several fields including value sensitive

design, requirement engineering, software development processes and sustainable development. His work focuses on developing methods and processes for value oriented software development. In his current research project, Till is comparing several requirement identification approaches in terms of their ability to deliver quality, ethical and human-centered requirements. Till is especially interested in the context of navigation applications. He is an IEEE member participating in the IEEE P7000 standardization effort for ethical engineering.

Moderator. Alan Borning is Professor Emeritus in the Paul G Allen School of Computer Science & Engineering at the University of Washington, USA, where he was a faculty member from 1980 to 2016. He was also an Adjunct Professor in the Information School. He received a B.A. from Reed College in Mathematics (1971) and a Ph.D. from Stanford University in Computer Science (1979). He has done research in and around value sensitive design for several decades, and mentored multiple PhD students working with value sensitive design in a range of domains, including accessibility, civic engagement, implantable medical devices, public transportation, and urban simulation.

Panel Structure. The 90-minute panel will be organized as follows:

1. Introduction of Panel Topic and Panelists (10 min)
2. Remarks by Individual Panelists (4 - 6 min each; 24 min total)
3. Comments and Prompts for Panelists from the Moderator (20 min)
4. Comments and Questions from the Audience (40 min)

KEYWORDS: doctoral education, research practice, value sensitive design.

REFERENCES

Davis, J. (2006). Value Sensitive Design of Interactions with UrbanSim Indicators. Ph.D. dissertation, Dept. of Computer Science & Engineering, University of Washington.

Davis, Janet & Nathan, Lisa. (2015). Value Sensitive Design: Applications, Adaptations, and Critiques. *Handbook of Ethics, Values, and Technological Design: Sources, Theory, Values and Application Domains*. 11-40. 10.1007/978-94-007-6970-0_3.

Deibel, K. (2011). Understanding and Supporting the Adoption of Assistive Technologies by Adults with Reading Disabilities. PhD dissertation, Dept. of Computer Science & Engineering, University of Washington.

Friedman, B. (1996). Value-sensitive design. interactions, *III* (6), 17–23.

Friedman, B., & Hendry, D.G. (2019). *Value Sensitive Design: Shaping technology with moral imagination*. Cambridge, MA: MIT Press.

Nathan, L. P. (2009). Ecovillages, Sustainability, and Information Tools: An Ethnography of Values, Adaptation, and Tension. Ph.D. dissertation, Information School, University of Washington.

Van Wynsberghe, A. (2012). Designing Robots with Care. Ph.D. dissertation, University of Twente.

Yoo, D. (2018). Designing with (Political) Complexity: Understanding Stakeholders, Emotion, Time, and Technology in the Case of Medical Aid-in-dying. Ph.D. dissertation, Information School, University of Washington.

STUCK IN THE MIDDLE WITH U(SERS): DOMESTIC DATA CONTROLLERS & DEMONSTRATIONS OF ACCOUNTABILITY IN SMART HOMES

Dr Lachlan Urquhart, Dr Jiahong Chen

University of Edinburgh (Scotland), Univeristy of Nottingham (England).

lachlan.urquhart@edinburgh.ac.uk; jiahong.chen@nottingham.ac.uk

EXTENDED ABSTRACT

The value of 'accountability' has a key place in designing legally compliant, trustworthy smart homes. Article 5(2) of the General Data Protection Regulation (GDPR) lays down the 'accountability principle', whereby data controllers who are handling personal data are required to demonstrate their compliance with data protection (DP) principles documented in Article 5(1). This demonstration needs to be to data subjects, data protection authorities, and beyond. These include provisions such as data minimisation, data security / integrity and ensuring lawful processing (e.g. consent or fulfilling a contract). When the accountability principle is read in conjunction with Article 24 of GDPR, we see a broader set of obligations for data controllers under GDPR. Namely, they need to put in place technical and organisational safeguards to comply with GDPR. Considering the rapid development of the Internet of Things (IoT), there is a clear and pressing need to clarify how data controllers in this sector are expected to demonstrate full compliance with GDPR under the accountability principle.

Urquhart, Lodge and Crabtree (2019) explored the accountability principle, its relationship with Article 24 GDPR, the challenges it faces in the context of governing the IoT in homes. It was concluded that, while the GDPR is "technologically neutral", it will become a driving force for certain technological developments that seek to safeguard personal data processing. With the growing interest in privacy enhancing technologies (PETS) and privacy engineering, we think there are opportunities to regulate the IoT by design. In this paper, we build on this analysis but focus on how the accountability principle might play out when DP law is increasingly stepping into the domain of the home. One of the difficulties in applying the accountability principle to the home stems from the domestic power dynamics disrupted by IoT technologies (Tolmie et al, 2002). This shift been conceptualised in recent work by Chen, Edwards, Urquhart and McAuley (2019), where they examine how Court of Justice of the European Union case law is narrowing exemptions in DP law around household processing, and broadening notions of joint data controllership in case law (Edwards, Finck, Veale and Zingales, 2019). This includes in the following cases: *Fashion ID* (Case C40/17), *Wirtschaftsakademie* (Case C-210/16), *Ryneš* (Case C-12/13) and *Jehovan todistajat* (C-25/17). This leads to DP law applying in domains it was never intended to apply in (i.e. domestic spaces). The involvement of multiple actors as well as the blurring physical and relational boundaries has rendered the social-technical landscape of a smart home remarkably different from a traditional, "non-smart" one. The complicated legal, technological and social relationships involved in a smart home requires a nuanced analysis of how responsibilities should be shared between a range of actors involved in the operation of domestic IoT devices.

Yet, it is questionable whether some regulatory assumptions underpinning the current DP regime are indeed transferrable to a smart home setting. Despite the law moving in this direction, holding those who install IoT devices to account for data processing, the reality does not map well onto these legal definitions and categories. How should parents operating an Amazon Echo enable data portability for conversation logs created with family visitors? Should flatmates running smart security cameras be contacting the data protection regulator within 72 hours if there is a data breach? Should parents need

to anonymise logs of when their children use their fob on the smart lock? What harms might emerge if the right to erasure is not instrumented properly on a smart fridge when a lodger moves out? This paper explores the legal shifts in data protection law that are seeing such questions become real practical issues that need a response.

To address this issue, we will explore the changing nature of responsibilities for DP governance in smart homes, focusing specifically on the accountability principle. The "traditional" (legally fictitious) framing of the relationship between the data controller and data subject is increasingly challengeable. The controller is not always a (commercial or public) organisation who may be setting the purposes, mechanisms and nature of data collection. Instead, in smart homes, individuals are often operating these technologies, and may find themselves in a position of being a joint data controller, or what we term, a "domestic data controller" (DDC). As the household exemption in DP law has been narrowing, and notions of controllership expand, we are left with questions around how domestic data controllers might handle data in a manner that is accountable, in accordance with Article 5(2). One example could be using personal information management systems (PIMS) as technical models that demonstrate what a demonstration of accountability to data subjects and regulators could be (Crabtree et al, 2018).

The difficulty here remains that domestic data controllers are fundamentally still users, perhaps being in a position of authority within the home or managing the device accounts (Goulden 2019). However, in practice they are unlikely to have greater resources, skills, or awareness of their responsibilities than the data subjects (who may be their children/spouse/flatmates). As such, there is a challenge here where domestic data controllers are increasingly responsible for processing under GDPR, yet they have little power to actually determine the nature of processing and to make it more legally compliant (and more broadly, ethical). They are using technologies as they come out of the box, with little scope to actually change how they function. Thus, we need to think about the IoT system design process, and how legal values like accountability can be realised in the domestic setting (Friedman et al, 2006).

Consequently, the power dynamics between IoT vendors (who will have an ongoing service relationship with these customers) and the users (who will also have domestic data controller functions) remains unresolved. On the one hand, there may be a need for more support from IoT vendors (who may also be joint data controllers) to domestic data controllers. This is reflected in Recital 78 which states: "when developing, designing, selecting and using applications, services and products that are based on the processing of personal data or process personal data to fulfil their task, *producers of the products, services and applications should be encouraged to take into account the right to data protection* when developing and designing such products, services and applications and, with due regard to the state of the art*, to make sure that controllers and processors are able to fulfil their data protection obligations*"(emphasis added). Thus, there is work for vendors to do in supporting DDCs, but similarly, reflection on what DDCs need to do themselves if creating DIY smart home systems will also be important.

Nevertheless, we do not feel it is appropriate for DP law to step into the home and making DDCs responsible. If the law is continuing in this direction, there is a pragmatic need to consider the nature of demonstrations of accountability not just from major firms such as Google or Amazon, but also between household members. We also need to consider how DDCs might demonstrate they are processing their peers' data in a compliant manner.

From a data ethics point of view, this means, at least in a smart home context, the approaches to demonstrating accountability may markedly differ from one type of data controller to another depending on their actual roles in the use of personal data. Importantly, if DDCs should assume only part of the full spectrum of responsibilities imposed by DP law – those proportionate to the level of

control they exercise over the operation of IoT devices – then the ways they are required to demonstrate compliance should also reflect such technical and legal realities.

We are particularly concerned about DP responsibilities being pushed entirely to DDCs, which appears rather absurd, given they are effectively still just users. Considering the technologies vendors' heavy involvement in determining the means and purposes of data processing, and thus their likely status as joint controllers, it seems we need to find ways to interpret the accountability principle in a way that would sufficiently mirror their overarching influence on the system.

Accordingly, in this paper, we consider the nature of shared responsibility of DDCs and IoT vendors for demonstrating accountability to data subjects. We examine gaps in current practice and consider what both parties may need to do in designing for the social realities of the smart home, balanced against legal obligations in GDPR. In particular, we reflect on practical recommendations for implementing systems that demonstrate accountability, both for IoT vendors and DDCs.

KEYWORDS: technology law, smart homes, accountability, data protection, domestic data controllers, human computer interaction.

REFERENCES

Crabtree, A et al. (2018) Building Accountability into the Internet of Things: The IoT Databox Model *Journal of Reliable Intelligent Environments.* 4(1) 39–55. Retrieved from https://link.springer.com/article/10.1007%2Fs40860-018-0054-5

Chen, J., Edwards, L. Urquhart, L., & McAuley, D. (2019) Who is Responsible for Data Processing in Smart Homes? Reconsidering Joint Controllership and the Household Exemption. *Forthcoming* Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3483511

Edwards, L. Finck, M. Veale, M. Zingales, M (2019) Data subjects as data controllers: a Fashion(able) concept? Internet Policy Review 13 Jun 2019 Retrieved from https://policyreview.info/articles/news/data-subjects-data-controllers-fashionable-concept/1400

Friedman, B., Kahn Jr., P. H. and Borning, A. (2006) 'Value sensitive design and information systems', in Zhang, P. and Galletta, D. (eds) *Human-computer interaction and management information systems: Foundations*. (Armonk, NY: M.E.Sharpe) 348–372.

Goulden, M. (2019), *'Delete the family':* platform families and the colonisation of the smart home, *Information, Communication & Society* Retrieved from https://www.tandfonline.com/doi/full/10.1080/1369118X.2019.1668454

Tolmie, P. Pycock, J, Diggins, T, MacLean, A. and Karsenty, A. (2002) Unremarkable Computing, *ACM SIGCHI 2002.* 399-406. Retrieved from https://dl.acm.org/citation.cfm?doid=503376.503448

Urquhart, L., Lodge, T., & Crabtree, A. (2019). Demonstrably doing accountability in the Internet of Things. *International Journal of Law and Information Technology*. 27(1). 1–27 Retrieved from https://doi.org/10.1093/ijlit/eay015

TEACHING VALUES IN DESIGN IN HIGHER EDUCATION TOWARDS A CURRICULUM COMPASS

Wolmet Barendregt, Elisabet M. Nilsson, Daisy Yoo, Rikke Toft Nørgård, Tilde Bekker, Annemiek Veldhuis, Eva Eriksson

Eindhoven University of Technology (The Netherlands), Malmö University (Sweden), Aarhus University (Denmark), Aarhus University (Denmark), Eindhoven University of Technology (The Netherlands), Eindhoven University of Technology (The Netherlands), Aarhus University (Denmark)

wolmet.barendregt@gu.se; elisabet.nilsson@mau.se; dyoo@cc.au.dk; rtoft@tdm.au.dk; m.m.bekker@tue.nl; a.h.m.veldhuis@tue.nl, capo@ucn.dk, evae@cc.au.dk

EXTENDED ABSTRACT

In this paper, we discuss some of the fundamental questions and pillars we need to consider when teaching design and engineering students about values in design. First, we review existing approaches to addressing values in design. Next, we suggest three main pillars identified relevant for teaching values in design – 1) Ethics and Human Values, 2) People and Stakeholders, and 3) Technology and Context. We describe each pillar and discuss why and how these pillars are important to consider when teaching students about values. Finally, building on these three pillars, we aim to further structure how a learner's understanding of values develops from a simple to a more complex level, using established taxonomies of learning. Eventually, the teaching activities should allow the learner to engage in the curriculum through 'coming to know', 'becoming able to act' and 'obtaining an identity' as a caring designer.

Scholars in a wide range of disciplines have acknowledged that values are embedded in technical systems and devices, consciously or unconsciously (Knobel & Bowker, 2011). Knowles and Davis (2016) therefore also argue that "technology affects values regardless of whether the designer has any explicit intention to do so" (p. 62), meaning that designers and developers have the possibilities to make powerful changes in society through the design of digital technologies, which may both embed and affect values. In line with this, almost two decades ago, Suchman (2002) criticized the design education that designers are encouraged in their training to be ignorant of their own positions within the social relations that comprise technical systems, to view themselves as creators of technological objects. However, designers are always biased by a particular way of seeing the world and by their sociocultural backgrounds (Haraway, 1988). Design never derives from nowhere, and the designers are never value-neutral (Søndergaard & Kofoed, 2017; Suchman, 2002; Verbeek, 2011). Design and engineering professionals thus play an important role in shaping society, but without always being explicitly aware of this. Consider Facebook's CEO Mark Zuckerberg, for example, who brought a technology to life without being critically aware of the major societal consequences of its use. Teachers in higher education who care about tomorrow's society thus have an obligation to make students aware of their responsibility and impact as future practicing designers.

In what follows we discuss the approach we are taking in the Value Sensitive Design in Higher Education (VASE) project. VASE is a cross-European project aimed at developing educational resources for teaching students about the role values play in design, as well as for providing students the knowledge and skills to become responsible and caring designers of future technologies.

Many scholars have thoughtfully considered values in design, so the first step in the VASE project was to review the different approaches taken, including value sensitive design (Friedman, 1996; Friedman & Hendry, 2019), values in design (Nissenbaum, 2005), values at play (Belman et al., 2009), values-led participatory design (Iversen et al., 2012a, 2012b), and worth-centred design (Cockton, 2006). Each approach provides a different lens, for example, whether they focus on *values* of moral import or *worth* writ large, whether they focus more on values in the design *process* or on values in the designed *product*, and whether they focus more on *designers'* values or *stakeholders'* values. All these approaches can provide theoretical directions and fruitful resources for teaching about values in design. However, most of these approaches has been developed for research purposes rather than for teaching purposes. The primary challenge for the VASE project is thus to consider how we may leverage such research techniques (e.g., design methods and tools) to develop effective teaching materials and activities (e.g., teaching patterns and curriculums). To do so, we also strive to leverage existing design education tools, such as the Ethics for Designers Toolkit (Gispen, 2017) and the Product Impact Tool (Dorrestein, 2019), and available online lectures, such as the TED Talk by Sebastian Deterding on "What your designs say about you" (2011).

Given that there are so many potential resources out there, we need to carefully select and present materials and activities in such a way that it can be easily accessed and used by teachers working across multiple disciplines (e.g. industrial design, computer science, educational technology), engaging with students on different levels (e.g. bachelor and master), and dealing with different sets of constraints (e.g., time, location, person power, budget). Currently, we are working on the creation of a *curriculum compass*, a structural guidance that can help organize teaching activities together with relevant materials and tools, by employing educational design patterns as development framework (Goodyear, 2005; Mor & Winthers, 2008). For this structure, we have identified three main pillars for teaching about values in design: 1) Ethics and Human Values, 2) People and Stakeholders, and 3) Technology and Context. Building on these three pillars, we aim to further structure how a learner's understanding of values develops from a simple to more complex level. To do so, we are drawing from established taxonomies of learning, such as the SOLO taxonomy (Biggs & Collis, 1982) and the Bloom taxonomy (Bloom, 1956) to address different levels of competences. Finally, our overarching goal is to make sure that our students become caring and responsible designers of the future society in a holistic and grounded manner. To this end, our project not only focuses on developing conceptual knowledge about values and ethics and gaining practical skills to design in a value-sensitive way, but more importantly, on becoming a reflective and responsible designer. Therefore, the structure encompasses teaching activities that allow the learner to engage in the curriculum through 'coming to know', 'becoming able to act' and 'obtaining an identity' (Barnett & Coate, 2005; Barnett, 2009) as a caring designer.

KEYWORDS: values in design, teaching, higher education.

REFERENCES

Barnett, Ronald (2009). Knowing and becoming in the higher education curriculum, *Studies in Higher Education*, 34:4, 429-440.

Barnett, Ronald; and Coate, Kelly (2005). *Engaging the Curriculum in Higher Education*. Berkshire: Open University Press.

Belman, Jonathan; Flanagan, Mary; and Nissenbaum, Helen (2009). Instructional Methods and Curricula for Values Conscious Design. *Loading: The Official Journal of the Canadian Game Studies Association*, 3(4).

Biggs, John B.; and Collis, Kevin F. (1982). *Evaluating the Quality of Learning: the SOLO taxonomy*. New York: Academic Press.

Bloom, Benjamin S. (red.) (1956). *Taxonomy of educational objectives: the classification of educational goals. Handbook 1, Cognitive domain*. New York: David McKay.

Cockton, Gilbert (2006). Designing worth is worth designing. Conference proceedings of *the 4th Nordic conference on Human-computer interaction: changing roles*, Oslo, Norway.

Deterding, Sebastian (2011). What your designs say about you. TED talk. Retrieved 2019 Dec. 12 https://www.ted.com/talks/sebastian_deterding_what_your_designs_say_about_you?language=en

Dorrestein, Steven. Product impact tool. Retrieved 2019 Dec. 12 from https://productimpacttool.org/nl/portal/,

Gispen, Jet (2017). Ethics for designers. Retrieved 2019 Dec. 12 from https://www.ethicsfordesigners.com/

Goodyear, Peter (2005). Educational design and networked learning: Patterns, pattern languages and design practice. *Australasian Journal of Educational Technology*, 21(1).

Friedman, Batya (1996). Value-sensitive design. *interactions, III*(6), 17–23.

Friedman, Batya; and Hendry, David G. (2019). *Value Sensitive Design: Shaping technology with moral imagination*. Cambridge, MA: MIT Press.

Haraway, Donna (1988). Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective. *Feminist Studies*, 14(3), 575–599.

Iversen, Ole Sejer; Halskov, Kim; and Leong, Tuck Wah (2012a). Values-led participatory design. *CoDesign: International Journal of CoCreation in Design and the Arts, 8*(2-3), 87–103.

Iversen, Ole Sejer; and Leong, Tuck Wah (2012b). *Values-led participatory design: mediating the emergence of values.* Conference proceedings resented of the 7th Nordic Conference on Human-Computer Interaction: Making Sense Through Design.

Knobel, Cory; and Bowker, Geof (2011). Values in Design. *Communications of the ACM,* 54(7), 26–28.

Knowles, Bran; and Davis, Janet (2017). Is Sustainability a Special Case for Persuasion?. *Interacting with* Computers, 29(1), 58–70.

Mor, Yishay; and Winters, Niall (2008). Participatory design in open education: a workshop model for developing a pattern language. *Journal of Interactive Media in Education, 1(12).*

Nissenbaum, Helen (2005). Values in technical design. In Carl Mictham (Ed.), *Encyclopedia of science, technology, and ethics* (p. 66–70). New Work: Macmillian.

Suchman, Lucy (2002). Located accountabilities in technology production. *Scandinavian Journal of Information Systems*, 14(2), 91–105.

Søndergaard, Marie Louise Juul; and Koefoed Hansen, Lone (2017). Designing with Bias and Privilege? *Nordes 2017*, 1–8.

Verbeek, Peter-Paul (2011). *Moralizing Technology: Understanding and Designing the Morality of Things.* Chicago/London: The University of Chicago Press

THE FUTURE OF VALUE SENSITIVE DESIGN

Batya Friedman*, David G. Hendry, Steven Umbrello, Jeroen van den Hoven, Daisy Yoo

University of Washington (USA), University of Washington (USA), Institute for Ethics and Emerging Technologies (Italy); TU Delft (The Netherlands); Aarhus University (Denmark)

batya@uw.edu; dhendry@uw.edu; steve@ieet.org; M.J.vandenHoven@tudelft.nl; dyoo@cc.au.dk

EXTENDED ABSTRACT

In this panel, we explore the future of value sensitive design (VSD). The stakes are high. Many in public and private sectors and in civil society are gradually realizing that taking our values seriously implies that we have to ensure that values effectively inform the design of technology which, in turn, shapes people's lives. Value sensitive design offers a highly developed set of theory, tools, and methods to systematically do so.

In short, value sensitive design is an approach for foregrounding human values in the technical design process (Friedman and Hendry, 2019; van den Hoven, 2013). First developed in human-computer interaction (HCI), value sensitive design has now been applied in a wide range of computing and related fields including artificial intelligence (Umbrello and De Bellis, 2018), biomedical and health informatics (Mueller and Heger, 2018), civilian drones (Cawthorne and Cenci, 2019), computer security (Denning, et al., 2010), computer supported cooperative work (Harbers and Neerincx, 2017), data science (Winkler and Spiekermann, 2019), multi-lifespan design (Friedman and Nathan, 2010; Yoo et al., 2016) nanotechnology (Timmermans et al., 2011; Umbrello, 2019), natural language processing (Bender and Friedman, 2018), participatory design (Friedman and Hendry, 2012; Yoo, Huldtgren, Woelfer, and Friedman, 2013), and robotics (Santoni de Sio and van den Hoven, 2018; Cheon and Su, 2018; van Wynsberghe, 2013) to name a few.

Since its inception in the early 90s (Friedman, 1996), value sensitive design has continued to expand, develop and adapt as new work and issues have emerged. Notably, in 2012 Borning and Mueller (Borning and Mueller, 2012) proposed four topics for next steps in the evolution of value sensitive design, including (1) adopting a pluralistic position on values; (2) contextualizing lists of values that are presented as heuristics for consideration; (3) strengthening the voice of the participants in publications describing VSD investigations; and (4) making clearer the voice of the researchers themselves writing about VSD investigations. Many of those have now been achieved and integrated into the core of value sensitive design theory and practice. For example, it became a best practice for VSD researchers to include a section called "Researcher Stance" in their publications, in which the researchers self-disclose their background, relation to the participants in the study, and relevant personal values that may be important for readers in evaluating the research.

Continuing with this self-reflective process, a workshop in Aarhus, Denmark in 2015 and a second workshop at the Lorentz Centre in Leiden, The Netherlands in 2016 began the discussion about the next decade for value sensitive design. A set of 12 grand challenges emerged from those conversations. A special issue of the journal *Ethics and Information Technology* was devoted to this topic, comprised of a broad range of short thought pieces on novel applications and theoretical directions (in progress). An international network of research centers in the United States, Australia, China, Denmark, Germany, The Netherlands, and Sweden has been formed to share research findings as well as exchange lessons learned, best practices, and findings from projects undertaken with industry and government organizations.

While value sensitive design has experienced much success with regard to its adoption and appropriation in the research community, as it makes its foray into industry appropriation much is yet to be done to support widespread, meaningful adoption. The time is now ripe to ask this question: What near term next steps for value sensitive design? And what longer term vision?

Panel Focus. To convey the focus of this panel, we provide a list of some of the questions the panel takes up, including:

- What are the key grand challenges researchers and practitioners working within a VSD approach should take up?
- How does VSD speak to and differ from other design-for-values approaches to technologies that are referred to and supported in the literature as well as by industry?
- What are near term next steps for VSD?
- Regarding the theme of ETHICOMP 2020, "Paradigm Shifts in ICT Ethics: Societal Challenges in the Smart Society," what are the key challenges faced in the smart society? Given the strong interdependency between technology and policy in the smart society, how can VSD enable policy design and technical design proceed in tandem?
- How can VSD handle apparently disparate, yet converging technologies that are essential to the fourth industrial revolution (i.e., AI, AR/VR, exoskeletons, etc.)?
- What lessons can be learned from the diverse fields in which VSD has been applied, particularly with how to account for a plurality of contexts, concerns and values?
- As VSD continues to develop and be appropriated in industry and universities, what would computer science practice and education look like 20 years from now?
- How will we know if a VSD approach is improving computer science practice? What metrics can we use? What data should we be collecting now as baseline data to enable assessments 5, 10, and 20 years from now?
- Given the merits of VSD approach as well as how it aims to seamlessly integrate in existing design practices, how can we make VSD more accessible and easier to understand by a wide range of engineers and technologists as well as by non-specialists and non-designers?

Panelists and Moderator. Panelists are comprised of two senior—Batya Friedman and Jeroen van den Hoven—and two younger—Steven Umbrello and Daisy Yoo—scholars working in value sensitive design. As a group, they represent a diversity of expertise including applied moral philosophy, computer science, design, ethics, and information. They also represent countries in Europe and North America and are comprised of a balance of women and men. The senior scholars pioneered value sensitive design; they will be positioned to speak to VSD's early years and development to date as well as their hopes and visions for the future of VSD. The younger scholars came of age in an intellectual landscape in which VSD was established and have taken VSD further in their respective work; they will be positioned to speak to where they see VSD's opportunities and challenges for younger scholars as well as their hopes and visions for the future of VSD. Thus, the panel is poised to discuss VSD's future within a multi-generational light. David Hendry, the panel moderator is also a long-term member of the VSD community.

Batya Friedman is a Professor in the Information School at the University of Washington where she co-directs the Value Sensitive Design Lab. She pioneered VSD in the 1990s. Her 2019 MIT Press book co-authored with Dave Hendry is *Value Sensitive Design: Shaping Technology with Moral Imagination*.

David Hendry is an Associate Professor in the Information School at the University of Washington where he co-directs the Value Sensitive Design Lab. Dave is currently at work on new ideas for teaching value sensitive design through tech policy case studies – the joint consideration of policy and technical design.

Steven Umbrello is the Managing Director of the Institute for Ethics and Emerging Technologies where his primary research focus is on autonomous weapon systems, responsible innovation and the general ethics of emerging and transformative technologies.

Jeroen van den Hoven is the University Professor in Ethics and Technology at Delft University of Technology and the scientific director of the Delft Design for Values Institute. He is a permanent member of the European Group on Ethics and Editor-in-Chief of *Ethics and Information Technology*.

Daisy Yoo is a Postdoctoral Research Fellow and a member of the Value Sensitive Design in Higher Education (VASE) project at the Aarhus University, Denmark. Dr. Yoo completed her Ph.D. at the University of Washington, where she worked on the Voices from the Rwanda Tribunal project to investigate multi-lifespan design.

Panel Structure. The 90-minute panel will be organized as follows:

1. Introduction of Panel Topic and Panelists (6 min)
2. Remarks by Individual Panelists (6 min each; 24 min total)
3. Comments and Questions from the Audience (30 min)
4. Audience Small Group Work to Discuss and Record Audience Visions for VSD (30 min)

KEYWORDS: applied ethics, computing education, computing practice, grand challenges, responsible innovation, value sensitive design.

REFERENCES

Bender, E. M., & Friedman, B. (2018). Data statements for natural language processing: Toward mitigating system bias and enabling better science. *Transactions of the Association for Computational Linguistics, 6*, 587–604.

Borning. A., & Mueller, M. (2012). Next steps for value sensitive design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)* (pp. 1125–1134). ACM.

Cawthorne, D., & Cenci, A. (2019). Value sensitive design of a humanitarian cargo drone. In *2019 International Conference on Unmanned Aircraft Systems (ICUAS)* (pp. 1117–1125). IEEE.

Cheon, E., & Su, N. M. (2016). Integrating roboticist values into a Value Sensitive Design framework for humanoid robots. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 375–382). IEEE.

Friedman, B. (1996). Value-sensitive design. *interactions, III*(6), 17–23.

Friedman, B., & Hendry, D.G. (2019). *Value Sensitive Design: Shaping technology with moral imagination*. Cambridge, MA: MIT Press.

Friedman, B., & Hendry, D. (2012). The envisioning cards: a toolkit for catalyzing humanistic and technical imaginations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)* (pp. 1145–1148). ACM.

Friedman, B., & Nathan, L. P. (2010). Multi-lifespan information system design: a research initiative for the hci community. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)* (pp. 2243–2246). ACM.

Denning, T., Borning, A., Friedman, B., Gill, B. T., Kohno, T., & Maisel, W. H. (2010). Patients, pacemakers, and implantable defibrillators: Human values and security for wireless implantable medical devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)* (pp. 917–926). ACM.

Harbers, M., & Neerincx, M. A. (2017). Value sensitive design of a virtual assistant for workload harmonization in teams. *Cognition, Technology & Work, 19*(2-3), 329–343.

Mueller, M., & Heger, O. (2018). Health at any Cost? Investigating Ethical Dimensions and Potential Conflicts of an Ambulatory Therapeutic Assistance System through Value Sensitive Design. In *ICIS 2018 Proceedings*. Association for Information Systems.

Santoni de Sio, F., & Van den Hoven, J. (2018). Meaningful human control over autonomous systems: a philosophical account*. Frontiers in Robotics and AI, 5*, 15.

Timmermans, J., Zhao, Y., & van den Hoven, J. (2011). Ethics and nanopharmacy: Value sensitive design of new drugs. *Nanoethics, 5*(3), 269–283. https://doi.org/10.1007/s11569-011-0135-x

Umbrello, S. (2019). Atomically precise manufacturing and responsible innovation: A value sensitive design approach to explorative nanophilosophy. *International Journal of Technoethics (IJT), 10*(2), 1–21. https://doi.org/10.4018/IJT.2019070101

Umbrello, S., & De Bellis, A. F. (2018). A value-sensitive design approach to intelligent agents. In *Roman Yampolskiy (Ed.), Artificial Intelligence Safety and Security* (pp. 395–410). CRC Press. https://doi.org/10.13140/RG.2.2.17162.77762

van den Hoven, J. (2013). Value sensitive design and responsible innovation. In *R. Owen, J. Bessant, & Heintz, M. (Eds.), Responsible innovation: Managing the responsible emergence of science and innovation in society*. John Wiley & Sons, LTD. https://doi.org/10.1002/9781118551424.ch4

van Wynsberghe, A. (2013). Designing robots for care: Care-centered value-sensitive design. *Science and Engineering Ethics, 19*(2): 407–433.

Winkler, T., & Spiekermann, S. (2019). Human Values as the Basis for Sustainable Information System Design. *IEEE Technology and Society Magazine, 38*(3), 34–43. https://doi.org/10.1109/MTS.2019.2930268

Yoo, D., Derthick, K., Ghassemian, S., Hakizimana, J., Gill, B., & Friedman, B. (2016). Multi-lifespan design thinking: two methods and a case study with the Rwandan diaspora. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '16)* (pp. 4423-4434). ACM.

Yoo, D., Huldtgren, A., Woelfer, J. P., Hendry, D. G., and Friedman, B. (2013). A value sensitive action-reflection model: Evolving a co-design space with stakeholder and designer prompts. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)* (pp. 419-428). ACM.

UNIVERSALITY OF HOPE IN PATIENT CARE: THE CASE OF MOBILE APP FOR DIABETES

Majid Dadgar, K.D. Joshi

University of San Francisco (USA), University of Nevada, Reno (USA)

mdadgar@usfca.edu; kjoshi@unr.edu

EXTENDED ABSTRACT

"Hope" is the thing with feathers -

That perches in the soul -

And sings the tune without the words -

And never stops - at all -

I've heard it in the chillest land -

And on the strangest Sea -

Yet - never - in Extremity,

It asked a crumb - of me.

- Emily Dickinson

In this paper we investigate the human value of hope in the self-management systems used by the patients with diabetes to manage their chronic health conditions. We use value sensitive design (VSD) framework to uncover the value instances revealed in our interviews with patients with diabetes. The value instances identified in the interview transcript map to components of hope theory: goal, agency, and pathways. We recommend technology features that allow patients with diabetes to achieve their goals in life while managing their chronic conditions.

1. INTRODUCTION

As information and communication technologies (ICTs) advance, their uses and applications become more diverse and complex. These complex technologies are designed and used by humans and therefore, need a human-centric approach. The human-centric ICTs in the healthcare context play a major role in improving patients' lives (Bardhan, Chen, & Karahanna, 2017). These ICTs should be sensitive to the values of the patients (Dadgar & Joshi, 2018).

The value sensitive design (VSD) framework has proven to be an effective tool in identifying and explaining the human values of technology users and their development and change over time (Friedman, Howe, & Felten, 2002). In this paper we investigate the value of hope in the patients with diabetes. Specifically, we identify the instances of the value of hope for the patients with diabetes and recommend technology features that could support them.

2. HOPE AND SELF-MANAGEMENT

Hope in the theory of hope is defined as the perceived capability to derive pathways to desired goals, and motivate oneself via agency thinking to use those pathways (Snyder, 2000). Setting and attainment of goals are central in how the construct of hope is conceptualized by Snyder. People have higher hope when they believe their goals are attainable. Pathways to desired goals are necessary for hopeful thoughts. People who can realize and pursue pathways toward their desired goals stay hopeful over time. The sense of agency in achieving their goals through purposeful pathways motivates and empowers patients. The agency and pathway components of hope are distinct but entangled. At difficult times when people face barriers towards their goals, the strong sense of agency enables them to tackle the barriers. Positive and negative emotions are the result of the perceived success in achieving goals. Perceived success in achieving goals creates positive emotions in people and perceived failure triggers negative emotions.

We investigate how these components of hope theory can be supported using ICTs. We use VSD to identify value instances of hope for the patients with diabetes who use mobile app to self-manage their chronic conditions. The value instances identified in the interviews bridge the support needed from hope interventions implicated in technology features (see Table 1).

Table 1 An instance of the value of hope extracted from interview data, hope components mapped to the value instance, and technology features that can support this value instance and hope components

Value Instances	**Hope components**	**Technology features**
"I felt upset [when I knew I was diagnosed with prediabetes] because all my life I had done the right things. I had exercised, I had eaten right and even when I had to stop exercising I still ate right and so I was very disappointed. I was angry at my muscle disease and I was upset, I almost started crying because I was just … One more thing that has gone wrong with my health because of my other disease, so yeah, I was upset. I at first told the doctor that I didn't want to take anything. I was mad. I didn't want to do this because it was admitting that I had diabetes or pre-diabetes or whatever."	Goal: healthy life style Agency: lack of agency reflected in negative emotions – "I was very disappointed", "I almost started crying", "I was upset", "I was mad". Pathway: exercise and eating right – "I had exercised, I had eaten right and even when I had to stop exercising I still ate right"	Digital coaches are intelligent technology-based services that simulate human coaches and reinforce patients on their pathways toward goals and enhance patients' agency by providing motivational messages, techniques, and resources in real time.

3. METHOD

We have used VSD to develop interview strategies and criteria that will reveal the values of the patients with diabetes (Friedman & Hendry, 2019). We interviewed 20 patients with diabetes. In the first meeting, patients were introduced to a mobile app that they could use to manage their diabetes, its symptoms, and life style changes. After the first meeting, patients used the mobile app to manage their diabetes for one week. In the second meeting, patients were interviewed about their experience

with the diabetes mobile app and their needs and concerns. Interviews were transcribed and analyzed based on VSD to identify value instances that map to the hope components. Next we make recommendations that how technology can support these values instances and hope components necessary to create and maintain hope in the patients with chronic diseases and conditions.

4. RESULTS AND DISCUSSION

Hope instances identified and extracted from interviews with patients with diabetes illustrate how this value manifests in different variations in patients' lives. These value instances could be supported effectively by technology features. The hope components with one example of value instance and technology features are provided in Table 1.

In Table 1 an example of a value instance mapped to hope components with support of technology features is provided to illustrate how value-sensitive technologies support goal-oriented agency and pathways in patients with diabetes. A patient diagnosed with diabetes expressing and describing negatives emotions indicates an underlying issues with patient's agency and available pathways. The available pathways towards a healthy life style for this patient have not been effective in achieving her goals to live a healthy life style. The ineffective pathways undermine patients' feeling of agency. The patient questions her abilities in achieving goals and develops negative emotions of being upset and mad. Digital coaches enhance patients' agency by motivating patients along the way in pursuit of their goals. An empowered patient with higher agency can tackle barriers and negative emotions in achieving goals. Digital coaches designed in the diabetes mobile app provide guidance, resources, and emotional support. The real time and on-demand access to digital coaches increases patients' motivations in achieving their goals by reinforcing and reaffirming patients' thoughts towards their goals.

5. CONCLUSION

In this work in progress study we being to explore the role of ICTs in supporting and enhancing patients' hope to self-manage their diabetic chronic conditions. We use VSD to design interview strategies and questions for patients with diabetes to identify their needs and desires to use a diabetes mobile app and self-manage their chronic conditions. We use hope theory components of agency, pathways, and goal to translate value instances of hope into supportive technology features.

This study provides guidance and recommendations for the healthcare providers and system developers to assist patients with diabetes self-manage their chronic conditions. The paper instantiates value of hope in the context of ICT-enabled self-management of diabetes and illustrates how system developers can design and develop technology features and healthcare providers to use those technology features to enhance the feeling of agency in the patients and provide effective pathways towards their goals.

KEYWORDS: Value Sensitive Design, Hope, Self-management, Healthcare, Diabetes, Agency, Pathways.

REFERENCES

Bardhan, I., Chen, H., & Karahanna, E. (2017). The Role of Information Systems and Analytics in Chronic Disease Prevention and Management. *MIS Quarterly*, (Call for Papers MISQ Special Issue).

Dadgar, M., & Joshi, K. D. (2018). The Role of Information and Communication Technology in Self-Management of Chronic Diseases: An Empirical Investigation through Value sensitive design. *Journal of the Association for Information Systems (JAIS)*, *19*(2), 86–112.

Friedman, B., & Hendry, D. G. (2019). *Value Sensitive Design: Shaping Technology with Moral Imagination*. Cambridge, MA: The MIT Press.

Friedman, B., Howe, D. C., & Felten, E. (2002). Informed consent in the Mozilla browser: Implementing value-sensitive design. *Proceedings of the 35th Annual Hawaii International Conference on System Sciences*. Presented at the Proceedings of the 35th Annual Hawaii International Conference on System Sciences, Big Island, HI.

Snyder, C. R. (2000). *Handbook of Hope: Theory, Measures, and Applications*. San Diego, Calif: Academic Press.

VALUE SENSING ROBOTS: THE OLDER LGBTIQ+ COMMUNITY

Adam Poulsen, Ivan Skaines, Suzanne McLaren, Oliver K. Burmeister

Charles Sturt University (Australia), Newcastle Pride (Australia), Charles Sturt University (Australia), Charles Sturt University (Australia)

apoulsen@csu.edu.au; iskaines@ozemail.com.au; smclaren@csu.edu.au; oburmeister@csu.edu.au

EXTENDED ABSTRACT

LGBTIQ+ older adults are an under-researched community in aged care (Fredriksen-Goldsen, Kim, Barkan, Muraco, & Hoy-Ellis, 2013). This community is experiencing loneliness at higher rates than the general older adult population (Fredriksen-Goldsen, 2016; Hughes, 2016). The value sensitive design (VSD) and use of social care robots (SCRs) provides an innovative advance toward equity for older LGBTIQ+ adults at risk of loneliness. Good care is person-centred, culturally competent, and follows principles of descriptive ethics, consistent with the intentions of VSD (Friedman & Hendry, 2019). The completed pilot of a larger, ongoing mixed-methods study exploring these subjects is discussed here.

LITERATURE REVIEW

VSD is a popular method for investigating stakeholder values and designing systems to account for those values (Friedman & Hendry, 2019). Recent VSD works (e.g., Jacobs & Huldtgren, 2018; Manders-Huits, 2011), attempt to move the methodology towards normative ethics, aiming to establish a standardised design decision framework to create technologies. In contrast, VSD pioneers were careful not to suggest that *values are either entirely normative nor descriptive*. Friedman et al. (2006) state that "each value has its own language and conceptualizations within its respective field, and thus warrants separate treatment" (p. 366) and that no list of values is comprehensive. Furthermore, Friedman, Kahn, Borning, and Huldtgren (2013) maintain that some values are universal and normative.

Similarly, good care practice is neither entirely normative nor descriptive. What each person and community needs and values in care matters (Abma, Molewijk, & Widdershoven, 2009). Descriptive principles of care hold instrumental value for individuals, and they should be considered in VSD. At the same time, there are normative principles in care that are intrinsically good and valuable, including safety and wellbeing, as identified by duty of care, professional ethics, and law.

SCRs play a role in social support/care by enabling, assisting in, or replacing social interactions. For good robot-delivered care, SCRs need to ensure both normative intrinsic values and descriptive instrumental values found in real care practices. Moreover, just as good care is determinative in practice (Beauchamp, 2004), SCRs must account for changing and emerging values in care. *Value sensing robots* (i.e., robots which attempt to learn user values and adapt behaviour to suit those values) may work towards this.

METHODOLOGY

To encapsulate key concepts in the design of value sensing robots, values in motion design (VMD) was conceptualised (Poulsen & Burmeister, 2019; Poulsen, Burmeister, & Kreps, 2018). Chiefly, VMD aims to account for the pluralistic and evolving nature of values through the design of SCRs which make

explicit value-driven decisions to govern actions; these decisions are shaped to the values of the user in situ, when it is safe to do so and only within a framework of intrinsic values implicitly embedded into the design. As a starting point in VMD, designers aim to capture community values that the care robot can then shape to the individual during run-time to provide person-centred care.

To test VMD, an interpretivist pilot study was conducted with five LGBTIQ+ older adults (three gay men, one gay gender-fluid person, and one lesbian non-binary person). Through semi-structured interviews, participants were questioned about the LGBTIQ+ experience of ageing, aged care, social isolation, and loneliness, as well as the older LGBTIQ+ community's values. These interviews were transcribed and analysed using content analysis. Ethics approval from the university and from participating LGBTIQ+ communities, from which participants were recruited for this and the larger study, was obtained.

FINDINGS

Using content analysis, the values of the LGBTIQ+ older adults interviewed were derived (see Figure 1). Table 1 features examples of how the LGBTIQ+ older adults conceptualise values compared to the literature. The value conceptualisations in Table 1 were derived from the interviews using content analysis.

Figure 1. The older LGBTIQ+ community's values found in five pilot study interviews. Only those values which were referenced by three or more persons have been included

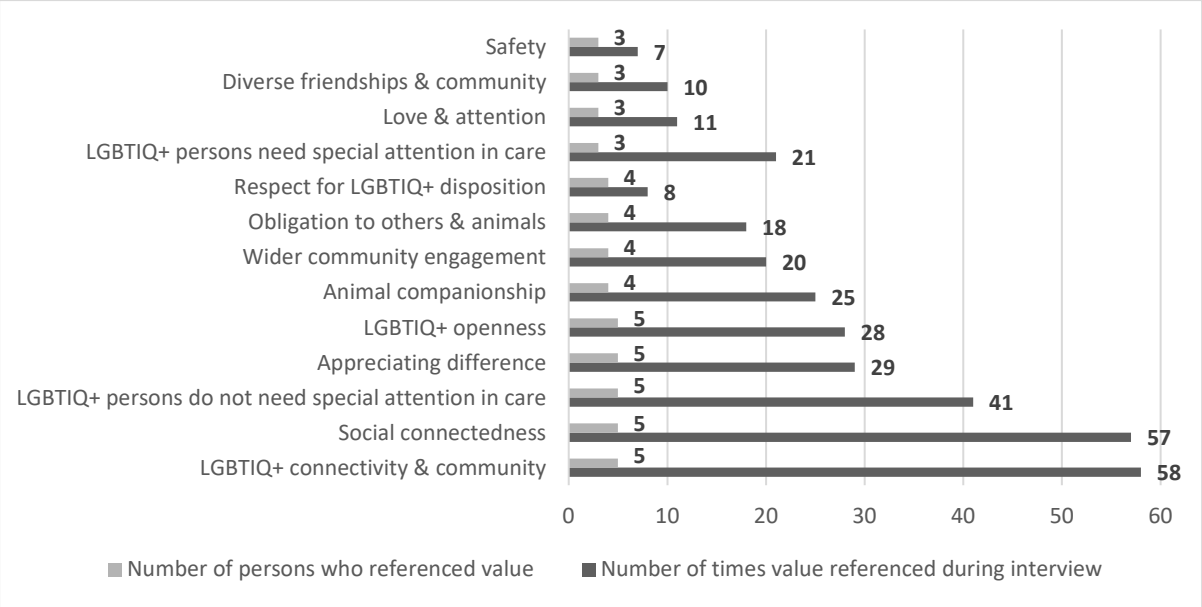


Table 1 Exemplary pilot study values compared to values found in the literature

LGBTIQ+ older adult conceptualisation of values	**Equivalent values found in the literature**
Appreciating difference	Inclusivity
LGBTIQ+ connectivity & community	Community
Diverse friendships & community	Cultural diversity
LGBTIQ+ openness	Freedom of expression
LGBTIQ+ persons do not need special attention in care	Equality
Obligation to others & animals	Being needed
LGBTIQ+ persons need special attention in care	Equity
Respect for LGBTIQ+ disposition	Respect

DISCUSSION

Figure 1 shows how LGBTIQ+ older adults prioritise values. Value sensing SCRs for LGBTIQ+ older adults can be configured with these values as a generalised value framework. Thereafter, using adaptive functions, the SCR can reprioritise those values and learn new ones in situ with the user to provide person-centred care. To explain value sensing adaptive functions, by analogy, consider the current care robot Elli-Q[39] which examines an image, recognises the objects in an image, and provides a verbal translation of what is featured in the image. Value sensing could examine, recognise, and translate user values in a similar way.

For example: An LGBTIQ+ older adult who uses an SCR is sitting with another person, but they are no longer conversing. The SCR ought to be able to understand what user values are being impacted. Is the user desiring social connectedness, but they have exhausted conversation topics? Are the user and the other person struggling to socially connect due to cultural differences? Does the user enjoy the silence and feel adequately socially connected?

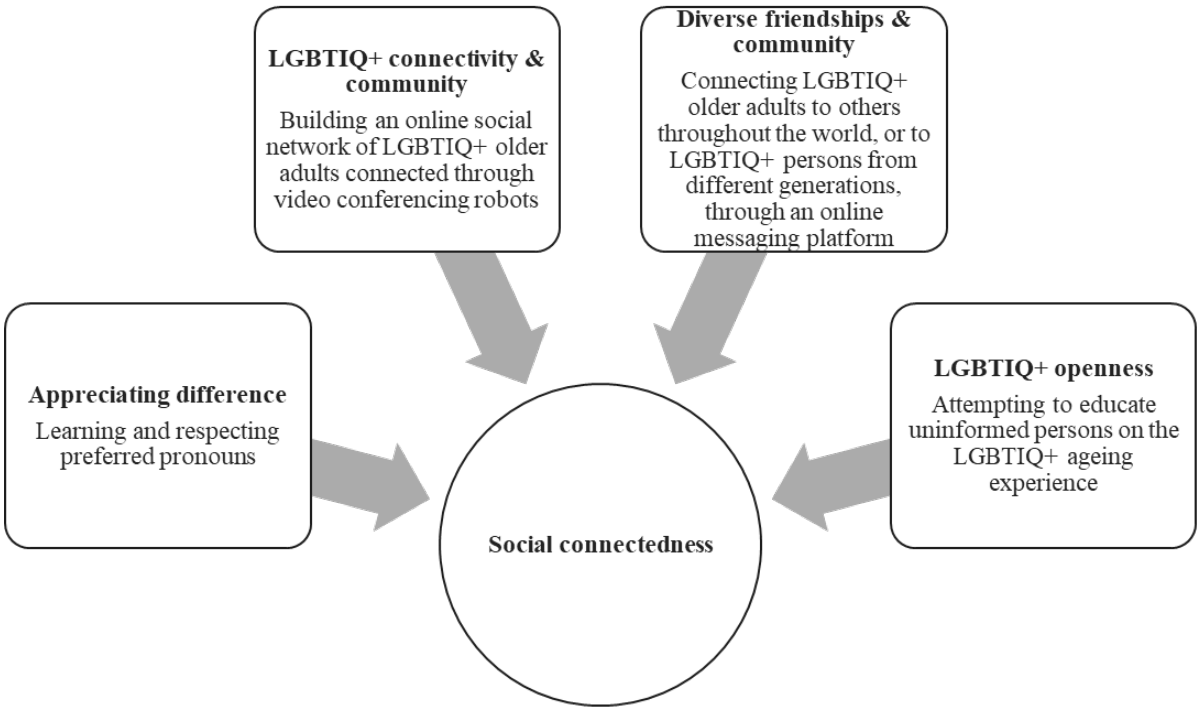
Knowing what user values are being impacted, and why, helps the SCR hone its delivery of care. If the SCR understands that running out of conversation negatively impacts the value *social connectedness*, then it will be able to support the user to better exercise this value in the future. For instance, the SCR could suggest conversation topics. However, even with this social support, perhaps the LGBTIQ+ older adult is still not feeling socially connected (e.g., giving short answers or often looking away) because the SCR is suggesting conversation topics which are not relatable for the LGBTIQ+ older adult (e.g., family or children). Arising from Table 1, observing that LGBTIQ+ older adults conceptualise respect as *respect for LGBTIQ+ disposition*, the SCR is negatively impacting this value. A value sensing robot ought to understand the values and value conceptualisations of different communities and individuals to provide person-centred care.

Figure 2 further demonstrates how social connectedness might be achieved with the value conceptualisations of LGBTIQ+ older adults in mind. In run-time, value sensing robots could adapt these values to better suit the values of individual LGBTIQ+ older adults. For instance, consider a video

[39] *See* https://elliq.com/

conferencing robot which plays a role in social care by hosting video calls across an online LGBTIQ+ social network. If the user does not utilise the existing functions designed to ensure *LGBTIQ+ connectivity and community*, then it might instead adapt this value (and subsequent behaviours) to schedule cafe meetups with other local LGBTIQ+ older adults connected to the online social network.

Figure 2. Designing SCR components with the older LGBTIQ+ community's value conceptualisations in mind, each working to ensure the normative intrinsic value social connectedness



CONCLUSION

Value sensing SCRs need to adapt to the values of LGBTIQ+ older adults in a *person-centred care mode* to help overcome the loneliness that is presently widespread in this community. With adaptive functionality, SCRs can be designed to make dynamic, value-driven decisions in situ to customise the level of care down to the person-centred level within duty of care limits.

KEYWORDS: Healthcare robotics, community, LGBTIQ+ ageing, value sensitive design.

REFERENCES

Abma, T., Molewijk, B., & Widdershoven, G. (2009). Good Care in Ongoing Dialogue. Improving the Quality of Care Through Moral Deliberation and Responsive Evaluation. *Health Care Analysis,* 17, 217-235.

Beauchamp, T. L. (2004). Does ethical theory have a future in bioethics? *The Journal of Law, Medicine & Ethics,* 32(2), 209-217.

Fredriksen-Goldsen, K. I. (2016). The Future of LGBT+ Aging: A Blueprint for Action in Services, Policies, and Research. *Generations (San Francisco, Calif.),* 40(2), 6-15. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/28366980

Fredriksen-Goldsen, K. I., Kim, H.-J., Barkan, S. E., Muraco, A., & Hoy-Ellis, C. P. (2013). Health disparities among lesbian, gay, and bisexual older adults: results from a population-based study. *American Journal of Public Health,* 103(10), 1802-1809.

Friedman, B., & Hendry, D. G. (2019). *Value Sensitive Design: Shaping Technology with Moral Imagination*. Cambridge, MA: MIT Press.

Friedman, B., Kahn, P. H., Borning, A., & Huldtgren, A. (2013). Value Sensitive Design and Information Systems. In N. Doorn, D. Schuurbiers, I. van de Poel, & M. E. Gorman (Eds.), *Early engagement and new technologies: Opening up the laboratory* (pp. 55-95). Dordrecht: Springer Netherlands.

Friedman, B., Kahn, P. H. J., & Borning, A. (2006). Value Sensitive Design and Information Systems. In P. Zhang & D. Galletta (Eds.), *Human-computer interaction and management information systems: Foundations* (pp. 348-372). New York: M. E. Sharpe.

Hughes, M. (2016). Loneliness and social support among lesbian, gay, bisexual, transgender and intersex people aged 50 and over. *Ageing and Society,* 36(9), 1961-1981.

Jacobs, N., & Huldtgren, A. (2018). Why value sensitive design needs ethical commitments. *Ethics and Information Technology,* 1-4.

Manders-Huits, N. (2011). What values in design? The challenge of incorporating moral values into design. *Science and Engineering Ethics,* 17(2), 271-287.

Poulsen, A., & Burmeister, O. K. (2019). Overcoming carer shortages with care robots: Dynamic value trade-offs in run-time. *Australasian Journal of Information Systems,* 23.

Poulsen, A., Burmeister, O. K., & Kreps, D. (2018). The ethics of inherent trust in care robots for the elderly. In D. Kreps, C. Ess, L. Leenen, & K. Kimppa (Eds.), *This Changes Everything – ICT and Climate Change: What Can We Do?* (pp. 314-328). doi:10.1007/978-3-319-99605-9_24

VALUE SENSITIVE DESIGN AND AGILE DEVELOPMENT: POTENTIAL METHODS FOR VALUE PRIORITIZATION

Till Winkler

Vienna University of Economics and Business (Austria)

till.winkler@wu.ac.at

EXTENDED ABSTRACT

Software is a crucial element of digital technology and has become an integral part of our society. However software is not neutral, but acts as a mediator for human values and biases, molding its own operational context, which in return shapes human perception and actions, creates new practices and subsequently ways of living (Verbeek 2008). Mediation through software can have potential negative effects such as biases introduced by algorithms (Obermeyer et al. 2019). This makes it obvious, that software developers have the responsibility to consider human values, potential biases and ethical concerns during the development process. The idea to consider values during technology development is as old as technology itself, but is often limited to instrumental values, such as efficiency and reliability (van de Poel 2015). Value sensitive Design (VSD) takes a stand for the integration of values with ethical importance into technology design (Friedman et al. 2013; van de Poel 2015). Since its initial conceptualization, VSD has spored an uncountable amount of value-oriented approaches and continues to evolve its own unique methods (Friedman et al. 2017). Unfortunately neither VSD nor any other value-oriented approach has found widespread deployment in the industry (Miller et al. 2007). This fact might be explained by a lack of light-weight methods compatible with agile development, a shortage of methods for important tasks and a lack of consistent methodological description (Miller et al. 2007; van de Poel 2015; Burmeister 2016).

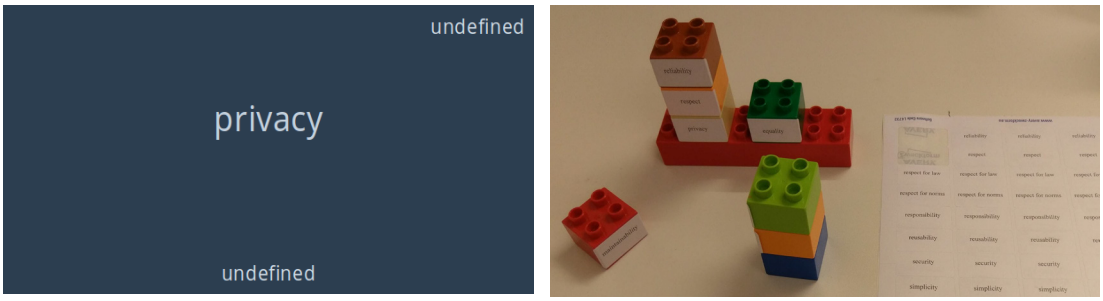
In the extended version of this paper, we examine in detail the touching points and contractions between VSD and AD. Furthermore, we present two new candidates for unique VSD methods, which we tested in the context of mobile navigation applications. Both methods have the overall aim to support the integration of VSD into an agile development process.

The once popular waterfall development process is based on the fallacy, that software requirements can be fully specified and optimal design solutions found upfront. Instead, software development is a complex undertaking, involving many uncertainties outside of a team's control; therefore flexibility is vital and at the same time hindered by upfront planing or predefined design. Additionally, upfront planning is nearly impossible due to the fast evolution of technology and the emergence of new practices during project realization (Schmidt 2016). While the integrative and iterative nature of VSD (Friedman et al. 2017) provides a high degree of flexibility, this benefit is rarely utilized. In many cases VSD is uses as a mixture of upfront requirement engineering (based on values) and design activity before the actual development. We advocate to integrate VSD similar to the way essential requirement engineering activities (elicitation, analysis and negotiation, documentation, validation) are included in agile practices; mingled together and performed iteratively throughout the development cycle (Ramesh et al. 2010). agile development starts with a rough approximation to final requirements and adds details iteratively during the whole development process (Rees 2002). In a similar fashion, VSDs tripartite methodology could add details to some important values during the whole development. In agile development requirements are considered decouple from each other, allowing a flexible order of implementation (Silliti & Succi 2005). Completely decoupling values from each other prevents

solving value tensions (Friedman et al. 2017), instead we suggest to implement one value after the other in accordance to their priority. Potential value tension needs to be solved during the implementation of an additional value. In general, strong requirement prioritization and feedback based on face-to-face collaboration with all stakeholder is key to the success of agile development (Silliti & Succi 2005; Ramesh et al 2010). Similarly values need to prioritized at the beginning of each agile development cycle together with all stakeholders in a fair and transparent manner.

Not only our proposed integration of VSD into agile development calls for a suitable value prioritization methods but also the possibility to elicit more than 360 known values from various backgrounds (Winkler & Spiekerman 2019). For the context of mobile navigation application, we let 264 user rate how important they perceived 48 values. In general, users rated more instrumental and popular discussed values such as „efficiency“ and „reliability“ as higher compared to values of potential ethical importance such as „autonomy“, „environmental protection“ and „health“. As a follow up we let 184 participants conceptualize six different values, two previously rated high, two rated medium and two rated with low importance. Analyzing the amount of mentioned aspects (or concepts) revealed a correlation between number of mentioned concepts and previously (by others) rated importance. It is easier for participants to come up with aspects for a highly rated value than for lower rated value. This suggests a bias, in form of a availability heuristic, when it comes to prioritizing values. An availability heuristic leads stakeholder to consider recently discussed information (or values) as more important, because these can be recalled easier (Wänke et al. 1995). Such a bias towards recently discussed and popular values might distort the aims of VSD, to go beyond instrumental values, and calls for mitigation methods. In the extended version of this paper we present the results of two additional value prioritization methods. The first is based on the implicit association test (IAT), a well established method in psychology to asses the subconscious associations between concepts in a participants memory (Greenwald et al. 1998). To achieve that a computer-based setup was developed measuring among participants elapse time in ms for value decisions. The other method is based on the ideas of serious play and is well suited for co-design or focus groups sessions (Garde & van der Voort 2016). Both methods were tested for the context of mobile navigation applications. Figure 1 shows the initial setup for the two value prioritization methods.

Figure 1. Two new value prioritization methods



a) IAT for value decisions b) Value bricks

We suggest to integrate VSD into an agile development process to increase its recognition outside of academia. Furthermore, we present results from three value prioritization methods to facilitate such an integration. It is our aim to foster a discussion within the VSD community about an integration into agile development and the legitimacy to prioritize values.

KEYWORDS: value sensitive design, human values, agile development, value prioritization.

REFERENCES

Burmeister, O. K. (2016). The development of assistive dementia technology that accounts for the values of those affected by its use. *Ethics and Information Technology*, 18(3), 185-198.

Friedman, B., Hendry, D. G., & Borning, A. (2017). A survey of value sensitive design methods. *Foundations and Trends® in Human–Computer Interaction*, 11(2), 63-125.

Friedman, B., Kahn, P. H., Borning, A., & Huldtgren, A. (2013). Value sensitive design and information systems. *In Early engagement and new technologies: Opening up the laboratory* (pp. 55-95). Springer, Dordrecht.

Garde, J. A., & van der Voort, M. C. (2016). Could LEGO® Serious Play® be a useful technique for product co-design? *Design Research Society*.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: the implicit association test. *Journal of personality and social psychology*, 74(6), 1464.

Miller, J. K., Friedman, B., Jancke, G., & Gill, B. (2007, November). Value tensions in design: the value sensitive design, development, and appropriation of a corporation's groupware system. *In Proceedings of the 2007 international ACM conference on Supporting group work* (pp. 281-290). ACM.

Obermeyer, Z., & Mullainathan, S. (2019, January). Dissecting Racial Bias in an Algorithm that Guides Health Decisions for 70 Million People. *In Proceedings of the Conference on Fairness, Accountability, and Transparency* (pp. 89-89). ACM.

Ramesh, B., Cao, L., & Baskerville, R. (2010). Agile requirements engineering practices and challenges: an empirical study. *Information Systems Journal*, 20(5), 449-480.

Rees, M. J. (2002, December). A feasible user story tool for agile software development?. *In Ninth Asia-Pacific Software Engineering Conference*, 2002. (pp. 22-30). IEEE.

Schmidt, C. (2016). Agile software development teams. Springer International Publishing.

Sillitti, A., & Succi, G. (2005). Requirements engineering for agile methods. *In Engineering and Managing Software Requirements* (pp. 309-326). Springer, Berlin, Heidelberg.

van de Poel, I. (2015). Design for values in engineering. *Handbook of Ethics, Values, and Technological Design: Sources, Theory, Values and Application Domains*, 667-690.

Verbeek, P. P. (2008). Morality in design: Design ethics and the morality of technological artifacts. *In Philosophy and design* (pp. 91-103). Springer, Dordrecht.

Wänke, M., Schwarz, N., & Bless, H. (1995). The availability heuristic revisited: *Experienced ease of retrieval in mundane frequency estimates.* Acta Psychologica, 89(1), 83-90.

Winkler, T., & Spiekermann, S. (2019). Human Values as the Basis for Sustainable Information System Design. *IEEE Technology and Society Magazine*, 38(3), 34-43.

VALUE SENSITIVE DESIGN EDUCATION: STATE OF THE ART AND PROSPECTS FOR THE FUTURE

David G. Hendry, Eva Eriksson, Anisha Thilini Jessica Fernando, Irina Shklovski, Dylan Cawthorne, Daisy Yoo

Unviersity of Washington (USA), Aarhus University (Denmark) , University of South Australia (Asustralia), IT University of Copenhagen (Denmark), University of Southern (Denmark), Aarhus University (Denmark)

dhendry@uw.edu, evae@cc.au.dk; anisha.fernando@mymail.unisa.edu.au; irsh@itu.dk; dyca@mmmi.sdu.dk; dyoo@cc.au.dk

EXTENDED ABSTRACT

In recent years the importance of responsible innovation in engineering and technical education has grown. Value sensitive design provides theory, method, and practice to account for human values in a principled and systematic manner throughout the design process. Accordingly, value sensitive design is an approach for responsible innovation in engineering and technical design. In this panel, with panellists from Australia, Europe, and North America we propose to take up the question: How can value sensitive design be used such that social, ethical, and policy considerations are interwoven into technical education, in engineering, informatics, and related fields? The panelists will give brief seed presentations and the panel moderator will elicit questions from the audience. The panel will end with a brief synthetic presentation that seeks to summarize the panel-audience discussion. This panel will help to develop a community of practice around technical education and values.

Responsible innovation crystalizes a number of concerns related to the human-technology relationship. Perhaps most fundamentally is the goal of making moral progress (van den Hoven, 2013). That is to say, responsible innovation ought to lead to technology that leads to greater justice, more human dignity and well-being, and better odds for human survival, over the next 100 years, in the gathering climate emergency.

Value sensitive design (Friedman & Hendry, 2019; Friedman, Hendry & Borning, 2017) provides theory, method, and practice to account for human values in a principled and systematic manner throughout the design process. With a commitment to "*progress, not perfection*," value sensitive design is intended to be appropriated and used to augment or extend existing technical design processes. Accordingly, value sensitive design is an approach for responsible innovation in engineering and technical design.

The aim of this panel session is to bring together educators from North America, Europe, and Australia to take up the question:

> *How can value sensitive design be used such that social, ethical, and policy considerations are interwoven into technical education, in engineering, informatics, and related fields?*

This question is relevant for a number of reasons. First, there appears to be widespread interest in this question and more broadly how to teach ethics in engineering, informatics, and related fields. One outstanding example, and a leader in responsible innovation, with many best practices in education, is the Department of Values, Technology, and Innovation, UT Delft, The Netherlands (UT Delft, 2019).

More recently, several influential American computer science departments, in response to the pernicious and widespread impact of Artificial Intelligence on society, have committed to transforming computer science education so that ethical questions from a societal lens are considered systemically (Massachusetts Institute of Technology, 2019; Stanford, 2019). Seeking such a transformation, Grosz et al. (2019) report on a curriculum that embeds ethical and moral reasoning throughout the undergraduate computer science curriculum. This is an approach that goes well beyond a unit or two on professional ethics (Association of Computer Machinery, 2019). In addition, Frauenberger & Purgathofer (2019), have developed a curriculum that positions entry-level informatics students to engage computer science problems, broadly, through varied modes of thinking. They argue that computer science "is inherently social and no social aspects can be meaningfully separated from CS" (p. 60). We also note that identifying ethical and social dilemmas is becoming a part of explicit learning goals in a growing number of courses in Scandinavian universities.

Second, professional associations are currently developing standards for considering human values in the development process. The IEEE P7000 standard, for example, seeks to insert specific milestones related to human values into best-practice software engineering processes (IEEE Standards Association, 2019a). Related to P7000, is the work on "ethically-aligned design," which, in the design of autonomous and intelligent systems, seeks to prioritize human well-being in extension to engineering and economic values (IEEE Standards Association, 2019b). Another example deriving from working with a value-based approach is the ISO 21801 Guideline on Cognitive Accessibility in technical systems (ISO/TC173, 2019), which can be used as a situated complement to the more technically focused WCAG 2.1 Web Content Accessibility Guideline (W3C, 2017). A critique of a standardisation approach is that it ignores the context and the situation; instead, it is better suited to planning and regulation than design. However, with ISO21801, a focus on people and processes complement the technical aspects.

These and many other related developments (e.g., Ethical CS, Values in Computing, 2019; Values and Ethics in Responsible Technology in Europe, 2019; Value Sensitive Design in Higher Education, 2019; etc.) seem to indicate an inflection point in engineering and technical education, where materials related to the social impacts of technology are embedded in technical education.

PANEL QUESTIONS

Pedagogical practices

1. How have you placed value sensitive design within technical education?
2. What challenges have you encountered, what discoveries have you made?
3. How have you taught theory and specific methods?
4. How do you adjust to different level of expertise in value sensitive design (beginner/advanced)?
5. How do you conceptualize and teach "skilful practice?"
6. What VSD-based artefacts can be developed to embed skilful practice and progress teaching and learning outcomes?

Curriculum design

1. How do social, ethical, and policy considerations impact the curriculum?

2. How do new and emerging engineering standards impact the curriculum?
3. By engaging with values, what is given up, what is gained?
4. What are the learning goals?
5. How does the field cultivate a community of practice, focused on value sensitive design?

Organisational and learning cultures

1. How should the humanities, social sciences, and engineering be organized to support educational goals in human values and engineering?
2. How will teaching incentives, rewards, and support need to be restructured?

FORMAT: THREE PHASES, FOCUSED ON AUDIENCE ENGAGEMENT

Phase-I: Seed presentations (30-40 minutes)

Panelists will give brief presentations of their experience with value sensitive design education and give a provocative claim or question.

Phase-II: Audience questions (45-55 minutes)

Questions from the audience will be elicited. To keep the questioning dynamic and interesting, the panel chair will have set of questions for the panelists and audience, ready to ask.

Phase-III: Summary (5 minutes)

The panel chair will seek a synthetic summary of the audience questions and comments, drawing on the above set of questions. Next steps and prospects for the future will be emphasized.

PANEL MEMBERS

The 4-5 panel members will represent educators from North America, Europe, and Australia, and represent various areas of experience in teaching value sensitive design, such as PhD supervision, entry levels at bachelor, and master levels in fields such as engineering, interaction design and computer science. The chair of the panel will be Daisy Yoo.

KEYWORDS: value sensitive design; engineering and technical design; education; skillful practice; ethics; moral reasoning and technology.

REFERENCES

Association of Computer Machinery (2018). *ACM Code of Ethics and Professional Conduct*. Retrieved from https://www.acm.org/binaries/content/assets/about/acm-code-of-ethics-and-professional-conduct.pdf

Ethical CS (2019). *Ethical CS*. Retrieved from https://ethicalcs.github.io/

Frauenberger, C., & Purgathofer, P. (2019). Ways of thinking in informatics. *Communications of the ACM*, *62* (7), 58-64. DOI: https://doi.org/10.1145/3329674

Friedman, B. & Hendry, D. G. (2019). *Value Sensitive Design: Shaping Technology with Moral Imagination*. Cambridge, MA: MIT Press.

Friedman, B., Hendry, D. G., & Borning, A. (2017). A survey of value sensitive design methods. *Foundations and Trends in Human-Computer Interaction*, *11* (23), 63-125.

Grosz, B. J., Grant, D., G., Vredenburgh, K., Behrends, J., Hu, L., Simmons, A. & Waldo, J (2019). Embedded EthiCS: integrating ethics across CS education. Co*mmunications of the ACM*, 62(8), 54-61. DOI: https://doi.org/10.1145/3330794

van den Hoven, J. (2013). Value sensitive design and responsible innovation. In *Responsible Innovation* (pp. 75–83). John Wiley & Sons, Ltd, 2013. ISBN 978-1-118-55142-4.

IEEE Standards Association (2019a). *IEEE P7000: Model Process for Addressing Ethical Concerns During System Design*. Retrieved from https://standards.ieee.org/project/7000.html

IEEE Standards Association (2019b). *EEE Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems*. Retrieved from https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead_v2.pdf

ISO/TC173, ISO 21801-1 (2019). *Cognitive accessibility -- Part 1: General guidelines*.

Massachusetts Institute of Technology (2019). MIT Schwarzman College of Computing. Retrieved from https://computing.mit.edu

Shah, J. A, & Nobles, M. (2019, August 5). *MIT Schwarzman College of Computing Task Force Working Group on Social Implications and Responsibilities of Computing Final Report*. Cambridge, MA: MIT.

Spiekermann, S. (2015). *Ethical IT Innovation: A Value-Based System Design Approach*. Boca Raton, FL: Auerbach Publications.

Stanford University (2019). Human-Centered Artificial Intelligence. Retrieved from https://hai.stanford.edu.

Values in Values in Computing (2019). Retrieved from http://www.valuesincomputing.org/.

Values and Ethics in Responsible Technology in Europe (2019). Retrieved from https://blogit.itu.dk/virteuproject/what-is-virt-eu/.

Value Sensitive Design in Higher Education (2019). Retrieved from https://vase.mau.se/.

UT Delft (2019). Retrieved from https://www.tudelft.nl/ethics/ .

W3C, “WCAG, Web Content Accessibility Guidelines 2.1,” (2017). Retrieved from: https://www.w3.org/TR/WCAG21/.

VALUES AND POLITICS OF A BEHAVIOR CHANGE SUPPORT SYSTEM

Janet Davis, Buyaki Nyatichi

Whitman College (USA)

[davisj,nyaticmb]@whitman.edu

EXTENDED ABSTRACT

Just Not Sorry is a Gmail plug-in that highlights when the user writes words of apologies such as "sorry," hedge words such as "just," and intensifiers such as "very" (Def Method, 2019b). Red underlines appear as if the words had been misspelled. For each underline, a motivational quote appears as a tooltip. One such quote follows as an example: "Using 'sorry' frequently undermines your gravitas and makes you appear unfit for leadership - Sylvia Ann Hewlett."

In this case study, we consider *Just Not Sorry* as an example of a persuasive technology—that is, a technology designed to change attitudes and behaviours (Fogg, 2002). *Just Not Sorry* came to our attention through a survey of technologies designed to influence speech and writing (Twersky & Davis, 2017). When we describe *Just Not Sorry* to others, it elicits strong and opposing reactions. Some say, "I need that!" while others say, "I would never use a tool like that." Our research question for this case study: What might explain such strong, opposing reactions?

We adopted value sensitive design (Friedman, Kahn, & Borning, 2006) as our guiding theory and methodology, intertwining conceptual, technical, and empirical investigations. In our initial conceptual investigations, we considered ethical principles, stakeholders, and implicated values. In parallel technical investigations, we each installed *Just Not Sorry* and used it for at least one month. We examined the source code on GitHub (Def Method, 2019a), as well as the system image presented in the Chrome Web Store (Def Method, 2019b). To understand the intentions behind *Just Not Sorry*, we also read Tami Reiss's (Reiss, 2015) story of the tool's conception and design. Finally, in our empirical investigations, we searched the Web for the keywords "gmail 'just not sorry.'" Amongst over 100 articles that mention *Just Not Sorry*, we identified 27 that were published in blogs, magazines, or newspapers considered notable by *Wikipedia* and that express an opinion about the tool. We are in the process of coding these opinions to confirm or augment our analysis of stakeholders and values.

First we consider direct and indirect stakeholders, and implicated values.

Just Not Sorry was inspired by a conversation among women in leadership positions (Reiss, 2015). They talked about recent satire that exaggerated women's stereotypical overuse of words such as "just" and "sorry." According to Reiss, the group agreed these stereotypes were true: "The women in these rooms were all softening their speech in situations that called for directness and leadership. We had all inadvertently fallen prey to a cultural communication pattern that undermined our ideas" (Reiss, 2015). Reiss then proposed a tool to help women change their behavior. In this way, *Just Not Sorry* was designed by and for women, as a tool to promote the status of women. However, the description of *Just Not Sorry* on the Chrome Web Store (Def Method, 2019b) does not mention gender. Some commentators point out that men can use the extension as well (e.g., Erikson, 2016), while others do not mention gender (e.g., Ye, 2016).

Therefore, while gender and thus *identity* is highly salient to the design of *Just Not Sorry*, direct stakeholders include both men and women. We find that *Just Not Sorry* is designed to enhance users' *achievement* and *social power* through its support for behavior change. From the Chrome Web Store

description, as well as our inspection of the tool, we see that *Just Not Sorry* also supports users' *autonomy* in that it suggests changes without requiring them. It protects users' *privacy* in that it ensures recipients cannot tell that *Just Not Sorry* was used to compose the email.

Indirect stakeholders include email recipients, colleagues, and feminists. Email recipients may perceive an email composed in accordance with the suggestions of *Just Not Sorry* as rude or abrupt. As observed by several commentators (e.g., Minter, 2016), the value of politeness or *courtesy* is thus implicated. If *Just Not Sorry* does indeed promote workplace success, colleagues who do not use the tool may find their own achievement and social power lessened. Changes in communication style across a workplace or industry may enhance either *collaboration* or *competition*.

Finally, while still more indirect, feminists constitute a large stakeholder group with a substantial interest in *Just Not Sorry*. Reiss (2015) clearly had feminist motivations for creating *Just Not Sorry.* Bucholtz (2014) offers the following definition of feminism:

> A diverse and sometimes conflicting set of theoretical, methodological, and political perspectives that have in common a commitment to understanding and challenging social inequalities related to gender and sexuality.

Hence, considering feminists as a stakeholder group implicates the value of *equality* with respect to identity. Of the 27 opinions we read, all but a few address gender, and over half address sexism, feminism, or equality.

Having identified feminists as a key stakeholder group, we refined our research question as follows: Could conflicting perspectives on gender equality explain strong, opposing reactions to *Just Not Sorry?*

We find that it is one particular kind of feminism that motivates *Just Not Sorry*. Reiss (2015) seeks to enhance women's position within existing structures by helping them address a perceived deficiency in their attitudes and behaviors. This is a textbook example of a liberal feminist approach: "Given its concern to bring women into men's spheres, liberal feminism has generally aimed to eradicate gender inequality by eradicating or at least reducing gender difference" (Bucholtz, 2014). Gill & Orgad (2017) further identify *Just Not Sorry* with the contemporary movement they call "confidence culture," in which women's failure to achieve equality in the public sphere is attributed to individual shortcomings that can be addressed through projects of self-improvement. Where Bucholtz (2014) finds that liberal feminism "is often no longer recognized as feminism at all," in the present day Gill and Orgad (2017) describe confidence culture as a popular, postfeminist remaking of feminism. We find that some supporters of *Just Not Sorry* (e.g., Lastoe, 2016) are writing from the taken-for-granted perspective of liberal feminism and confidence culture.

While liberal feminism seeks equality through the reduction of gender differences, cultural feminism views women's communication styles as distinctive and as having their own value. Bucholtz (2014) divides cultural feminism further into liberal cultural feminism and radical cultural feminism. "Liberal cultural feminism seeks acknowledgment of the equal value of what are seen as women's distinctive practices," Bucholtz writes, while radical cultural feminism "elevates women's practices over men's." We find examples of both perspectives among those who criticize *Just Not Sorry* in the popular media (Cauterucci, 2015; Minter, 2016).

Liberal and cultural feminism are the most commonly espoused forms of feminism outside of academia (Bucholtz, 2014), so it is not surprising that these are the viewpoints most easily found represented in the popular media. In our ongoing analysis, we also seek to identify references to concepts from other feminist approaches discussed by Bucholtz, such as the radical feminist concept of patriarchy, to

intersectionilty with respect to class or race, and to gender as performative rather than essential. We also seek to identify critiques of *Just Not Sorry* that do not come from a feminist perspective, but address other values such as autonomy.

In conclusion, we contribute a value sensitive design case study of a persuasive technology concerned with language and gender. We appeal to differing conceptions of feminism and gender equality to understand some of the controversy surrounding the technology. For our empirical investigation, we conduct a qualitative analysis of written opinions in the popular media, an approach which we do not believe has been previously used under the value sensitive design framework. We were able to take this approach because we are conducting a retrospective analysis of an existing tool, rather than prospective design of a new tool. However, we believe that analysis of popular media opinions may be a fruitful and low-cost approach for value sensitive design of other persuasive technologies, as public discussion of problematized behaviors may often precede or even inspire the design of new behavior change support systems.

KEYWORDS: Value sensitive design, persuasive technology, email, language, gender, feminism.

REFERENCES

Bucholtz, M. (2014). The Feminist Foundations of Language, Gender, and Sexuality Research. In S. Ehrlich, M. Meyerhoff, & J. Holmes (Eds.), *The Handbook of Language, Gender, and Sexuality* (pp. 23-47). John Wiley & Sons.

Cauterucci, C. (2015, December 29). New Chrome App Helps Women Stop Saying “Just” and “Sorry” in Emails. Retrieved June 12, 2019, from Slate Magazine website: https://slate.com/human-interest/2015/12/new-chrome-app-helps-women-stop-saying-just-and-sorry-in-emails.html

Def Method. (2019a). *Defmethodinc/just-not-sorry* [JavaScript]. Retrieved from https://github.com/defmethodinc/just-not-sorry

Def Method. (2019b, March 28). Just Not Sorry—The Gmail Plug-in (v1.6.0). Retrieved December 2, 2019, from Chrome Web Store website: https://chrome.google.com/webstore/detail/just-not-sorry-the-gmail/fmegmibednnlgojepmidhlhpjbppmlci?hl=en-US

Erikson, J. (2016, January 5). Sorry, but Gmail’s New Plug-in Isn’t Just for Women. Retrieved December 11, 2019, from CafeMom website: https://thestir.cafemom.com/good_news/194598/sorry_but_gmails_new_plugin

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

Friedman, B., Kahn, P. H., & Borning, A. (2006). Value sensitive design and information systems. In P. Zhang & D. Galletta (Eds.), *Human-computer interaction in management information systems: Foundations,* (pp. 348–372). Armonk, New York; London, England: M.E. Sharpe.

Gill, R., & Orgad, S. (2017). Confidence culture and the remaking of feminism. *New Formations*, *91*(91), 16–34. https://doi.org/10.3898/NEWF:91.01.2017

Lastoe, S. (2016, January 4). You’re Actually Not Sorry—And This App Will Help You Stop Saying it So Much in Email. Retrieved June 12, 2019, from The Muse website: https://www.themuse.com/advice/youre-actually-not-sorryand-this-app-will-help-you-stop-saying-it-so-much-in-email

Minter, H. (2016, January 14). The Just Not Sorry app is keeping women trapped in a man's world. *The Guardian*. Retrieved from https://www.theguardian.com/women-in-leadership/2016/jan/14/the-just-not-sorry-app-is-keeping-women-trapped-in-a-mans-world

Reiss, T. (2015, December 22). Just Not Sorry! (The backstory). Retrieved June 12, 2019, from The Medium website: https://medium.com/@tamireiss/just-not-sorry-the-backstory-33f54b30fe48

Twersky, E., & Davis, J. (2017). "Don't Say That!" *Persuasive Technology*, 215–226. Springer.

Ye, L. (2016, January 13). The "Just Not Sorry" App Will Improve Your Sales Emails in 5 Seconds Flat. Retrieved December 11, 2019, from HubSpot website: https://blog.hubspot.com/sales/just-not-sorry

VALUES IN PUBLIC SERVICE MEDIA RECOMMENDERS

Maaike Harbers, Lotte Willemsen, Paul Rutten

Rotterdam University of Applied Sciences (The Netherlands)

m.harbers@hr.nl; l.m.willemsen@hr.nl; p.w.m.rutten@hr.nl

EXTENDED ABSTRACT

1. INTRODUCTION

Recommendation systems, recommenders in short, selecting and filtering content are widely used by companies in order to provide suggestions for items to users (Ricci et al., 2011). These 'items' range from songs (e.g., Spotify), series (e.g., Netflix), and movies (e.g., YouTube) to messages (e.g., Facebook), job vacancies (e.g., LinkedIn) and products (e.g., Amazon). Public Service Media (PSM) organizations, publicly funded organizations that offer radio and television content to a general audience, can also benefit from recommenders by using them to bring their audience in contact with new content. However, whereas recommenders used by commercial parties often aim to maximize profit or engagement, which is often achieved by recommending items in line with the user's views and interests, PSM organizations have other goals, such as informing the public and exposing them to a balanced mix of different views and perspectives, that could conflict with these commercial recommendation practices. The European Broadcast Union (EBU) acknowledges the tension between serving the audience with recommenders and the responsibilities of PSM organizations (EBU, 2017).

Recently, increasing attention has been paid to the development of recommenders for PSM (Sørensen et al., 2017; Fields et al., 2018; Van den Bulck et al., 2018; Sørensen, 2019). However, though the need for PSM recommenders is acknowledged, research into their design and development is still in its infancy. One of the open questions is what metrics (e.g., diversity or serendipity) PSM recommenders should optimize for (Fields et al., 2018). As a first step towards answering this question, following a Value Sensitive Design (VSD) approach (Friedman et al., 2019), this extended abstract describes a value source analysis (Friedman, 2017), in which an overview of the most important values at stake in the design of PSM recommenders is provided, including a description of where these values come from. The overview is based on a literature study and empirical investigations performed at NPO, the Dutch national public broadcasting organization (NPO, 2019a). Furthermore, some observations regarding the (value-sensitive) design of information systems in general are made.

2. VALUES AT STAKE - LITERATURE

The first set of values relevant to PSM recommenders can be found in literature on PSM. One of the most prominent lists of values for PSM is provided by UNESCO, consisting of universality, diversity, independence and distinctiveness (UNESCO, 2001). *Universality* refers to the accessibility of media content to all citizens in the country, *diversity* involves diversity in content, audience targeted, and subjects discussed, *independence* involves the freedom to express ideas and circulate information, and *distinctiveness* refers to the distinction of one PSM organization from other media organizations.

In addition to PSM values, there are values related to the use of the technology underlying recommenders. As public organizations such as PSM generally have the goal to 'serve the public', they often take public values into account (Jørgensen et al., 2007). Multiple values have been identified as

relevant to the responsible design of information systems (Winkler et al., 2019). In relation to recommenders, most notably, this could mean a responsible use of personal data to protect *privacy* (Hoepman, 2014), and responsible use of machine learning, a technology often used in recommenders, supporting the values of values of *fairness*, *accountability* and *transparency* (ACM FAT).

3. VALUES AT STAKE - IN PRACTICE

We studied values at stake in PSM recommenders in a real-world setting at NPO, the organization that oversees public broadcasting services in the Netherlands. One of the ways in which NPO brings content produced by public broadcasters to the Dutch audience is via its website NPO Start (www.npostart.nl), which makes limited use of a recommendation algorithm. Most of the recommendations on NPO Start are manually curated, but for website visitors with an account (the minority of the visitors), a small part of the recommendations is personalized and generated by an algorithm. NPO is currently working on improving and expanding their recommender. For our study, we attended several meetings at NPO in which the design of the new recommender was discussed, conducted interviews with stakeholders within and out of NPO, and studied project documentation and reports produced by NPO.

NPO's mission is to connect and enrich the Dutch audience with content that informs, inspires and entertains (NPO, 2019a, 2019b), which is broadly in line with the PSM values described in the previous section. Project documentation showed that the most prominent value in the recommender design project was *pluriformity* (the 'explicitly supported value' in VSD terminology). In meetings, a lot of time was spent on discussing what, exactly, pluriformity means with respect to the recommender to be designed. Other values that came up during the meetings were accuracy, privacy and transparency. *Accuracy* of recommendations was deemed important, as users receiving too many recommendations that are not interesting to them would disengage. With respect to *privacy*, it was agreed upon that the recommender should not collect explicit personal information such as age, gender or ethnicity, but only use watching behavior. *Transparency* to users about the origin of recommendations was also deemed important.

In an interview with the head of the development team, responsible for the implementation of the recommendation algorithm (and also part of the project team), we learned that the current algorithm weights five factors: novelty, clickthrough rate, personalization, fraction watched and public values. The last factor, public values is composed of users' ratings of content based on eight values, out of which one is pluriformity (p.62, NPO, 2019c). There is thus a discrepancy between the focus on pluriformity in the redesign project and the (minor) role of pluriformity in the current recommender. With respect to the planned increased importance of pluriformity, the development team neither knew how to translate pluriformity into an implementation, nor did they see it as their responsibility.

Interviews with users, people who watch content produced by public broadcasters on NPO Start, revealed that the majority of users is interested in personalized recommendations, but that most of them were not or only vaguely familiar with the term pluriformity.

4. DISCUSSION

Several insights can be drawn from the results so far. There are several values that the organization wants to embed in the new recommender, most notably pluriformity. Yet, problems are encountered in translating these values into a concrete implementation. Whereas the development team refers to others to operationalize pluriformity so that they can implement the algorithm, other members of the project team have trouble providing such an operationalization, partly because they have limited

programming knowledge and have troubles imagining what developers need. At the same time, the term pluriformity does not appeal to users of the recommender, which may be problematic in providing transparency (another value at stake) about the system. These differences seem to indicate a mismatch between knowledge, culture and languages spoken by different groups of people: (most members of) the project team, developers and users.

A mismatch in understanding of the design challenge and its implicated values between teams in the organization is possibly reinforced by an organizational structure in which employees with different expertise and backgrounds are organized different teams. This is problematic when the goal is to embed values in technology. For example, embedding the value of transparency in a recommender has implications for both the recommender's algorithm and its user interface. On the technical, backend side, the algorithm should be explainable, which may imply avoiding certain deep learning algorithms (Samek et al., 2017). On the user-facing, front-end side, there should be a way to communicate explanations to users in the interface, e.g. a textbox or a button for requesting an explanation for why an item was recommended (Tintarev et al., 2011). If the system does not meet requirements on both of these sides, it will not support transparency. In order to align different components of a system, teams responsible for the creation of these different components need to be aligned as well.

The insights above lead to a more general observation. In VSD analyses, when describing value implications, 'technology' is often treated as a single system and 'the designer' is often treated as a single role (Friedman et al., 2019). However, this is a simplification of (the creation of) a lot of technologies, as systems often consist of different components, which are developed by different teams, consisting of a variety of individuals, with different backgrounds and cultures. We believe that a VSD process could benefit from a more nuanced view on the 'technology' and 'designer', doing justice to their complexities. This may be particularly relevant for complex and intelligent systems, which have a heavy technical component, as well as a user interface.

5. FUTURE WORK

This paper forms a first step towards designing a PSM recommender. Next steps involve analyzing value tensions (Miller et al., 2007); selecting metrics based on these values (Fields et al., 2018); operationalizing these metrics, including weighing them against each other; and designing and evaluating prototypes. This process will be performed iteratively, involving multiple cycles of prototyping and collecting user feedback. In this process, attention will be paid to the multifaceted nature of recommenders as well as their designers.

KEYWORDS: Public service media, recommendation system, recommender, values, value sensitive design, pluriformity.

ACKNOWLEDGEMENTS: the authors thank npo for their collaboration and support in this research project. This extended abstract reflects the authors' interpretations of information and events, and the authors are solely responsible for the contents of this extended abstract.

REFERENCES

ACM FAT. Conference on Fairness, Accountability, and Transparency. Retrieved from: https://fatconference.org/.

EBU (2017). *Big data initiative report: time to invest.* Technical report. European Broadcast Unit (EBU).

Fields, B., Jones, R., & Cowlishaw, T. (2018). The case for public service recommender algorithms. Proceedings of *FATREC Workshop on Responsible Recommendation*.

Friedman, B., & Hendry, D. G. (2019). *Value sensitive design: Shaping technology with moral imagination*. Mit Press.

Friedman, B., Hendry, D. G., & Borning, A. (2017). A survey of value sensitive design methods. *Foundations and Trends® in Human–Computer Interaction*, 11(2), 63-125.

Hoepman, J. H. (2014). Privacy design strategies. In *IFIP International Information Security Conference* (pp. 446-459). Berlin, Heidelberg: Springer.

Jørgensen, T. B., & Bozeman, B. (2007). Public values: An inventory. Administration & Society, 39(3), 354-381.

Miller, J. K., Friedman, B., Jancke, G., & Gill, B. (2007). Value tensions in design: the value sensitive design, development, and appropriation of a corporation's groupware system. In *Proceedings of the 2007 international ACM conference on Supporting group work* (pp. 281-290). ACM.

NPO (2019a). Nederlandse Publieke Omroep. Retrieved from: https://over.npo.nl/.

NPO (2019b). Jaarverslag 2018. Annual report. Retrieved from: https://over.npo.nl/organisatie/onze-waarde-voor-nederland/jaarverslag.

NPO (2019c). Terugblik 2018. Technical report. Retrieved from: https://over.npo.nl/organisatie/onze-waarde-voor-nederland/terugblik.

Ricci, F., Rokach, L., & Shapira, B. (2011). Introduction to recommender systems handbook. In *Recommender systems handbook* (pp. 1-35). Boston, MA: Springer.

Samek, W., Wiegand, T., & Müller, K. R. (2017). Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models. arXiv preprint arXiv:1708.08296.

Sørensen, J. K. (2019). Public Service Media, Diversity and Algorithmic Recommendation: A Europe-wide Implementation Study. In *RecSys 2019: 13th ACM Conference on Recommender Systems*.

Sørensen, J. K., & Hutchinson, J. (2017). *Algorithms and public service media*. Public Service Media in the Networked Society RIPE, 91-106.

Tintarev, N., & Masthoff, J. (2011). Designing and evaluating explanations for recommender systems. In Recommender systems handbook (pp. 479-510). Springer, Boston, MA.

UNESCO (2001). *Public broadcasting: Why? How?* Technical Report (pp. 1-28). Paris: UNESCO.

Van den Bulck, H., & Moe, H. (2018). Public service media, universality and personalisation through algorithms: mapping strategies and exploring dilemmas. *Media, Culture & Society*, 40(6), 875-892.

Winkler, T., & Spiekermann, S. (2019). Human Values as the Basis for Sustainable Information System Design. *IEEE Technology and Society Magazine*, 38(3), 34-43.

10. Monitoring and Control of AI Artifacts

Track chair: Yukari Yamazaki, Seikei University, Japan

AN EMPIRICAL STUDY FOR THE ACCEPTANCE OF THE ORIGINAL NUDGES AND HYPERNUDGES

Yukari Yamazaki

Seikei University (Japan)

yyamazak@econ.seikei.ac.jp

EXTENDED ABSTRACT

Over the past decade, the notions of behavioural economics, especially for nudges have been applied to public policy making and medical judgement. Behavioural science teams can be found in dozens of countries, including Australia, the Netherlands, Canada, Ireland, Denmaadrk, Mexico, Germany, and Qatar. There is a great deal of activity elsewhere, particularly in the regulatory domain (OECD, 2017).

Despite nudge (Thaler and Sunstein, 2008) that steer peoples' behaviour in desirable directions through milder choice interventions has drawn attention to, it also has received blistering critiques of ethicality (i.e., Goodwin, 2012), diminishing human wisdom (Furedi, 2011), the troubles and pitfalls (Bovens, 2009), and manipulations (Wilkinson, 2012). In response to these misunderstood critiques, Sunstein who is one of the advocates of the nudge has discussed the validity and considered the beneficence and ethicality of nudges (Sunstein, 2015; 2016). According to his consideration, there are neither neutral ways to present options, nor choices made in a vacuum, and one cannot avoid the choice architectures which influence choice in many ways. It might be easy to promote purchasing by altering the presentation order of alternatives and attributes, easiness to pick them up, and the selection of defaults, as well as naming just a few of the design options available.

Several studies have surveyed the acceptance, trustiness, and consensus for variety types of nudges in various countries (i.e., Sunstein et al., 2018). These surveys have appeared that, on one hand, citizens in various countries hesitated the nudges that perceive to be inconsistent with their interests or values of most choosers (Reisch and Sunstein, 2016), on the other hand, they generally tend to approve of almost all nudges. The contexts of health and safety nudges would be approved for people and the levels of acceptance of nudging techniques depended on the countries of participants, as well as the depth, types, contexts, and prosociality of nudges.

Back to the ethicality and beneficence of nudges, quoting Sunstein's consideration (2015), "when nudges are fully transparent and subject to public scrutiny, a convincing ethical objection is less likely to be available." In addition, it is also stated that, "if people have not consented to them; such nudges can undermine autonomy and dignity" (p.1). Furthermore, Sunstein (2018) insisted that some notorieties of nudges, such as an excessive trust in government and the ideas that nudges are covert, are manipulative, and exploit human behavioural biases because of irrationality are misconceptions. He said, "Nudges always respect, and often promote human agency; because nudges insist on preserving freedom of choice, they do not put excessive trust in government; nudges are generally transparent rather than covert or forms of manipulation" (p.1). According to the above considerations, two of the prominent elements in ethical nudge should be transparency and autonomy.

Recently artificial intelligence and machine learning (AI/ML) has drawn attention amongst mass media and academic fields not only because of attractive, tremendous, and hyper functions as well as efficiency and effectiveness, but also the ethicality and riskiness. The IEEE, for example, has taken the ethicality of AI designing, utilizing, and prevalence as serious problem and given an alert for AI systems

as nudging tools. (IEEE, 2018). In addition, nudges through AI/ML-driven new technologies are coined as “hypernudges” (Yeung, 2017) or “digital nudges” (Weinmann et al., 2016). While it has gradually appeared that AI/ML systems have several beneficial traits, there are several specific features as hypernudges. Some of them are, for example, self-tracking of past behaviour (in some cases) without getting agreement for utilizing it, presenting immediate feedback based on self-tracking and big data as recommendations for each user, making some judgment behalf on human autonomy, and steering people to use the AI artefacts repeatedly. These typical features with AI/ML-driven hypernudges might make users blind and depend too much on them. Therefore, the manipulative aspects of data-driven personalized communication, big date utilization, and behavioural targeting in the online realm has been regarded as problems (e.g., Lanzing, 2018).

In another argument, however, whereas various services by AI/ML such as vehicle navigation systems, position information of digital map, recommendation based on purchase history, personalized chatting with bots, various apps, and so on have spread amongst people recently, it is hard to ascertained these new nudges as the original ethical nudges which has discussed the validity in Sunstein (2015) and (2016) because of lack of autonomy and transparency (Yamazaki, 2019).

As noted, the acceptance of the original nudges has been surveyed, it has still been veiled the acceptance and consensus for hypernudges. Therefore, based on the prior studies which surveyed the acceptance of the original nudges, this study investigated and compared the acceptance levels for the original nudges and hypernudges focusing on the difference of sociodemography of participants, as well as the depth (campaign or mandatory), types (conscious or unconscious), contexts (i.e., health and ecology), and prosociality (toward personal or social) of nudges.

Participants were asked 16 questions that half of them were the original nudges and the other half were hypernudges. Responses were given as “agree” or “disagree” and some comments with nudges by AI, optionally.

The results indicated that overall, the original nudges were more accepted than hypernudge, and the acceptance level of the original nudge and hypernudge was different in the sorts of nudges. As for comparison with the original nudges and hypernudges for acceptance percentages, except 4 pairs of nudges (No. 8, 10, 13, and 16), the majority support (8 pairs) of the original nudges have observed against to hypernudges (Table 1).

Table 1 Cross Tabulation with each nudge pairs

%	O1/H1		O2/H2		O3/H3		O4/H4		O5/H5		O6/H6		O7/H7		O8/H8	
Agree	62.7	37.3	54.7	45.3	55.8	44.2	52.6	47.4	50.3	49.7	59.2	40.8	70.2	29.8	43.2	56.8
Disagree	22.0	78.0	42.8	57.2	38.4	61.6	46.8	53.2	49.8	50.2	37.5	62.5	24.2	75.8	56.4	43.6
χ^2	169.062**		16.243**		32.007**		3.930*		.030		54.576**		248.254**		20.964**	

%	O9/H9		O10/H10		O11/H11		O12/H12		O13/H13		O14/H14		O15/H15		O16/H16	
Agree	53.6	46.4	47.4	52.6	49.1	50.9	57.2	42.8	47.2	52.8	48.6	51.4	53.0	47.0	48.3	51.7
Disagree	46.2	53.8	54.0	46.0	50.8	49.2	40.2	59.8	53.1	46.9	50.8	49.2	48.8	51.2	54.9	45.1
χ^2	6.500*		5.065*		.336		33.710**		4.119*		.518		1.799		3.974*	

Chi-square value significant at alpha * $p < 0.05$

** $p < 0.01$

The results drew by this study indicated that the approval or disapproval for hypernudges were dramatically different from the original nudges. This suggested a kind of alert for the introduction, utilization, and spreading of hypernudges. We should consider how depth and what types, contexts, and prosociality of hypernudges would be acceptable for uses, as well as what kinds of invisible influences would occur by hypernudges. This study might serve as an onset in prevalence for AI-driven artefacts which takes into consideration the relevant speculation of AI acceptance and its riskiness.

KEYWORDS: AI-driven artefact, original nudge, hypernudge, acceptance.

REFERENCES

Bovens, L. (2009). The Ethics of Nudge, In T. Yanoff-Grüne and S. O. Hansson. (Eds.) *Preference Change* (pp. 207–209). Dordrecht: Springer.

Bruns, H., Kantorowicz-Reznichenko, E., Klement, K., Jonsson, M. and Rahali, B. (2018). Can Nudges Be Transparent and Yet Effective? *Journal of Economic Psychology*, 65, 41-59.

Furedi, F. (2011). "Defending moral autonomy against an army of nudgers", *Spiked*. Retrieved from http://tinyurl.com/6kfafka

Goodwin, T. (2012). Why we should reject 'nudge' Policy and Politics, 41(2), 159–182.

IEEE standard association (2018). Affective Computing, Ethically Aligned Design, ver. 2nd., 162-181. Retrieved from https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead_v2.pdf

Lanzing, M. (2018). Strongly Recommended" Revisiting Decisional Privacy to Judge Hypernudging in Self-Tracking Technologies, *Philosophical Technolog*y, 32(3), 549–568.

OECD (2017). Behavioural Insights and Public Policy: Lessons from Around the World. Paris: OECD Publishing.

Reisch, L. A. and Sunstein, C. (2016). Do Europeans Like Nudges?, *Judgment and Decision Making*, 1(4), 310–325.

Sunstein, C. R. (2015). *Nudging and Choice Architecture: Ethical Considerations*, (Harvard John M. Olin Discussion Paper Series Discussion Paper No. 809, Jan. 2015, Yale Journal of Regulation.

Sunstein, C. R. (2016). The Ethics of Influence: Government in the Age of Behavioral Science. CUP, New York.

Sunstein, C. R., Reisch, L. A. and Rauber, J. (2018). A worldwide consensus on nudging? Not quite, but almost, *Regulation & Governance*, 12, 3–22.

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

Yamazaki, Y. (2019). Certified Requirements of Nudging by AI Artifacts: Hypernudges Driven by AI/ML Artifacts will not be Recognized as Nudges, *XXVIII AEDEM International Meeting Tokyo*.

Yeung, K. (2017). "'Hypernudge': Big Data as a mode of regulation by design, *Information Communication and Society*, 20(1), 118-136.

Weinmann, M., Schneider, C. and Brocke, J. V. (2016). Digital Nudging, *Business and Information Systems Engineering*, 58(6), 433-436.

White, M. D. (2008). Behavioral law and economics: the assault on consent, will, and dignity. In G. Gaus, C. Favor, & J. Lamont (Eds.), *Essays on philosophy, politics and economics: Integration and common research projects* (pp. 201–224). Palo Alto, CA: Stanford University Press.

Wilkinson, T. M. (2012). Nudging and Manipulation, *Political Studies*, 61(2), 341-355.

APPROACH TO LEGISLATION FOR ETHICAL USES OF AI ARTIFACTS IN PRACTICE

Yasuki Sekiguchi

Hokkaido University (Japan)

seki@econ.hokudai.ac.jp

EXTENDED ABSTRACT

Ethics principles or guidelines of artificial intelligence (AI) or AI artifacts are discussed broadly and proposed by EC (AI HLEG, 2019), IEEE (2019) and others. R&D and production of AI artifacts are expected to comply with these ethics requirements. However, any technological artifacts, regardless of how ethically they are designed, developed and implemented (i.e., supply-ethics), can be used both ethically and unethically. In order to keep the coming AI society ethical, ethical use of AI artifacts must be maintained in practice (i.e., use-ethics). This implies that legislation and regulation on practical uses of AI artifacts are necessary. Thus, we need to find out focal points needing legislation and regulation. It is our objective here to propose some essential elements to do this.

Our living space is sometimes labelled as “real world (RW)”. It is the physical world we live and recognize. The space that is described in data and processed with the information and communication technology is labelled as “cyber world (CW)”. These two worlds have been different and separate in the sense that the CW was in the outside of our everyday life. We “used” objects in the CW. This is why ethics in the CW have not been worried about up to recently.

The internet and mobile devices have changed the situation. It seems not rare that an individual consumes a few hours a day to communicate with social network services, roll playing games, smart speakers, etc. The CW has become an inevitable part of our everyday life. Thus, our living space has become a mixture of the RW and the CW, i.e., the mixed world (MW). The MW must be ethical as the RW is. This is why legislation and regulation for practical uses of AI artifacts are needed.

Remember then that the currently existing acts and regulations barely keep the RW ethical. Because the MW is an extension of the RW into the CW, it seems reasonable to expect some extension of the existing acts and regulations would be effective too for keeping the MW ethical. This is one of points proposed here.

AI artifacts should be defined so as to make it easy to find out when they are used in practical situations. Considering the fact that the term “AI” has been defined and used variously, when we simply use “AI”, it implies a program or a machine which works in ways that if a person behaves similarly he or she is called intelligent (Floridi & Cowls, 2019). More specifically, terms used in this study are defined as follows:

An AI artifact is one that cannot function without at least one “AI system”. An AI system is an information system which has at least one “learning&etc system” and it is a system made by adding an application program interface to its learning&etc systems. A learning&etc system is a program of some learning&etc algorithms and includes input and output databases for the algorithms. A learning&etc algorithm makes it possible to change its own outputs or programs in the process of the utilization, by learning from data, information and knowledge, by inferences based on them, and/or by interactions with the environment through sensors, actuators etc. An AI artifact behaves

intelligently because of the learning&etc algorithms composing its AI system (The Conference toward AI Network Society, 2017).

An AI artifact might use plural AI systems, and their effects on ethics in the MW might be different from each other. Therefore, when ethical effect of an AI artifact is assessed, each AI system in it must be separately evaluated. This is also one of our proposals.

Developers of an AI artifact suppose its use context in order to specify technological requirements. A resultant AI artifact might not be used in the supposed context, but often used in different contexts. The supply-ethics by R&D is effective only in the supposed context, and the AI artifact in practical use can happen to have unethical effects on the MW.

The situation of the traditional technological artifacts such as motor vehicles and services based on them is similar. Some legislation and regulation are necessary for motor vehicles to be ethically used even though they are ethical in the supposed context. Let us see the case in Japan.

There are at least two streams of acts and regulations. One stream includes the Act on Vehicles for Road Transportation, and it intends to permit only standard and safe motor vehicles on roads. The other stream includes the Road Traffic Act, and it intends to prevent road hazards and otherwise ensure the safety and fluidity of traffic, as well as to contribute to preventing blockages arising from road traffic. This stream enforces ethical uses of motor vehicles in practice.

From studying in detail the second stream, it becomes clear that the following four factors must be specified concerning the situation on which some regulations are planned to impose: the type of artifacts to be regulated, the concerned parties, the function of concerned operations of the artifact and the purpose of use.

Moreover, it is very important that technology necessary to enforce the acts and regulations, and also such devices as the speed meter and the automatic speeding camera, have been developed.

A set of acts and regulations exist in the RW to promote protecting and fostering sane minors in Japan. Looking into those acts and regulations, three methods are found: To ban businesses from using the concerned artifact to minors, to ban minors form using the concerned artifact and to obligate possession registration and/or license for use. The first method above seems to be applicable to protecting and fostering sane minors in the MW.

There are many evidences that suggest negative effects of mobile devices and their applications (e.g., Carr, 2017; Cooper, 2017). Table 1 exhibits some ubiquitous AI systems used in such applications.

Table 1 Example of AI Systems Capitalizing on User Vulnerability

Function	Data for learning&etc	Targeted to	Values of	Example
Use incentive	Use record, individual response record	Individual user (User group)	Service provider	Like!, Share, Retweet
Item sale	Use record, Personal data	Individual user (User group)	Service provider	On-line game, Social game
Ads targeting	Use record, Personal data, Ads specs	Individual user	Service provider /Advertiser	SNS, EC
Recommendation	Use record, Personal data, Goods data	Individual user	Service provider /Seller	EC

Source: self-elaboration based on literature survey

Minors constitute one of representative vulnerable groups. This is why we propose the following three plans of legislation and regulation:

1. To ban AI artifacts from using personal profile data for use incentive, items sale, ads targeting, and recommendation.
2. To ban applications such as SNSs opened to minors from implementing use incentive such as validation feedback.
3. To obligate service providers to display the lowest allowable age on each application and content, and to implement a mechanism to prevent minors below the age from using them.

This study proposes the following: a view that our living space is changing from the RW to the MW, using a functional definition of AI, referencing RW cases to plan legislation and regulation in the MW, and focusing functions, data for learning&etc algorithms, targets and values of AI systems. As an example of effectiveness of our proposals, we show what to regulate for protecting and fostering sane minors in the MW. New technology might be necessary in order to enforce these regulations. However, developing new technology for regulation was needed in the RW, too.

KEYWORDS: use-ethics, mixed world, legislation, AI artefact.

REFERENCES

AI HLEG (2019, April 8). Ethics Guidelines for Trustworthy AI. Retrieved from https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60419

Carr, N. (2017, Oct. 6) How Smartphones Hijack Our Minds. Wall Street Journal.

Floridi, L. & Cowls, J. (2019). A Unified Framework of Five Principles for AI in Society. Harvard Data Science Review Issue 1. Retrieved from https://hdsr.mitpress.mit.edu/pub/l0jsh9d1

Cooper, A. (2017, April 9) What is 'Brain Hacking'? Tech insiders on why you should care. Retrieved from https://www. cbsnews.com/news/brain-hacking-tech-insiders-60-minutes/

The Conference toward AI Network Society (2017). Report 2017-To Activate International Debate on AI Networking (in Japanese). Retrieved from http://www.soumu.go.jp/main_content/000499624.pdf

The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems (2019). Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems, First Edition. IEEE. Retrieved from https://standards.ieee.org/content/ieee-standards/en/ industry-connections/ec/autonomous-systems.html

ARTIFICIAL INTELLIGENCE AND MASS INCARCERATION

Leah Rosenbloom

The Workshop School (USA)

leah.rosenbloom@gmail.com

EXTENDED ABSTRACT

Artificial intelligence (AI) is now common throughout the criminal justice system. Police use predictive algorithms to target locations and individuals for surveillance. Judges use risk assessment algorithms to determine whether defendants should be granted bail or parole. Prosecutors use the results of forensic analysis algorithms to accuse and convict defendants of crimes, including those punishable by death. These algorithms are considered intellectual property and are closed off from public scrutiny.

In this paper, we explore the impact of AI on mass incarceration. A comprehensive survey of existing practice reveals the ways in which algorithms perpetuate systemic injustice and violate defendants' legal rights. We argue the need for solutions that integrate technical and legal perspectives, including novel ways to shift the focus of the algorithms from punitive to restorative practices. With due oversight, transparency, and collaboration between experts in technology, law, and government, we can leverage existing algorithms to combat systemic injustice.

1. INTRODUCTION

While existing literature concerning algorithms in criminal justice applications is extensive, research is scattered between the scientific and legal communities. In order to form a complete picture of the impact of AI on mass incarceration, which touches problems in machine learning, data science, law, and governance, it is necessary to integrate these perspectives. To that end, this paper explores sources, issues, and proposed solutions in each area of research.

Several ethical concerns emerge from the survey of existing literature. First is the issue of data that reflect existing racial and socio-economic bias in the criminal justice system. Data science experts have proposed statistical models that remove racial bias from the data, which leads us to consider the implications of "objective" black-box algorithms operating within contexts of deeply entrenched bias. We argue that the use of black-box solutions to racial discrimination encourages law enforcement and judiciaries to defer their responsibilities, preferring automatic arrests and convictions over critical consideration.

Without transparency and oversight, it is impossible to examine the underlying mechanisms that process and objectify the data. Furthermore, even if the underlying data is scrubbed clean, the "correctness" metrics and implementation of the algorithms may still reflect systemic bias. Comprehensive solutions to problems with algorithms in criminal proceedings must include technical, practical, and legal components. We introduce novel, integrated analyses for each application of artificial intelligence in the criminal justice system: predictive policing, risk assessment, and machine testimony.

2. MACHINE LEARNING

Artificial intelligence can be reduced to a machine's ability to learn (Russell and Norvig, 1995). Machine learning is defined as the ability to process input data such that the categorization of new data is correct to some degree of approximation. While a specific analysis of closed-source algorithms is regrettably impossible, all machine learning algorithms must necessarily "learn" from pre-existing data. Therefore, we can use our understanding of the data to draw conclusions about the effectiveness of these algorithms in practice.

Any errors, inconsistencies, and biases in the underlying training data will carry over into the algorithms. We know from existing problems that existing training data is far from accurate and unbiased. Law enforcement has a serious and long-standing problem with racist policing (Langan, 1995; Lum & Isaac, 2016). Judges are known to make racially biased decisions about bail, sentencing, and parole (Goel et al., 2018). Forensic evidence is often contaminated, and analysts are known to make mistakes and collude with prosecutors to guarantee convictions (DiFonzo, 2005). Rather than acknowledging and examining these issues, law enforcement and legal systems continue to plow forward with the integration of machine learning into criminal justice proceedings (Danner et al., 2016; Saunders et al, 2016). The result, as we will discuss in the rest of our paper, is the blind perpetuation of injustice.

3. PREDICTIVE POLICING

The U.S. National Institute of Justice (NIJ) describes predictive policing as a law enforcement approach that "leverages computer models…for law enforcement purposes, namely anticipating likely crime events and informing actions to prevent crime" (2014). These computer models are trained on existing reports of criminal and police activity. One of the most widely-used algorithms, PredPol, uses only the "three most objective data points" of time, location, and type of previously-reported crime in each precinct's regional area (PredPol, 2020). Others, like Chicago's "Strategic Subjects List", focus on identifying groups and individuals (Saunders et al., 2016). The NIJ confirms that predictive algorithms can focus on "places, people, groups, or incidents" (NIJ, 2014). Each of these models has been shown to perpetuate existing racial and socio-economic bias (Saunders et al., 2016; Lum & Isaac, 2016).

While predictive policing algorithms are currently employed to inform policing, it is possible to use the same algorithms for community healing and restorative justice (Marshall, 1999). These algorithms reveal bias: we can use them to identify communities that are likely to have broken relationships with law enforcement. Until we see movement towards restorative justice, predictive policing, and policing in general, will continue to plague communities in need.

4. RISK ASSESSMENT

After someone is arrested, a judge determines the conditions of that person's release. Typically, the judge makes some kind of "risk assessment" to determine how likely the defendant is to commit more crimes. These assessments can influence bail, sentencing, and parole. While a judge gets the final say, risk estimates have been increasingly performed by machine learning algorithms.

Similarly to how predictive policing algorithms run on biased arrest data, risk assessment algorithms run on biased arrest data *and* biased judicial data. The data used in risk assessment, however, is uniquely biased by selective outcome representation; if the defendant in the input data set was not released on bail, there is no way to determine whether or not they would have committed an offense if they had been released. This would suggest that if a particular group was disproportionality arrested

and detained, or detained for inconsistent reasons, the algorithm would be more unpredictable and less accurate for that group.

5. MACHINE TESTIMONY

Prosecutors have become increasingly reliant on algorithms that classify forensic evidence, especially DNA, to secure convictions. Unlike more traditional methods, where an analyst might compare two forensic samples in a lab, machine learning algorithms allow analysts to compare samples against millions of other samples in a DNA database, and obtain the probabilistic estimates of various matches. Problems with traditional forensics carry over into the millions of samples in DNA databases. Rogue actors in crime labs could further compromise results if they were to exploit flexibility or vulnerability in the algorithm's input parameters.

There is one problem unique to forensic algorithms that poses a grave threat to defendants' right to a fair trial. Traditionally, analysts and experts would be able to testify to each step of forensic analysis in detail. Forensic analysts that handle biological and chemical samples are understandably educated in biology and chemistry; they are not educated in machine learning, and cannot attest to the reliability of machine learning tests. Moreover, even machine learning experts cannot attest to the reliability of these tests, because the details of the algorithm are obscured behind copyrights and corporate policy. Without the source code, it is impossible for defendants to hear, understand, and question the evidence against them.

KEYWORDS: artificial intelligence, predictive policing, risk assessment, machine testimony, restorative justice, technology and the law.

REFERENCES

Danner, M., VanNostrand, M. & Spruance, L. (2016). Race and gender neutral pretrial risk assessment, release recommendations, and supervision: VPRAI and PRAXIS revised. *Luminosity, Inc.* Retrieved from https://www.dcjs.virginia.gov/sites/dcjs.virginia.gov/files/publications/corrections/race-and-gender-neutral-pretrial-risk-assessment-release-recommendations-and-supervision.pdf

DiFonzo, J. (2005). The Crimes of Crime Labs. *Hofstra Law Review*, 34(1) 1-12. Retrieved from https://scholarlycommons.law.hofstra.edu/cgi/viewcontent.cgi?article=2372&context=hlr

Goel, S., Shroff, R., Skeem, J, & Slobogin, C. (2018). The Accuracy, Equity, and Jurisprudence of Criminal Risk Assessment. *Social Science Research Network* [SSRN]. Retrieved from https://ssrn.com/abstract=3306723

Langan, P. (1995). *The Racial Disparity in U.S. Drug Arrests.* Washington, DC: Bureau of Justice Statistics, U.S. Department of Justice. Retrieved from https://bjs.gov/content/pub/pdf/rdusda.pdf

Makarios, M., Steiner, B., & Travis III, L. (2010). Examining the Predictors of Recidivism Among Men and Women Released from Prison in Ohio. *Criminal Justice and Behavior*, 37(12), 1377-1391. Retrieved from https://journals.sagepub.com/doi/abs/10.1177/0093854810382876

Marshall, T. (1999). *Restorative Justice: An Overview.* London: Home Office. Retrieved from http://www.antoniocasella.eu/restorative/Marshall_1999-b.pdf

National Institute of Justice [NIJ] (2014). *Predictive Policing*. Washington, DC: National Institute of Justice. Retrieved from https://www.nij.gov/topics/law-enforcement/strategies/predictive-policing/Pages/welcome.aspx

PredPol (2020). Overview. Retrieved from https://www.predpol.com/about/

PredPol (2020). Proven Crime Reduction Results. Retrieved from https://www.predpol.com/results/

Russell, S. & Norvig, P. (1995) *Artificial Intelligence: A Modern Approach*. Prentice-Hall. Retrieved from https://pdfs.semanticscholar.org/bef0/731f247a1d01c9e0ff52f2412007c143899d.pdf

Saunders, J., Hunt, P., & Hollywood, J. (2016). Predictions put into practice: a quasi-experimental evaluation of Chicago's predictive policing pilot. *Journal of Experimental Criminology*, 12(3), 347-371. Retrieved from https://link.springer.com/article/10.1007/s11292-016-9272-0

Sewell, A. & Jefferson, K. (2016). Collateral Damage: The Health Effects of Invasive Police Encounters in New York City. *Journal of Urban Health*, 93(1), 42-67. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4824697/

CAPTURING THE TRAP IN THE SEEMINGLY FREE: CINEMA AND THE DECEPTIVE MACHINATIONS OF SURVEILLANCE CAPITALISM

Fareed Ben-Youssef, Kiyoshi Murata, Andrew Adams

Texas Tech University (USA), Meiji University, Centre for Business Information Ethics (Japan),

Meiji University, Centre for Business Information Ethics (Japan)

fbenyous@ttu.edu; kmurata@meiji.ac.jp; aaa@meiji.ac.jp

EXTENDED ABSTRACT

Shoshana Zuboff's Big Other concept offers a means to understand the paradigm shifts provoked by surveillance capitalism—by social networking systems collecting user data with little government oversight or end user understanding. The Big Other represents "an intelligent world-spanning organism" which brings with it "new possibilities of subjugation... as this innovative institutional logic thrives on unexpected and illegible mechanisms of extraction and control that exile persons from their own behavior" (Zuboff 85). The Big Other's inescapable annihilating power comes in part for how it evades legibility. In her analysis, Zuboff does not fully historicize the Big Other's rise, only broadly comparing its logic of total conquest to that of former imperial powers. Our paper disrupts the pervasive illegibility of the Big Other using three key examples in global cinema. In the process, we productively fill in historical blind spots in Zuboff's framework.

To underline the new subjugations of the Big Other, our interdisciplinary paper traces the line between what constitutes just and the unjust surveillance within business. Our examples feature enthused surveillance capitalists as well as confused, even terrified end users. We end by framing the historical roots of such an unquestioned form of mass surveillance by situating studies about the role of bureaucracies and Big Data in the Holocaust against Quentin Tarantino's WWII film *Inglourious Basterds* (2009). Our comparison illustrates the troubling consequences of the Big Other's emergence: to be reduced to data is to accept the possibility of being deleted. Cinema, we will ultimately show, is especially well-primed to visualize the trap in such seemingly free services, to make visible the often-invisible machinations of surveillance capitalism.

Our study employs a methodology which combines theories from social science with humanities-style close reading. Our framework offers us the opportunity to engage with the formal construction of these media texts. We tease out the tensions in these cinematic examples. In so doing, we show how these films are not simple entertainment; rather, they frame a contradiction—the allures of unregulated surveillant power as well as the root horror of its dehumanizing potential.

The Circle – The Attraction of Surveillance Capitalism

Our analysis first explores the attraction of surveillance capitalism for the business world as expressed in the film adaptation of *The Circle* (2017) directed by James Ponsoldt and written by the novel's author Dave Eggers. A scene shows a Steve Jobs-like executive introducing a line of hidden cameras with the slogan: "Knowing is good, knowing everything is better!" Even as the film captures the laudatory Silicon Valley rhetoric around such practices, it also winks at data mining's costs. The executive's admission that he placed cameras in public sites without any permits evokes the unregulated reality in which many businesses operate. While the executive heaps praise upon his company's total technological

vision, the screen behind him shows images of a city on fire. With such a tension between the executive's valorizing words and the stark imagery of destruction, the satirical film gestures to the unseen devastation of such unquestioned surveillance.

Pulse – The Terrifying Deceptions of the End User

Our second reading shows how cinema has represented how end users are deceived by technology corporations. Like the Spanish conquistadors before them, Zuboff argues that early surveillance capitalists, "relied on misdirection and rhetorical camouflage, with secret declarations that we could neither understand nor contest" (qtd. in Naughton). Kiyoshi Kurosawa's horror film *Pulse* (2001) features a haunted internet browser which serves as a metaphor for unscrupulous corporations. We analyze a scene wherein the user must agree to the browser's Terms of Use Agreement. Before they begin their terrorizing, the ghosts behind the browser force the unsuspecting user to agree to a contract that he can neither understand nor contest. Indeed, he blindly clicks through the agreement. A popup message appears that declares "Have Fun!" further suggesting how corporations begin to haunt the consumer. They trick the consumer with the prospect of fun and overwhelm him with impenetrable legal language.

Inglourious Basterds – Uncovering the Big Other in the Holocaust

Our last reading focuses on Tarantino's *Inglourious Basterds* to underscore the insidious destructive potential of the Big Other. At first glance, the film seems very distant from AI or information ethics concerns. However, we show how the film permits us to historicize the development of the Big Other within Big Data's imbrication in the Holocaust. While IBM's complicity in the genocide has been well-documented, Zuboff only gestures to the destructive potential of the Big Other. Zuboff employs Karl Polanyi's idea of 'commodity fiction' where people are subordinated to the market. Polanyi noted that such fictions "disregarded the fact that leaving the fate of soil and people to the market would be tantamount to annihilating them" (qtd. in Zuboff 83). Zuboff continues, "in the logic of surveillance capitalism there are no individuals, only the world spanning organism and all the tiniest elements within it" (Zuboff 83). Here, while ignoring any specific historical examples, Zuboff points to an over-riding logic of annihilation where users are reduced into the Big Other's tiniest elements. Her language acts as a starting point for our analysis of how Tarantino allegorically depicts historian Raul Hilberg's Bureaucratic Process of Destruction. By culminating upon a film about the Holocaust, we show how being a good businessman under surveillance capitalism may have the same ethical ramifications of being a good bureaucrat in a destructive surveillance state.

Raul Hilberg has argued about the importance of bureaucracy within Nazi genocide noting, "at first sight the destruction of the Jews may have the appearance of an… impenetrable event. Upon closer observation it is revealed to be a process of sequential steps that were taken at the initiative of countless decision makers in a far-flung bureaucratic machine" (Hilberg 53). We analyze a scene in *Inglourious Basterds* where a Nazi colonel named Hans Landa transforms a farmhouse into an office when searching for hiding Jews. The sequence enacts the three steps of Hilberg's Bureaucratic process. First comes identification where the targeted people are bureaucratically identified on paper as Jewish. Second comes concentration, where the targeted group is trapped in a ghetto and is controlled "through the watchful eyes of the entire German population" (qtd. in Hilberg 50). Landa reveals his all-seeing eye when noting that he knows exactly where the Jews are hidden. Finally, the third step is annihilation. In the film, the Jewish populace is executed by the soldiers. To be visible is to be capable of being part of the destructive process. The film thus illustrates the stakes when surveillance, be it on

the state level or that of private enterprise, goes unchecked and unmanaged. It pushes us to reflect upon what banality of evil both the distant bureaucrat and the algorithm might share.

We lastly explore the scene's vital play with language which recalls the ignorance of the consumer in the film, *Pulse*. Landa sets up the execution of the hiding Jewish family by speaking in English so that his French victims cannot comprehend. The film metaphorically indicates that the bureaucratic system driving the murder remains similarly incomprehensible. After hearing a confession of the Jewish family's whereabouts, the colonel says, “I am going to switch back to French, and I want you to follow my masquerade.” He says “Adieu” while directing his soldiers to fire their guns. Those affected by the system’s violence cannot understand the true meaning behind the language of bureaucracy. They cannot make sense of the possibilities of subjugation that define mass surveillance by the state or in more unfettered forms of surveillance capitalism.

Conclusion

Our analysis renders newly legible the Big Other's illegible processes, highlighting how cinema can frame the allure and costs of such control. In so doing, these key films show how media drives home the paradigm shift of surveillance capitalism and unveil its under-explored history. We have moved from a substrate of mediated relations, a village society wherein the state had limited and exceptional access to encoded information, to a new stratum of communication with the emergence of social networks. Now we have platforms that can see everything, an unregulated entity that can access all. Cinema tracks these shifts and the ensuing danger when businesses follow mantras like: “Knowing is good, knowing everything is better!” In so doing, these films demand scholarly attention for how they offer the public a viscerally affecting and disruptive critical understanding: they permit viewers to see how our rights of privacy come to burn up in the light of seemingly free social networks. These films ultimately give scholars of surveillance, AI, and information ethics a new language to map out surveillance capitalism’s trap.

KEYWORDS: Surveillance Capitalism, Cinema, Big Other, Control, The Holocaust, Banality of Evil.

REFERENCES

Hilberg, Raul. (1985) *The Destruction of the European Jews: Student Edition*.

Naughton, John. “'The goal is to automate us': welcome to the age of surveillance capitalism.” *The Guardian*. 20 Jan. 2019. Web. Accessed 10 July 2019.

Zuboff, Shoshanna. (2015) “Big other: surveillance capitalism and the prospects of an information civilization.” *Journal of Information Technology* 30, pp. 75-89.

DIFFERENCES IN HUMAN AND AI MEMORY FOR MEMORIZATION, RECALL, AND SELECTIVE FORGETTING

Sachiko Yanagihara, Hiroshi Koga

University of Toyama (Japan), Kansai University (Japan)

sachiko@eco.u-toyama.ac.jp; koga@res.kutc.kansai-u.ac.jp

EXTENDED ABSTRACT

Artificial intelligence (AI) mimics the learning behavior of humans, but there are significant differences in the capabilities of AI and humans. Tasks exemplifying these differences are more appropriate for humans to perform. One of these significant differences is the capability of AI to store and use all information, in contrast to the imperfect information management of humans. In this paper, we examine the mechanisms of repeated memorization and selective forgetting of memory through the game of competitive karuta, which is a traditional Japanese card game. First, we outline the general differences between human memory and that of IT artifacts, including AI, confirming previous research of memorization and forgetting. Next, we outline competitive karuta and examine human memory in that context. Finally, we consider mechanical differences of memory between AI, which can remember everything, and human beings, who forget, in reference to the “intentional forgetting” of Gloding & Maecleod (1998) .

Tulving (1974) defines human "forgetting" as "the inability to recall something now that could be recalled on an earlier occasion”. One theory of forgetting is co-dependent forgetting, known in psychology as context-dependent forgetting, and defined by Tulving as “reflecting the failure of retrieval of perfectly intact trace information”. This theory is famous for describing episodic memory. Episodic memory is subjective and autobiographical, and it is characteristic of the mechanisms of memory for sentient beings. Excepting AI, IT artifacts cannot memorize episodically, and instead memorize all information through concrete commands given by a human. Therefore, they forget memorized information in only particular situations.

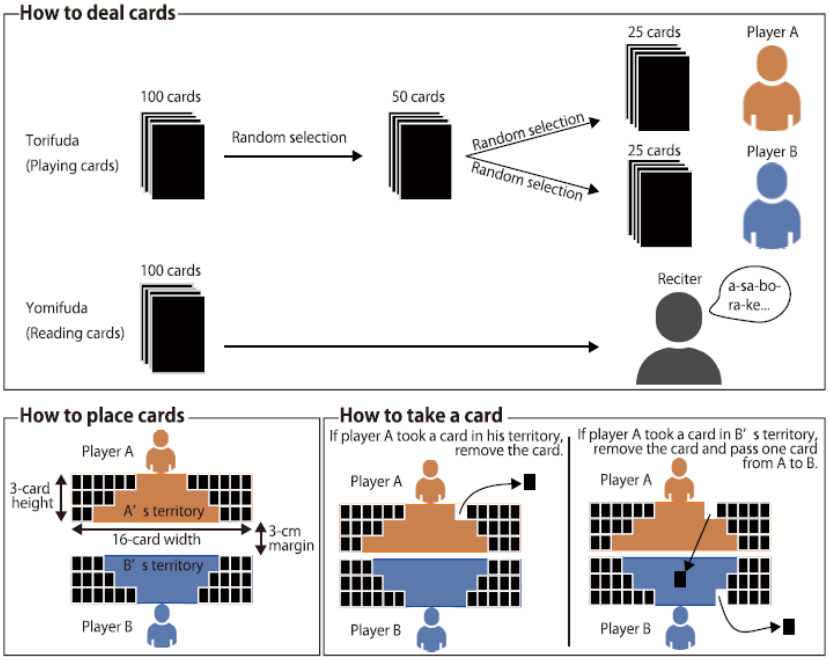
Hardware defines the only storage limitation of IT artifacts. Though forgetting is an everyday occurrence for humans, it is special situation in IT because IT artifacts cannot intentionally forget. Previous studies have shown that human memory requires "intentional” forgetting (Macleod & Golding, 1998). In addition, Ricoeur (2000) says “Forgetting is bound up with memory”, and “Forgetting can be considered one of the conditions for it”. Though IT artifacts do not intentionally forget, they share this relationship between memory and forgetting. Kluge and Gronau (2018) have defined “intentional forgetting” as “the motivated attempt to limit the future recall of a defined memory element”. Timm et al. (2018) explained further that “intentional forgetting” is a significant mechanism in an AI system. For example, Nuxoll et al. carried out a study for an algorithm of forgetting in order to artificially perform episodic memory (2010). The importance of this is demonstrated by technical methods for forgetting being studied considering the "Right to Be Forgotten" (Villaronga, Kieseberg, Li, 2018). To develop this idea further, this paper considers the processes of memory and forgetting used in competitive karuta as a way to observe the “intentional forgetting” that AI cannot do, but which humans can, through the repetition of memorization and forgetting.

Traditional games in Japan often challenge players to memorize and recall the state of a game, and this is especially true of competitive karuta. Competitive karuta was established based on a thirteenth

century literary work called “Ogura Hyakunin Isshu”, which translates to “One Hundred Poets, One Poem Each” in English. In competitive karuta, poems from Ogura Hyakunin Isshu, each comprised of 31 syllables, are written on cards called “karuta”. Of these, "yomifuda" cards contain an entire poem, while "torifuda" cards contain only the second half of the poem. The basic rules of the game are simple, and the playing area is as shown in Figure 1. After a yomifuda card is selected, a reciter reads the first half of the poem given on the card, and the players must select the corresponding torifuda card that contains the second half of the poem. The objective of the game is to reduce the number of cards in your territory to 25, and this is done by correctly remembering and identifying torifuda which match the recited yomifuda. When the correct torifuda is selected, the player who identified the correct torifuda either removes a card from their territory or adds a card their opponent’s territory. If a player selects a torifuda incorrectly, this is known as an “otetsuki”, or “foul” in English, and results in a card being taken from the opponent’s territory, increasing the mistaken player’s number of owned cards. Before the start of the game, the positions of the cards in the field are memorized in advance, and players attempt to memorize the cards by “kimariji”, which is the first syllable of a card by which a correct torifuda can be identified. However, because the placement and kimariji of the cards change as the competition progresses, it is necessary to re-memorize the placement and kimariji of the cards quickly. Because this process is easy for IT artifacts, game apps for competitive karuta have already been launched.

On the other hand, players must make decisions and move quickly to identify and select cards according to the progress of the game. This decision-making and action based on the game information has a direct impact on the outcome of the game. Though the physical aspect of the game can be performed reliably by a robot arm following the derivation of an optimal solution using an AI, the most important aspects of the game are memorizing the initial position and kimariji of the cards and then re-memorizing them as the game state changes over time. For this task, past memories are forgotten, and new information is repeatedly stored. No matter how fast the robot moves, it cannot win if its memory is weak. Even if an application is developed with total recall and accurate operation, a first-class player is difficult to surpass unless the application has the ability to hear slight differences in the voice of the reciter.

Figure 1. Rules of competitive karuta



Source: Yamada, Murao, Terada, and Tsukamoto (2018)

Players use only the essential information of past memories. However, previously memorized information is not used directly because using it becomes an obstacle to accessing current information and ensuring the accuracy of that information. This is not the case for IT artifacts, which have the capability of total memorization and recall. To accurately model human memory, though it is important to store information in an extended area that can be separated and forgotten, searched when necessary, and accessed, it is also important to block the access of some information to improve accessibility to prioritized information. This mechanism of intentional forgetting through omission can be observed in humans, and it improves memory reliability. We know that repeated learning prevents forgetting and improves long-term memory. However, it is important for humans to intentionally forget while repeatedly memorizing information, which differs from the behavior of AI. Modern IT artifacts are capable of total recall, and we recognize this capability as useful. Though humans would like the ability to memorize and recall at will, memory may be strengthened through intentionally forgetting.

When considering people working in organizations, humans are superior in activities that require collaboration to perform physical activities. This is a necessary consideration in the relationship between AI and humans. Though IT artifacts have a capacity for memory that surpasses that of humans, this difference between AI and humans is significant in modelling human intelligence accurately. To accurately model human intelligence, AI must make human-like decisions about whether to forget a given piece of learned information. Until AI has the ability to intentionally forget, AI intelligence cannot be considered analogous to that of human beings.

KEYWORDS: intentional forgetting, memory, competitive karuta, artificial intelligence.

REFERENCES

Golding, J. M. and Macleod, C. M. (1998). Intentional forgetting: interdisciplinary approaches. L. Erlbaum Associates, Mahwah, N.J.

Kluge, A. and Gronau, N. (2018). Intentional Forgetting in Organizations: The Importance of Eliminating Retrieval Cues for Implementing New Routines. *Front. Psychol.* 9:51. doi: 10.3389/fpsyg.2018.00051

Nuxoll, A., Tecuci, D., Ho, W. C., & Wang, N. (2010). Comparing Forgetting Algorithms for Artificial Episodic Memory Systems, Proceedings of the Remembering Who We Are- Human Memory for Artificial Agents Symposium, at the AISB 2010 convention, De Montfort University.

Ricoeur, P. (2000). LA MEMOIRE, L'HISTOIRE, L'OUBLI, Editions du Seuil, (Memory, History, Forgetting, Translated by Blamey,K,and Pellauer,D. The University of Chicago Press, 2004)

Timm, I. J., et al. (2018). Intentional Forgetting in Artificial Intelligence Systems: Perspectives and Challenges. In: Trollmann F., Turhan AY. (eds) KI 2018: *Advances in Artificial Intelligence*. KI 2018. Lecture Notes in Computer Science, vol 11117. Springer, Cham

Tulving, E. (1974). Cue-Dependent Forgetting: When we forget something we once knew, it does not necessarily mean that the memory trace has been lost; it may only be inaccessible. *American Scientist,* 62(1), 74-82. Retrieved from http://www.jstor.org/stable/27844717

Villaronga, E. F., Kieseberg, P., & Li, T. (2018). Humans forget, machines remember: Artificial intelligence and the Right to Be Forgotten. *Computer Law & Security Review*, 34-2, 304-313

Yamada, H., Murao, K., Terada, T., & Tsukamoto, M. (2018). A Method for Determining the Moment of Touching a Card Using Wrist-worn Sensor in Competitive Karuta. *Journal of Information Processing*, 26, 38-47

MONITORING AND CONTOL OF AI ARTIFACTS: A RESEARCH AGENDA

Hiroshi Koga, Sachiko Yanagihara

Kansai University (Japan), University of Toyama (Japan)

koga@res.kutc.kansai-u.ac.jp; sachiko@eco.u-toyama.ac.jp

EXTENDED ABSTRACT

The purpose of this paper is to find a future research agenda through examination of the concept of AI artifacts.

To that end, this paper is organized as follows: First, First, what AI artifacts are discussed. Next, the characteristics of AI artifacts are clarified, that is, the following two points. (1) AI artifacts contain organizational context and human agency, (2) AI artifacts fuse boundaries with natural objects. Finally, the impact is examined and future research agendas are proposed.

In recent information systems research, the search for the significance of IT artifacts is recognized as an important issue. IT artifacts are perceived as "those bundles of material and cultural properties packaged in some socially recognizable form such as hardware and /or software" (Orlikowski & Iacono, 2001, p.121). Furthermore, Orlikowski and Iacono (2000) offer the following five premises of IT artifacts; That is, (1) IT artifacts, by definition, are not natural, neutral, universal, or given. (2) IT artifacts are always embedded in some time, place, discourse, and community. (3) IT artifacts are usually made up of a multiplicity of often fragile and fragmentary components, whose interconnections are often partial and provisional and which require bridging, integration, and articulation in order for them to work together. (4) IT artifacts are neither fixed nor independent, but they emerge from ongoing social and economic practices. (5) IT artifacts are not static or unchanging, but dynamic.

Thus, IT artifacts tend to emphasize organizational aspects rather than technical characteristics. Therefore, it should be called "social artifact" or "socio-materiality". From such a perspective, Lee et al. (2015) referred to the subject in "Information System Research" as "IS artifacts", which are subclasses: (1) information artifacts, (2) technology artifacts, and (3) social artifacts. It points out the need to focus on the interaction between them (Lee, Thomas and Baskerville, 2015).

Furthermore, IT artifacts have been considered as a component of organizational practice with human agents. Leonardi (2012) named "imbrication" the structure in which human agencies and IT artifacts (material or technical agencies) are intertwined. Similarly, Orlikowski (2008) called such a structure "entanglement".

On the other hand, it is for AI artifacts that technical aspects are often emphasized. For example, "such as artificial neural networks, specifically focusing on deep neural networks" by Tuncali et.al. (2018. P.1) or "data and AI models being used in the process of AI system development" by Maksimov et.al. (2018, p.2).

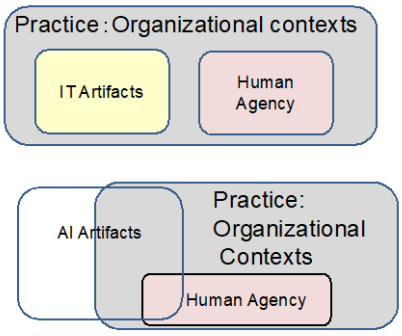
Behind such a strong technical orientation, it is thought that the organizational context is embedded in machine learning. In other words, embedding organizational context means that the intelligence activities that have been entrusted to human beings have been entrusted to artifacts.

Figure 1 is a simplified diagram of these differences. The top diagram in Figure 1 schematically illustrates the relationship between IT artifacts (material or technical agencies) and human agencies in organizational practice. The diagram below in Figure 1 shows that AI artifacts not only capture part of

human agency intelligence /decision-making activities, but also contain organizational context. The part that protrudes from the organizational context suggests an AI artifact “runaway (be out of control)”. AI artifacts need to be "monitored and controlled" so as not to deviate from the organizational context.

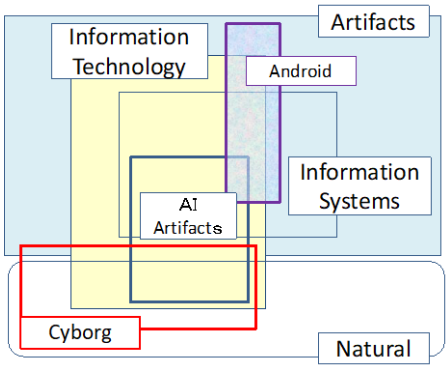
Another important difference between AI artifacts and IT artifacts is that AI artifacts can become hybrids with natural objects, as shown in Figure 2.

Figure 1. Differences between AI artifacts and IT artifacts



Source: Drawing by authors

Figure 2. Differences between AI artifacts and IT artifacts



Source: Drawing by authors

Delegation of intelligence / judgment functions creates the following problems. For example, where is the responsibility for accidents in autonomous driving? In the academic field of business administration, “responsibility” has been considered a key factor for the simultaneous development of individuals and organizations (Barnard, 1938). Therefore, if the responsibility is ambiguous, the organization may collapse.

Alternatively, there is a risk that data analysis generates “sensitive information”. For example, the invasion of privacy will be seen as a problem, such as an US company (e.g. Target Corporation) predicting the number of gestation weeks of customers. Needless to say, in the field of business administration, customers are also considered to be organizational members (contributors) (Barnard, 1938). This is because it is difficult to continue organizational activities if customer contributions are supported.

The hybrid of natural and artificial objects can be rephrased as a hybrid of real and virtual. Incidentally, as with cricket, baseball and ya-kyu (Japanese), there are regional differences in attitudes towards hybrids. In baseball, the United States, it is no exaggeration to think that privacy can be infringed if it can provide an excellent customer experience. In cricket, that is, in Europe, it is important not to infringe on individual rights such as the right to be forgotten. In any case, in these areas, it seems to understand that AI artifacts should be under human control. On the other hand, in ya-kyu, that is, in Japan, the attitude toward hybrids is affinity. For example, in Japan, industrial robots are given names (for example, names of female idols such as Momoe, Junko and so on). This is because, like humans, robots are considered "comrades". In Buddhism, it is considered "all things have the Buddha nature". Therefore, AI artifacts are also considered to be equal to humans and have little resistance to accepting AI artifacts as friends.

AI artifacts differ in nature from traditional IT artifacts. AI artifacts (1) merge with natural objects and the real world, and (2) come to include organizational context. Therefore, the danger of producing unintended results cannot be denied. This is a reason why AI artifacts need to be monitored and controlled. The following agendas can be pointed out as specific monitoring and control issues.

(1) Elucidation of the relationship between AI artifacts and responsibility, which is an element of organizational development

(2) Elucidation of problems in utilizing action history information of organization contributors

(3) International comparison of attitudes toward the integration of humans and AI artifacts

As it has been described above, we have definitely confirmed that more research is urgently needed to explore a variety of new phenomena of AI artifact monitoring and control.

KEYWORDS: IT artifacts, AI artifacts, responsibility, privacy.

ACKNOWLEDGMENT: This work was supported by JSPS KAKENHI Grant Number JP12345678 and by the Kansai University Fund for Domestic and Overseas Research Support Fund, 2019.

REFERENCES

Barnard, C.I. (1968). *The Functions of the Executive*. Harvard university press.

Leonardi, P.M. (2012). *Car crashes without cars: Lessons about simulation technology and organizational change from automotive design*. MIT Press.

Maksimov, Y.V., Fricker, S.A. and Tutschku, K. (2018). Artifact Compatibility for Enabling Collaboration in the Artificial Intelligence Ecosystem. *International Conference of Software Business*, pp.56-71.

Orlikowski, W.J. and Iacono, C.S. (2000). The Truth Is Not Out There: An enacted view of the digital economy. Brynjolfsson, E. and Kahin, B. eds. *Understanding the Digital Economy: Data, Tools, and Research*. MIT Press, 352–380.

Orlikowski, W.J. and Iacono, C.S. (2001). Research commentary: Desperately seeking the "IT" in IT research: A call to theorizing the IT artifact. *Information Systems Research*, 12(2), 121-134.

Rankin, T. (1987). The Turing paradigm: A critical assessment. Dialogue, 29(2-3), 50-55.

Tuncali, C.E., Ito, H., Kapinski, J. and Deshmukh, J.V. (2018). Reasoning about safety of learning-enabled components in autonomous cyber-physical systems. *2018 55th ACM/ESDA/IEEE Design Automation Conference* (DAC), pp. 1-6.

ON THE CHALLENGES OF MONITORING AND CONTROL OF AI ARTIFACTS IN THE ORGANIZATION FROM THE PERSPECTIVE OF CHESTER I. BERNARD'S ORGANIZATIONAL THEORY

Hiroshi Koga

Kansai University (Japan)

koga@res.kutc.kansai-u.ac.jp

EXTENDED ABSTRACT

The concept of “artificial intelligence:AI” that appeared about 60 years ago is now attracting attention again in the business world. On the other hand, currently, the application range of AI commercial applications is not so large. It is only used in areas such as image recognition, natural language processing and automatic translation. However, there is concern that the risk that the user will suffer increases as the AI artifact increases in the future. Of course, there is optimism about the risk of AI artifacts. The claim is the same as the risk to other technologies (Bryson and Kime, 1998). However, unlike conventional ICT, which supports humans, AI makes decisions with human support. In other words, the position of human agency is different. Therefore, the development of AI artifacts and their use in organizations are considered to involve risks different from those of conventional ICT. In this paper, we will consider such risks from organizational research, especially from the perspective of C.I. Barnard’s concept of organization.

Therefore, the purpose of this paper is to clarify why “the monitoring and control of AI artifacts” is necessary, and to discuss that the key to “monitoring and control” is “the moral leadership” claimed by Chester I. Barnard, the father of modern organization theory and management studies.

Furthermore, this paper is not a positivist study such as a statistical survey, but a theoretical study. In pursuit of rigor, a positivist research may be desirable. Nonetheless, we chose theoretical research because we wanted to show that the concepts proposed by Chester I. Barnard are valid as a reference framework for comprehensive discussion of AI artifact management issues. Then, the contribution of this paper is to consider the ethics of AI artifacts from the viewpoint of organizational theory and to discuss from the relevance to leadership.

First, an overview of C.I. Barnard’s argument are introduced. C.I. Barnard, president of an American telephone company, was famous as the father of modern organization theory. His landmark 1938 book, “The Functions of the Executive,” proposed a unique concept of organization. That is, the organization was defined as “coöperation as a functioning system of activities of two or more persons”. In his idea, organizations are generally not long-lived. In order to make the organization permanent, it is necessary to provide an incentive for contributors to continue to participate. Barnard called the balance between contribution and incentive “the organizational equilibrium”.

In addition, he introduced the concept of “zone of indifference.” He said “there exists a ‘zone of indifference’ in each individual within which orders are acceptable without conscious questioning of their authority” (Barnard, 1938, p.167).

Next, it is discussed why monitoring and control of AI artifacts is necessary. C.I. Barnard sees organizations as “systems of activity” rather than “structures”. The system of activities is based on contributions from people and material devices. Therefore, “practice” is the subject of research.

Furthermore, according to the sociomaterial perspective that Wanda J. Orikowski claims, human and material agencies are intricately intertwined and integrated in the course of activities and practices (Orklikowski, 2009). Such a perspective of organizational practice is referred to as the sociomateriality.

In traditional IT artifacts, material agencies have been used to support human agency activities. To that end, attention has been focused on the “unintended consequences” of material agencies (cf. Pickering, 2010).

On the other hand, in the case of AI artifacts, the configuration pattern of the practice in which it is incorporated (or embedded) is different. First, the AI agency is a “chimera” that is a half human agency and a half material agency because it incorporates some of the judgment activities of the human agency. Second, conventional IT artifacts are only “human assistance”, but AI artifacts are “autonomous” and do not require “entanglement” with human agencies. From the above points, the function of AI artifacts in organizational practice is unpredictable. Therefore (for the time being) monitoring and control of AI artifacts is essential.

Next, the author will discuss the challenges of AI artifact monitoring and control from two concepts presented by Barnard.

The first keyword is “organizational equilibrium”. Barnard considered the “contributor of the organization” to be a provider of activities in a broad sense, including shareholders and customers. Customers are allowed to maintain their organization as a contributor if they have enough rewards to provide activities and personal information. However, if the return is insufficient, stop contributing. If monitoring with AI artifacts weakens the motivation of contributors, the organization collapses. Therefore, it is necessary to consider the fairness and fairness of the return of contributors.

The second keyword is “the zone of indifference”. Whether the contributor accepts the judgment indicated by the AI artifact depends on the size of the indifference zone. The size of indifference is constantly changing. It is leadership that determines the size.

Thus, the third keyword will be introduced. It is “moral leadership”. Barnard points out that moral leadership is the key to injecting and sustaining value in an organization. Moral leadership greatly affects the fairness and fairness of contributors' rewards. In addition, it affects the size of the indifference zone. Therefore, the key factor for monitoring and controlling AI artifacts depends on the ability to demonstrate moral leadership. This is the conclusion of this paper

KEYWORDS: AI artifact, organizational equilibrium, zone of indifference, moral leadership, sociomateriality.

ACKNOWLEDGMENT: This work was supported by JSPS KAKENHI Grant Number JP12345678 and by the Kansai University Fund for Domestic and Overseas Research Support Fund, 2019.

REFERENCES

Barnard, C.I. (1938). *The Functions of the Executive*. Harvard University Press. OCLC 555075.

Bryson, J. J., & Kime, P. (1998). Just another artifact: Ethics and the empirical experience of AI, *Fifteenth International Congress on Cybernetics,* pp. 385-390.

Orlikowski, W.J. (2009). The sociomateriality of organisational life: considering technology in management research. Cambridge journal of economics, 34(1), pp.125-141.

Pickering, A. (2010). *The mangle of practice: Time, agency, and science*. University of Chicago Press.

POST-TRUTH SOCIETY: THE AI-DRIVEN SOCIETY WHERE NO ONE IS RESPONSIBLE

Kiyoshi Murata, Yohko Orito, Tatsuya Yamazaki, Kazuyuki Shimizu

Meiji University (Japan), Ehime University (Japan), Meiji University (Japan),Meiji University (Japan)

kmurata@meiji.ac.jp; orito.yohko.mm@ehime-u.ac.jp; peasesephiros@gmail.com; shimizuk@meiji.ac.jp

EXTENDED ABSTRACT

This study deals with the *post-truth society*, which would advent due to the widespread use of uncontrolled or uncontrollable artificial intelligence (AI) systems. In that society, people would be encased in filter bubbles in various aspects of their lives where what they know is unconsciously controlled by machine learning algorithms, and thus it would become very difficult for them to discover the real truth about the world. It's well known that personalised political advertisements delivered by Cambridge Analytica at the 2016 US presidential election and in the Brexit campaign brought about the political situation called the post-truth politics.

It would also be hard for them to successfully control their identities in that society because information on them including socially stigmatic one created by AI systems would remain accessible online for long periods of time, and many of those who access it could easily believe the contents of the information regardless of whether it is true or not. The truth about individuals, groups, organisations, nations and so on would become meaningless or worthless for the society, and people would be forced to live their post-truth lives in despair.

An actual example of an AI application which functions as a threat to personal identity is one to create a deepfake, a doctored video in which a person can be made to appear as if they are doing and saying anything (Cook, 2019a). Many people including politicians and famous figures have become victims of the AI applications to masterfully edit deepfakes, being distorted their digital identities. The utilisation of this sort of AI software which can be used to conceal the truth and replace it by fakes could threaten democracy and suppress individual freedom. When it comes to deepfake porn videos, the AI applications could lead to curtailing freedom of expression and violating human dignity – especially of women – although some takes a negative attitude towards regulating such contents ironically on the ground of the protection for freedom of expression. Eventually, deepfake AI applications have not been effectively regulated so far whereas technological efforts to fight against deepfake videos are continued (Kemeny, 2018). In addition, it is very difficult to find people responsible for the victims' damage created by deepfakes (Cook, 2019b).

The autonomous functioning of machine-learning-based AI systems, which leads to the unpredictability and uncontrollability of them, would provide parties relevant to the development and use of those systems, such as software engineers and system developers, with a good excuse to evade their responsibility for harm the systems cause. In fact, for example, it is not easy to decide who has to take a responsibility for a fatal traffic accident caused by an autonomous car. No one would be willing or able to be responsible for anything happen owing to the systems in the post-truth society. More than twenty years ago when a computerised society centred on the Internet was emerging, Nissenbaum (1996) pointed out the four factors which erode accountability in computing: many hands, bugs, computers-as-scapegoat and ownership without liability. Today, in the early days of an AI-driven

computerised society, the situation surrounding the four factors has become worse rather than better. Many people who are not necessarily personally identified have contributed to the development of AI systems, into which free/libre open source software (FLOSS) modules are often incorporated. It seems to be necessary to reconsider the meaning or definition of bug due to the unpredictability and uncontrollability of AI systems, and these characteristics may promote engineers' or developers' attitude of dodging responsibility by shifting the blame to AI systems. It has not been unusual that the source codes of AI modules developed by for-profit information and communication technology (ICT) companies are exposed as FLOSS. Consequently, it is now extremely difficult to fill the vacuum of accountability in AI computing.

The risk of the emergence of the post-truth society where no one is responsible demonstrates the social significance of the accountable management of AI artefacts through properly monitoring and controlling them, though this is really a tough challenge. However, if such management is failed, we would face the disruptions of social lives of individuals, the erosion in local communities, social fragmentation and the ruin of democracy.

It is not realistic that providing AI artefacts with legal personality and questioning their responsibility. Instead, of course, organisations which engage in the development and use of AI systems should take their responsibility and accountability for the technological and social quality of them. Nowadays, a large majority of ICT-based system developments and operations are conducted by business organisations. The speed of ICT developments is very fast often being referred to as dog year or mouse year, and cutting-edge ICT is rapidly deployed by business organisations without disclosing sufficient information about the deployment because it is conducted in a competitive environment. Therefore, unless business organisations develop and use ICT based on the idea of "ethics by design" taking their responsibility and accountability to the current and future generations, responses to the harm brought about by novel ICT can be made only afterwards. Those who are members of organisations which engage in the development and operation of ICT-based systems are required to recognise that even ethical – not to speak of legal and technological – responses to the harm are reactive, not proactive, if they are made by people outside the organisations.

However, major players in the ICT industry who lead the development and use of cutting-edge ICT including AI technologies seem not to willingly take their responsibility commensurate with the tremendous impact of their business activities on society. Actually, many ICT companies have maintained an attitude of "innovative first, consider consequences afterwards". But, if they fail to behave as professionals, their development and use of cutting-edge ICT may bring about serious social harm.

KEYWORDS: post-truth society, post-truth life, uncontrolled or uncontrollable AI system, responsibility, accountability.

REFERENCES

Cook, J. (2019a, June 12). Deepfake videos and the threat of not knowing what's real. Huffpost. Retrieved from https://www.huffpost.com/entry/deepfake-videos-and-the-threat-of-not-knowing-whats-real_n_5cf97068e4b0b08cf7eb2278.

Cook, J. (2019b, June 23). Here's what it's like to see yourself in a deepfake porn video: there's almost nothing you can do to get a fake sex tape of yourself taken offline. Huffpost. Retreived from

https://www.huffpost.com/entry/deepfake-porn-heres-what-its-like-to-see-yourself_n_5d0d0faee4b0a3941861fced.

Kemeny, R. (2018, July 10) AIs created our fake video dystopia but now they could help fix it: new software developed by artificial intelligence researchers could help in the fight against so-called deepfake videos. Wired. Retrieved from https://www.wired.co.uk/article/deepfake-fake-videos-artificial-intelligence.

Nissenbaum, H. (1996). Accountability in a computerized society. *Science and Engineering Ethics*, 2(1), 25-42.

REDISCOVERY OF AN EXISTENTIAL-CULTURAL-ETHICAL HORIZON TO UNDERSTAND THE MEANINGS OF ROBOTS, AI AND AUTONOMOUS CARS WE ENCOUNTER IN THE LIFE IN THE INFORMATION ERA IN JAPAN, SOUTHEAST ASIA AND THE 'WEST'

Makoto Nakada

University of Tsukuba (Japan)

nakadamakoto @msd.boglobe.ne.jp

EXTENDED ABSTRACT

In this paper I will make an attempt to find (rediscover) the potentially broader or alternative horizon to understand the meanings of robots, AI and autonomous cars which we encounter in the life in the information era in Japan, Southeast Asia and the 'West.' In this case, the broader or alternative (horizon) refers to the situations which would be beyond the narrowly interpreted views on human life under the strong influence of techno-determinism, Cartesian dualism of body and mind, the Western presuppositions to put an emphasis on the limited aspects of human existence (i.e. rationality, intelligence separated from bodily existence, individualism as another expression of isolation and the oblivion of vulnerabilities or finitude of life).

I will make this attempt mainly by focusing on the following points. 1) First, we will see the research findings which I gained by my researches performed in the past decade in Japan, Southeast Asia and some of the Western countries. Through this analysis we will see that people in Japan, Asia and the West tend to understand or evaluate the meanings of life by depending on their existential-cultural-ethical perspectives on 'what is a good-virtuous life?' This finding is a continuation of those findings in my previous papers in some ways but in this paper I will add another point to those: the analysis on a wholeness of human life including the aspects such as the finitude of life (death), sympathy for others' suffering from isolated death, sacrifice in the disasters and the accidents as well as the sensitivity to meanings of life leading to awareness of small happiness or beauty in life in transience. For example, one of the interesting findings through my analysis on the data is the fact that people tend to share the high evaluation on the views, 'People will become corrupt if they become too rich (Honest poverty)' in spite of differences in nationality, religion, language, degree of development of industries and informatization(the percentage of the respondents showing positive response to this view is: 75.7%(Japan in 2018), 81.6%(Indonesia in 2018), 87.0%(Vietnam in 2017), 68.7%(Germany in 2003), 57.4%(Sweden in 2019)). Similarly people tend to show the positive responses to these views: 'People have a certain destiny, no matter what form it takes'(Destiny); 'In our world, there are a number of things that cannot be explained by science'(Denial of natural science) and so on.

In a sense, this might be interpreted as a case of construction or emergence of new horizon which is made possible through my practice as presenting a new scheme or a potential frame of reference. If we combine this point with the problems of 'horizon' or 'passive synthesis' suggested by Husserl, 'Bewandtnis(a link of meanings constructed through our involvement in the world)' by Heidegger, 'Basho(a place as a thematic field of unity of the subject/the object or the perception/the understanding)' by Kitaro Nishida and 'reconstruction of horizon of perception through setting a new topicalization' by Gurwitsch or Jyunichi Murata, we might say that this is a case in which we can get a hint to think about 'how can we overcome the horizon determined by techno-determinism?' or 'how

can think about the possibility of enabling an alternative horizon to emerge?' My suggestion would be: the combination of 'a new schema about horizon as something to be made,' 'awareness of meanings of life as a kind of our experience,' 'an attempt to find a new link of meanings' and 'a new arrangement of meanings through a set of new topicalization' can be sources to create a new horizon or at least to help us to be aware of presence of an alternative horizon.

2) Secondly, we want to know how people's views on 'good and virtuous life' are related with their views on the robots, AI, autonomous cars or other various artifacts people encounter in their informatized surroundings. As we will see in this paper, one of the most interesting findings on this point in this paper is that people's awareness of meanings of death or meanings of victims in the disasters or accidents are correlated with their views on robots, AI and autonomous cars. For example, the degree of affirmative response to the view, 'I sometimes feel that I have to think more deeply about the important meanings of life when I hear the stories of persons who saved others at the cost of their own life in natural disasters and similar crises,' is found to be statistically significantly correlated with one of the views on autonomous car, 'Although automobile driving robot by artificial intelligence seems to be convenient, considering to leave judgment on life or death to the machine, there is a problem of use without much consideration' (Autonomous car's judgment for life) (the correlation coefficient =.329 with a level of significance under .01%)(data: research in Japan in 2018).

This means that people seem to encounter robots, AI and autonomous cars on a certain kind of horizon which depends upon or reflects people's awareness of meanings of a good-virtuous life as well as of a life as not separated from death or other vulnerabilities they must face sometime somewhere. This might be interpreted as a concrete case of creation of a potentially alternative horizon which goes beyond the views influenced by techno-determinism as views eliminating the questions on death or illness as part of human existence.

3)And thirdly, I will examine the discussions of various authors which might give us suggestions on the problem, 'how can we make our eyesight broader in order to see the meanings of life in the information era?' In my view, the discussions dealing with the wholeness of life including our potential awareness of those aspects of our life such as vulnerabilities of life, the finitude of life, embodiment and others would be very important. We might take into the consideration the importance of these discussions when we try to think about the meanings of robots, AI and autonomous cars, care by robots or decision on life and death by AI in a car or machine to diagnose illness in our life: the discussions on these aspects suggested by Husserl, Heidegger, Merleau-Ponty, Gadamer, Dreyfus, Introna, Capurro, Toombs, Todres, Galvin, Svenaeus and others.

In addition, I think that the discussions by some of Japanese authors (Nishida, Watsuji, Saigusa, Tokieda, Kimura, Ichikawa, Nakamura and others) would be useful in the sense that they tend to pay attention on the oneness of meanings of life, tension between Western logic and Japanese logic (the logic of predicate) or the sensitivity to meanings in the world in transience or in life on a journey beyond objective description.

Table 3 Correlation between 'sharing pity for others' death/ sacrifice' and 'various views on robots/autonomous car' in Japan. (Data: 2016HG)

	Problems of care robots	Sympathy for virtual creatures	Care robot for children	Autonomous car's judgment for life	Responsibility for autonomous car
flowers for lament	**.309****	**.145****	**.176****	**.182****	**.171****
being beautiful through transience	**.285****	**.277****	**.131****	**.273****	**.275****
sacrifice	**.344****	**.293****	**.100***	**.329****	**.351****
lonely death	**.291****	**.256****	**.169****	**.277****	**.216****
Astroboy's final episode	**.230****	**.271****	**.248****	**.152****	**.198****

1) The figures of the table show correlation coefficients. 2) **=p<0.01, *=p<0.05, without ** or *=ns= non (statistically) significant. 3) 'Flowers for lament' shows 'I can imagine clearly the figures of the victims or their family when I see the flowers for lament or sorrow at the traffic accidents or other accidents.' Similarly, 'being beautiful through transience' shows 'I can sometimes feel that the fireworks or the glow of a firefly in the summer are beautiful because they are transient or short-lived.' 'Sacrifice' shows 'I sometimes feel that I have to think more deeply about the important meanings of life when I hear the stories of persons who saved others at the cost of their own life in natural disasters and similar crises.' 'Lonely death' shows 'I sometimes feel that everyone must have had their own meaningful days even if he/she died alone and his/her death is called a case of 'lonely death' in the newspapers.' 'Astroboy's final episode' is 'I am moved when I know Astroboy's final episode as self-sacrifice for saving the earth.'

(Table 1,2,4,5,6 are omitted because of limited space.)

KEYWORDS: existential horizon, robots, autonomous car, meanings of life and death.

(PART OF)REFERENCES

Gurwitsch, Aron(1966).Phenomenology of Thematics and of the Pure Ego: Studies of the Relation Between Gestalt Theory and Phenomenology. In *Studies in Phenomenology and Psychology*. Evanston: Northwestern University Press.

Murata, Jyunichi(1995). *Tikaku to seikatsusekai*(Perception and life-world). Tokyo: University of Tokyo Press.

Todres, L., Galvin, K.T. and Holloway, I. (2009).The humanization of healthcare: A value framework for qualitative research. *International Journal of Qualitative Studies on Health and Well-being*, 4:2, 68-77.

Toombs, S.K. (1988). Illness and the Paradigm of Lived Body. *Theoretical Medicine* 9 (1988). 201-226.

SUPERIORITY OF OPEN AND DISTRIBUTED ARCHITECTURE FOR SECURE AI-BASED SERVICE DEVELOPMENT

Yoshiaki Fukami, Yohko Orito

Keio University (Japan), Ehime University (Japan)

yofukami@sfc.keio.ac.jp; orito.yohko.mm@ehime-u.ac.jp

EXTENDED ABSTRACT

The architecture of the Internet is distributed. There is no central server on the Internet. Each individual and organization issues identifiers, data and metadata. This distributed architecture enables the Internet to rapidly diffuse information and realize scalability. At the same time, such architecture creates a bottleneck that generates a massive amount of learning data for AI (Artificial Intelligence). The diversification of data specifications, including syntax and vocabulary, makes it difficult to utilize fragmented data on the Internet. In contrast, GAFA generates a massive amount of data with unified identifiers and specifications, which is one of reasons that they have established superiority in AI development.

The unified management of data has an advantage in the efficiency of AI-based service development and its applications. However, comprehensive personal information management by a few oligopolistic companies means that they may have substantial power of control over individuals. The benefit of the unified management of personal information tends to be based on the risk of abuse of power by large IT companies. However, is publicly unified management the best way to balance benefit creation with personal information protection by utilizing big data?

Medical information is a good use case to examine such architectural issues according to an ethical point of view because medical and health care services are mainly used as public services, and government plays an important role in these cases. It is not difficult to reach agreement on whether it is necessary to make medical and healthcare services more efficient with information technology. However, it is difficult to determine how much personal information should be used across multiple facilities. The case of Tamba city (Hyogo prefecture in Japan) is a suitable use case to evaluate how much and how integrated personal information is used for medical and healthcare services.

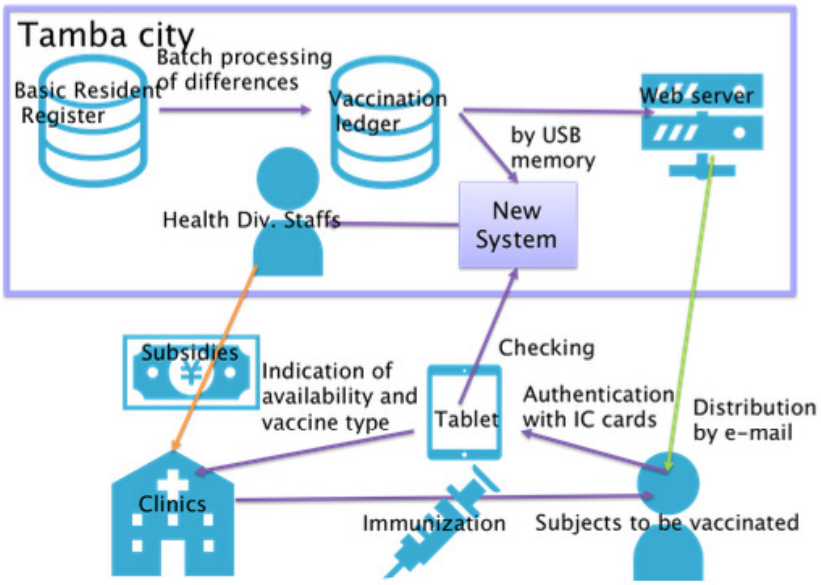
Many projects have been conducted to develop technologies to accumulate big data for AI learning. In particular, there have been many proof-of-concept projects to develop centralized electric health record (EHR) and unify distributed EHR and personal health record (PHR). Fragmented medical history data accumulation does not contribute to the improvement of the quality of medical services (Blechman, E. A. et al., 2012). There have been multiple concepts to digitalize medical records and to facilitate examination and compose prescriptions, such as computerized physician order entry (CPOE), which improves safety (Eslami et al., 2007). Clinical decision-support systems (CDSS) (El-Sappagh and El-Masri, 2007) are described as “any computer program designed to help healthcare professionals to make clinical decisions” (Shortliffe, 2011).

Tamba city, a small town located in a mountainous area has launched an immunization implementation determination system that has been implemented by linking medical and government data. Tamba city supports the costs of 15 types of vaccinations for children between 0 and 15 years old. Subsidies are paid by the city to the clinics. However, if target persons of vaccinations relocate out of the city, the subsidies are not paid, and the costs of the vaccines become the responsibility of the

clinic in that city. The medical association requested that the city eliminate the condition that the doctor bore the inoculation costs for children who were not covered by the government. As a result, Tamba city distributed tablets to clinics that were connected to a database synchronized with the basic resident register ledger via a mobile virtual network operator (MVNO) with a closed network. This enabled determination of whether a person is eligible for a grant target (Figure 1).

The immunization determination system succeeded in reducing the workload of the municipal office and eliminated mistakes. In this case, a system was developed with a simple and centralized architecture. The data resource is only the basic resident register ledger generated and is managed by the municipal office. The tablets distributed to clinics are owned by the municipal government. Personal information on the subjects is processed and managed within the facility of the municipal government.

Figure 1. Immunization determination system with a closed network in Tamba city



Tamba city and the stakeholders in the region have decided to extend the system to regional comprehensive care. This means that the system will process various types of data generated at multiple organizations and will exchange information such as prescriptions, caregiver visit records, results of a medical examination and healthcare directives.

The system was also developed with a closed network for the MVNO in compliance with the personal information protection law and is shared through tablets owned by the municipal office. Doctors working at the core hospitals in the Tamba region access computers for electric medical charts by way of exception. The architecture of the system is also centralized but used by diversified stakeholders, such as doctors. Data of prescription and actual medication history are transacted and handled only among doctors, co-medical staff and care givers.

However, the system has two ethical issues. One issue involves expanding information asymmetry between the government/medical staff and citizens. People cannot memorize all of their medical activities and their entire treatment history. Because the tablets are distributed among medical and nursing service providers, only service providers can access patient records, and citizens cannot access these records. Thus, as more records are accumulated, the asymmetry increases.

Medical records are not shared with patients. Even though citizens can accumulate more diversified data with wearable devices and smart phone applications for prescription management, they cannot manage their records because they do not have rights to access the system. While there are reasonable reasons for the limited disclosure and sharing of medical information with patients, their further engagement is needed for decisions on courses of treatment. It is important for patients and citizens to engage in determination of treatment policy, according to patient-centred medicine (Christine and Davidoff, 1996).

KPIs of regional comprehensive care must be diversified, and it is beneficial for patients to participate in the selection of metrics for KPIs because the cure rate and survival rate are inappropriate KPIs of long-term care even outside of hospitals. It is also desirable to introduce sensors chosen and owned by patients that enable multifaceted situational understanding according to patients' preferences.

Another issue is abuse of service providers. Compared to the immunization implementation determination system, much more diversified data from citizens is accumulated in the regional comprehensive care system, and a much greater diversity of engaged persons can access personal information. Regional comprehensive care information is accessed not only by medical staff but also by government employees, caregivers, and social workers. Services are provided outside of medical facilities, across the region and even in patients' homes for the long term. Therefore, the potential for fraud and blackmail based on the medical histories has increased. Even if data are shared among limited professional stakeholders, misuse of the information cannot be prevented.

How do we design a system to deter usage of shared data for purposes other than the original intents? Is it possible to monitor every activity of all service providers to control the use of data? Such types of solutions may result in other types of privacy invation by service providers. Because the system is developed with a closed network, it is impossible to monitor service providers' data usage from outside.

This tends to make the system closed to protect privacy. However, limiting access to information makes it difficult to regulate abuse by the few parties with information access. Simply developing a system with a closed network is not sufficient. The range of information to be shared and a method of access control need to be defined from patient-centric/citizen-centric perspectives.

To protect patients' privacy, a centralized and closed architecture needs to be adopted. However, such architecture does not necessarily have an advantage for the protection of human rights. An open and distributed architecture is better, even in the context of medical and nursing care, for AI-based decisions according to the concept of patient-centred medicine. This open and distributed architecture encourages self-determination in medicine and provision of appropriate care. The introduction of AI may change the rational architecture of systems where tailor-made services are developed with big data including personal information.

KEYWORDS: open and distributed architecture, secure AI-based service.

REFERENCES

Blechman, E. A., Raich, P., Raghupathi, W., & Blass, S. (2012). Strategic Value of an Unbound, Interoperable PHR Platform for Rights–Managed Care Coordination. *Communications of the Association for Information Systems*, 30, pp.83–100.

El-Sappagh, S. H., & El-Masri, S. (2011). A Proposal of Clinical Decision Support system Architecture for Distributed Electronic Health Records. In *Proceedings of the International Conference on Bioinformatics & Computational Biology (BIOCOMP)* (p. 1).

Eslami, S., Abu-Hanna, A., & de Keizer, N. F. (2007). Evaluation of Outpatient Computerized Physician Medication Order Entry Systems: A Systematic Review. *Journal of the American Medical Informatics Association*, 14(4), pp.400–406.

Laine, C., & Davidoff, F. (1996). Patient-centered Medicine: A Professional Evolution. *JAMA*, *275*(2), pp.152-156.

Shortliffe, E. H. (2011). Biomedical Informatics: Defining the Science and Its Role in Health Professional Education. In A. Holzinger & K. M. Simonic (Eds.), *Information Quality in e-Health. USAB 2011. Lecture Notes in Computer Science,* vol. 7058. (pp. 711–714). Springer, Berlin, Heidelberg.

THE ETHICS OF AUTONOMY AND LETHALITY

Elias Carayannis, Richard L. Wilson, Ion A. Iftimie, Michele C.A. Iftimie

European Union Research Center (U.S.A.), Hoffberger Center for Professional Ethics (U.S.A.), NATO Defense College (Italy), European Union Research Center (U.S.A.)

caraye@gwu.edu, wilson@towson.edu, iftimie@gwu.edu, falconqu@gmail.com

EXTENDED ABSTRACT:

This paper takes a critical look at the recent developments in AI and assess the impact of their military use from an anticipatory ethical perspective (Wilson 2019). We propose a new anticipatory ethics methodology that considers a quintuple helix ecosystem (Carayannis et al. 2018). Using this methodology, we discuss the government, industry, academia, civil society and environmental considerations, priorities and implications of using lethal autonomous weapons by the military. This anticipatory ethical discussion is critical to address as lethal autonomous weapons are expected to become the main way that we fight wars by 2040.

Asimov's three fundamental Rules of Robotics—that must be "built most deeply into a robot's positronic brain", or its artificial intelligence (AI)—state that 1) a robot may not injure a human being or, through inaction, allow a human being to come to harm; 2) a robot must obey orders given it by human beings except where such orders would conflict with the First Law; and 3) a robot must protect its own existence as long as such protection does not conflict with the First or Second Laws. This paper argues that the dismissal of Asimov's three laws by the Group of Governmental Experts (GGE) on lethal autonomous weapon systems (LAWS)—convened by the UN Conference on Certain Conventional Weapons (CCW) in Geneva from 25 to 29 March 2019—means that there is only a matter of time until lethal AI technologies become the new norms in combat environments.

Science fiction often portrays lethal autonomous weapons, or LAWS, as evil, completely out of control machines that are intent to destroy humanity, as we know it. In both the 1978 and the 2003 Battlestar Galactica series, Cylons manage to destroy the human civilization, forcing the few survivors of the human race into space. The 1984 Terminator opens up with a post-cataclysmic image of 2029 Los Angeles, where Skynet attempts to track and exterminate the remaining human survivors: "The drones have taken over, and it's futile to fight them […] Large killer battleship drones, their searchlights shining in the smoky ruins of what was once a city, float in the air, searching for humans below to kill with powerful cannons. Tank-like drones on the ground crush human skulls under their treads" (Algire et al 2013, 15). In the 2004 dystopian sci-fi action film 'I, Robot,' the 2035 NS-5 LAWS technology 'V.I.K.I.' is enhanced with advanced—logical but heartless—artificial intelligence that portrays humans as self destructive. V.I.K.I concludes that in order to protect the lives of their human masters, some people must be killed for the greater good. Finally, in the 2005 film Stealth, a 2016 U.S. Navy LAWS/drone enhanced with artificial intelligence technologies is struck by lightning, and becomes a rogue killer drone. In movie after movie, artificial intelligence and LAWS become Shelley's mental vision of the modern Prometheus. Despite this, globally, a new race for autonomous tactical weapons is taking place. In the new environment, fully autonomous tools are already being used to support "managerial and organizational cognition" (Carayannis 1999) on the battlefield and can be used to predict the actions of enemy combatants. Entering into the realm of preemption, hundreds of millions of dollars are now being invested (in the United States, alone) in the development of LAWs. At military academies and at defense conferences of the military–industrial complex, arguments are made that in

an uncertain, complex and highly-technological future, it will become impossible to defend our society from terrorists without the use of LAWS.

The United Nations (UN) is currently in the process of drafting its narrative on LAWs, which it defines as "weapon systems that, once activated, can select and engage targets without further human intervention" (Heyns 2013, 1). This definition differs significantly from that of a conventional weapon in *the selection of the target*. In the case of LAWS, algorithms imprinted on a microchip (rather than a human) *selects* the subject for the targeted killing, transcending "the role of information technology as an enabling agent" (Carayannis 1998). This raises many technological, ethical, legal and political questions about the use of LAWS in future warfare environments that the UN is currently attempting to address. More specifically, the Group of Governmental Experts (GGE) is debating whether LAWS could be programmed to "comply with the requirements of international humanitarian law and the standards protecting life under international human rights law" (Heyns 2013, 1). From a technological standpoint, UNHRC reports—on extrajudicial, summary, or arbitrary executions—are assessing that currently "no adequate system of legal accountability can be devised," and because of that, they recommend that "robots should not have the power of life and death over human beings" (Heyns 2013, 1).

The emergence of LAWS has taken "law, war, and military institutions on an uncharted path into the future at breakneck speed. Some of the transformational effects of these new weapons systems are clear; others are still emerging" (Beard 2009, 442). The use of drones brought the accountability and ethics of targeted killings under the spotlights not only in America, but also on the world stage. This paper will address the anticipatory ethical considerations behind the use of LAWS, starting with the first unsuccessful launch of 9,200 LAWS during World War II by the Japanese, to the discourses about LAWS in the contemporary security environment. By synthesizing the claims of the 1) *robots are better than humans* and of the 2) *humans are better than robots* schools of thought—and by confirming or denying their core assumptions—this paper will address the question of whether LAWS can be programmed to surpass humans in understanding context in complex situations, often described as *the fog of war*. From a legal perspective, without a clear assessment of whether autonomous robots can or cannot distinguish between legal combatants and civilians on the ground, one cannot assert whether they can be used to make life and death decisions, or whether LAWS can be used to target legal combatants. We address the applicability of *jus in bello* (unnecessary loss of life during war) in the LAWS debate, and question whether the use of LAWS will compound or alleviate the calamities of war while targeting lawful combatants during international armed conflict—between two or more states—or unlawful combatants during non-international armed conflicts—between state and non-state actors.

The principles of distinction and proportionality and the legality of targeted killings—made increasingly more salient by the use of drones and other technologies of enhanced lethality—are also discussed. This is needed in order to better understand the distinction between unlawful combatants and civilians, the balance between humanity and military necessity, and who (nature), when (use), where (location), and why (purpose) a lawful or unlawful combatant can be treated as a legitimate military objective by a lethal autonomous weapon. Finally, we address the ethics behind the political ends of using LAWS for the purpose of preventing the scourge of war (jus ad bellum) and the unnecessary loss of life during war (jus in bello). We look at the past use of drones by the United States and address when the political ends (such as cutting the human and financial costs of war) justify the means (the use of LAWS) during armed conflicts. We pose that these political considerations must always consider an anticipatory ethical analysis and that nation states and the international community have an ethical responsibility to ensure that LAWS (at all stages of development) do not become a means to terrorize civilians in the contemporary security environment. Military units charged with operating LAWS are

expected to operate by both legal and ethical considerations, and we expect that our anticipatory ethics methodology will help military commanders better navigate the dilemmas emerging from their use.

KEYWORDS: Artificial Intelligence, Lethal Autonomous Weapon Systems, Asimov Laws of Robotics, Anticipatory Ethics.

REFERENCES

6, P. (2001). Ethics, Regulation and the New Artificial Intelligence, Part II: Autonomy and Liability. Information, Communication and Society, 4:3, pp. 406-434.

Algire, D., Argentieri, C., Sullivan, M., & Rosenblatt, L. (2013). Drones: Are they Watching You? Source Interlink Media.

Amnesty International. (2012). United States of America Targeted Killing Policies Violate the Right to Life. Amnesty International Publications June. London, UK.

Asaro, P. (2007). How Just Could a Robot War Be? Presentation at 5th European Computing and Philosophy Conf., Twente, NL: June.

Asaro, P. (2013). On Banning Autonomous Weapon Systems: Human Rights, Automation, and the Dehumanization of Lethal Decision-making. International Review of the Red Cross.

Beard, J. M. (2009). Law and War in the Virtual Era. The American Journal of International Law. Volume 103. Pp 409-445.

Carayannis, E. G. (1998). "The strategic management of technological learning in project/program management: the role of extranets, intranets and intelligent agents in knowledge generation, diffusion, and leveraging." Technovation 18.11: 697-703.

Carayannis, E. G. (1999). Fostering synergies between information technology and managerial and organizational cognition: the role of knowledge management. Technovation, 19(4), 219-231.

Carayannis, E. G., Grigoroudis, E., Campbell, D. F. J., Meissner, D., & Stamati, D. (2018). “The Ecosystem as Helix: An Exploratory Theory-building Study of Regional Co-opetitive Entrepreneurial Ecosystems as Quadruple/Quintuple Helix Innovation Models.” R&D Management 48 (1).

Conley, C. W. (1968). The Great Japanese Balloon Offensive. Air University Review, January-February. Downloaded from http://www.airpower.maxwell.af.mil/airchronicles/aureview/1968/jan-feb/conley.html.

Marchant, G. (2011). International governance of autonomous military robots. Columbia Science and Technology Law Review, Volume XII.

Heyns, C. (2013). UNHRC (A/HRC/23/47): Report of the Special Rapporteur on extrajudicial, summary or arbitrary executions. 9 April 2013.

Kizza, J. M. (2013). New Frontiers for Computer Ethics: Artificial Intelligence. In Ethical and Social Issues in the Information Age, pp. 201-210: Springer London.

Kleffner, J. K. (2012). Section IX of the ICRC Interpretive Guidance on Direct Participation in Hostilities. Israel Law Review 45(1), pp. 35–52. Cambridge University Press and The Faculty of Law, The Hebrew University of Jerusalem.

Masters, J. (2013). Targeted Killings. Council on Foreign Relations. http://www.cfr.org/counterterrorism/targeted-killings/p9627. January 8, 2013.

Mena, J. (2011). Machine Learning Forensics for Law Enforcement, Security, and Intelligence. Boca Raton, FL: Taylor and Francis Group.

Shah, N., Holewinski, S., & Lucas, L. (2012). The Civilian Impact of Drones: Unexamined Costs, Unanswered Questions. Columbia Human Rights Clinic.

Sharkey, N. (2008). Grounds for Discrimination: Autonomous Robot Weapons. RUSI Defence Systems; available from http://rusi.org/downloads/assets/23sharkey.pdf

Sharkey, N. (2010). Saying 'No!' to Lethal Autonomous Targeting. Journal of Military Ethics. Vol. 9, Iss. 4. Pp. 369-383.

Sharkey, N., & Suchman, L. (2013). Wishful Mnemonics and Autonomous Killing Machines. AISB Quarterly. No. 136, May pp. 14-22.

Singer, P. (2008). Wired For War: The Robotics Revolution and Conflict in the 21st Century. New York, NY: Penguin Group.

Sparrow, R. (2006). Killer Robots. Journal of Applied Philosophy, Vol. 24, No.1.

Thurnher, F. S. (2012). No One at the Controls: Legal Implications of Fully Autonomous Targeting. Joint Force Quarterly 67 (4th Quarter, October 2012).

Von Clausewitz, C. (2008). On War. Princeton, NJ: Princeton University Press.

Wilson, R. L. (2019, October). Requirements for Making the MQ-9 fully Autonomous: An Anticipatory Ethical Analysis. In ECIAIR 2019 European Conference on the Impact of Artificial Intelligence and Robotics (p. 369). Academic Conferences and publishing limited.

Winnefeld, J., & Kendall, F. (2011). The Unmanned Systems Integrated Roadmap FY 2011-2036. U.S. Department of Defense.

Zenko, M. (2013). Reforming U.S. Drone Strike Policies. Council on Foreign Relations Press. Council Special Report No. 65. January. http://www.cfr.org/wars-and-warfare/reforming-us-drone-strike-policies/p29736

WHAT ALPHAGO BRINGS FOR THE PROFESSIONAL PLAYER IN THE GAME OF GO, AND NEAR FUTURE IN OUR SOCIETY?

Akira Uchino

Senshu University (Japan)

uchino@isc.senshu-u.ac.jp

EXTENDED ABSTRACT

The game of chess, Shogi (Japanese chess), and Go are the longest-studied domain in the history of artificial intelligence (AI). Especially, Go has long been viewed as the most challenging of classic game for AI owing to its enormous search space and the difficulty of evaluating board positions and moves. But AlphaGo, a computer program developed by Google DeepMind, defeated a human professional player in 2016, and AlphaGo Master again in 2017. Google DeepMind announced AlphaZero algorithm, starting from random play and given no domain knowledge except the game rules, defeated a world champion program in the games of chess, and Shogi, as well as Go. That means artificial intelligence exceeds human brain in such intellectual domain. The professional players have been responding to the emerging AI software and experiencing the new changing world.

We should record what has happened for these years, think about the reason why happed, the meaning of it for human-being, and so on. Here I emphasize what the “deep learning” brings for us and how to treat such a superpower AI system from now on.

1. INTRODUCTION

Artificial intelligence (AI) has the longest-studied domain in the intellectual game domain. Deep Blue by IBM won the world campion in 1997. After that Shogi and Go are the second target of AI. But in 1997 Shogi game soft was uppermost 3 dan which is intermediate advanced class amateur player far from professionals and Go game soft could enjoy only the beginners.

All games of perfect solution have an optimal value function, which determines the outcome of the game, from every board position or state. The game may be solved by recursively computing the optimal value function in tree search containing approximately possible sequence of moves.

The Table 1 shows the volume of search space and the score to top level player. We could have complete solution of Checker now, but exhaustive search is infeasible for Chess, Shogi, and Go. Shogi is similar type of game as Chess, using 9 by 9 board and 20 piece each player, far less than Chess. But you can use enemy’s piece when you capture (take) it, so 40 pieces are available through the game. The Chess becomes less complex because of decreasing pieces on the board in the end part of the game.

In Chess and Shogi, position evaluation is easy, and each piece has own value. So, program can be coded using value function. Go is another kind of Chess type game, in which the aim is to surround more territory than the opponent. Each black stone and white stone are the same one and same value, and you place a stone on an intersection on 19 by 19 board. You have huge options in vacant intersections to place your stone.

Abstract shows the game of Go has long been viewed as the most challenging of games for AI owing to its enormous search space and the difficulty of evaluating board positions and moves. But AlphaGo, a computer program developed by Google DeepMind, defeated a human professional player in the full-sized game of Go in 2016, and AlphaGo Master again in 2017. Google DeepMind announced AlphaZero algorithm, starting from random play and given no domain knowledge except the game rules, convincingly defeated a world champion program in the games of chess, and Shogi, as well as Go.

There are several works about artificial intelligence and intelligent game, which we must research.

1) We should record what has happened in this domain. The record is historical value for AI history.
2) We must check what kind of technics or methods develop the stages. We can recognize the effectiveness of them.
3) In the process of human brain lose gameplay to AI, there happens social interactions. What happens suggests for the recognition of relations between human-being and AI now on.
4) AlphaZero algorithm, starting from random play and given no domain knowledge except the game rules, is mechanical learning process from scratch. Human-being have studied *Joseki*, the standard moves in part of the games for several hundreds of years. Go AI soft learns *Joseki* by itself. The learning process shows sometimes meaningless moves for human-being, and we often find excellent new *Joseki*. Comparing the learning process of AI with human-being is interesting to research.

Table 1. Human-being vs Computer Soft

Volume of search space	Games to top-level human player
10 to the 30th power	***Chinook*** won the world champion in 1994. The soft retired to human players in 1996.
	The game proved if the players plays the best, the game is a draw.
10 to the 60th power	***Logistello*** won world cahmion Murakami 6 to 0 in 1997.
10 to the 120th power	***Deep blue*** by IBM won the the world champion in 1997.
10 to the 220th power	***Bonanza*** lose Title holder but led the game to the middle of the game in 2007.
	Soft won to professional players within 30 seconds game about 75 percentage in 2011.
	GPS Shogi won top professional plyayer Hitoyuki Miura in 2013.
	Ponanza won Shogi Chamion Amahiko Sato in 2017.
10 to the 360th power	MTCS(Monte Calo Tree Search) emerged in Go soft in 2006.
	Go soft reached 5 dan in amateur level but still need much time (5 to 10 years) to win the professional players in Feb. 2015
	Alpha Go paper published in Jan. 2016.
	Alpha Go won the world champion Lee Sedo 4 to 1 in March 2016.
	Alpha GO Master won the world rank number 1 player Ke Jie 3 to 0 in May 2017.
	Alpha GO Zero emerged without huma knowledge in Oct. 2017.
	Alpha Zero emerged in Dec. 2018. General Game soft won the existing top game softs.

2. HOW GO AI SOFT DEVEOPS TILL 2015

When Deep Blue won the world campion in 1997, Shogi game soft was uppermost 3 dan which is intermediate advanced class amateur player far from professional player and Go game soft could enjoy only the beginner. In Chess and Shogi, position evaluation is not so difficult, and each piece has own value. We can program game code for them to value function.

Go is another kind of Chess type game, in which the aim is to surround more territory than the opponent. When a chain of stones has more than two “eyes”, or two vacant points adjacent to a stone, a chain is alive and vacant points are counted as a territory. But when a chain is surrounded by opposing stones and it has no liberties, it is dead or killed, so that it is captured and removed from the board. Go has two characteristics, one for getting more territory and one for putting obstacle in the way of other territory, which means to fighting to live or dead. In Go, each black stone and white stone is the same one and same value, and you place a stone on an intersection on 19 by 19 board. You have huge options in vacant intersections to place your stone. It is so difficult to capture value function compare to Chess and Shogi; Go AI soft improved so slowly. In 2005 Crazy Stone by Remi Colom has appeared. Crazy Stone use Monte Carlo tree search (MTCS) which was regarded as impossible to use in Go game. Random choice to proceed the game with powerful computer resources became useful at that time. After Crazy Stones, Go AI soft have rapidly improved one year by one year, but early 2010s, the progress is nearly stopped. Go AI soft reached 5 dan in amateur level in Feb. 2015 but still need much time, 5 to 10 years to win the professional players.

3. HOW ALPAHGO DEVELOPS

Table 2 shows how AlphaGo develops. When first AlphaGo paper published on January 28th in 2016, Europa Champion was not top professional Go player so that the difference was 3 dan so that we prospected the world champion would win 4 to 1 or 5 to 0. But for several months the AlphaGo became stronger unexpectedly so that the score was 1 to 4. The world champion Lee Sedo won only one game, but it was weak point of the AlphaGo and the week point adjusted after the game. In 2016 human-being lost the game of Go by AI.

Table 2. Versions of AlphaGo

Version	Competition results	Technics, Metohds	Hardware specin competition
AlphaGo Fun	won Europa Champion in Oct. 2015 (Published on Jan. 28 in 2016)	[AlphaGo Paper]	176GPU 48TPU
		CNN(Convolutional neural network)	
		SL(supervised learning) policy network	
		RL(reinforced learning) policy network	
		Policy gradient method, DQN (deep Q-network) value network	
		MTCS (Monte Carlo Tree Search)	
AlphaGo Lee	won the world champion Lee Sedo 4 to 1 in March 2016	difference in AlphaGo Fun	176GPU 48TPU
		using policy network from self-play	
		over 14 layer neural network	
AlphaGo Master	consecutive 60 wins in the Go website in the early 2017	difference in AlphaGo Zero	4TPU
		MCTS, the same playout as AlphaGo Fan	
	won the world rank number 1 player Ke Jie 3 to 0 in May 2017		
AlphaGo Zero	won AlphaGo Master 89 to 11 (Published on Oct.19 in 2017)	[AlphaGo Zero Paper] using 39 residual network	4TPU
		dual netowork(integrate policy network and value network)	
		self-play reinforcement learning	
		MCTS without playout (using win rate prediction of dual network)	
Alpha Zero	Alpha Zero General reinforced algorithm (Published on Dec. 7 in 2018)	Chess: equivalent to ***Stockfish***	
		Shogi: over ***Elmo***	
		Go: equivalent to ***AlphaGo***	
		general self-play reinforcement learning ⇒ possibility of general application	

4. AFTER ALPHAGO IN PROFESSIONAL FIELDAI

In Shogi world, Shogi AI soft reached almost at the professional level in 2013. Japan Shogi Association tried to avoid the professional Shogi players losing the Shogi AI Soft and restricted professionals from playing soft in public. Shogi AI soft used mechanical learning include self-play reinforcement learning.

In Go world, at the beginning of 2015 we thought it takes 5-10 years to win the professional Go players. But it took only 2 years. AlphaGo uses “deep learning”, it is the key to success to win the professional Go player. Self-play reinforcement learning without “deep learning”, the Go AI soft could reach as strong as professional level but could not reach the God hand. Deep learning brings human-being to see God hand player in the game of Shogi and Go.

AlphaGo retired to match human-being. Other Go AI softs have developed by deep learning and self-play reinforcement learning. Now many professional Go players use Go AI softs to study. There are new styles and new *Joseki,* the standard moves in part of the games, effected by Go AI softs. If top professional Go player challenge 3,000 to 5,000 times to AlphaGo, I think the player would win one game; that means the Go AI soft is not 100% perfect.

5. REMAINING ISSUES

There are remaining issues we should research;

1) What has happened in the professional field in Go and Shogi?

   How do the professional players use AI soft?

   What brings the soft to the professional fields?

2) Are there problems about Go AI soft?

   The rule of the game of Go is not so difficult. But Go has mysterious aspects. There is the possibility that both players cannot proceed the game because the same situation continues eternity. Once the superko, like the threefold repetition rule of Chess, has emerges, can the AI soft control the situation? Go AI soft uses Chinese rule so that correct komi cannot be treated. Go AI soft uses winning percentage as value function and there is no concern how much maximum territory the player gets. In the end game, the soft doesn’t show the best next hand.

3) If “deep learning” start from random play and given no domain knowledge except the game rules, what kind of different learning process between human-being and “deep learning” are there in Joseki.

6. CONCLUSION

We use AI artefacts to improve our life. AI artefacts are surely useful and most of the case, the decision of the AI is better than human-being. But Go AI soft is not perfect. If top professional Go player challenge 3,000 to 5,000 times to the most powerful Go AI soft, the player would win one game. If the human-being entrust their life to AI artefacts, black box AI soft makes a mistake by chance. what brings to the table.

KEYWORDS: the play of Go, self-play, reinforcement learning, neural network, deep learning.

REFERENCES

David et al. (2016, January 28), "Mastering the game of GO with deep neural networks and tree search" *Nature* 529(484-489)

David et al. (2017, October 19), "Mastering the game of Go without Human Knowledge" *Nature* 550(354-355)

David et al. (2018, December 7), "A genreral reinfrocement learning algorithm that masters chess, shogi, and Go through self-play" *Science* 362(1140-1144)

11. Open Track

Track chairs: Kiyoshi Murata, Meiji University, Japan – Ana María Lara Palma, Universidad de Burgos, Spain – Yohko Orito, Ehime University, Japan – Alicia Izquierdo-Yusta, Universidad de Burgos, Spain

A META-REVIEW OF RESPONSIBLE RESEARCH AND INNOVATION CASE STUDIES - REVIEWING THE BENEFITS TO INDUSTRY OF ENGAGEMENT WITH RRI

Vincent Bryce

De Montfort University (United Kingdom)

p2558837@my365.dmu.ac.uk

EXTENDED ABSTRACT

Responsible Research and Innovation (RRI) provides a framework that may overcome computer ethics problems such as the increasingly ubiquitous nature of computing technologies, the global nature of innovation, and the need to consider accountability at different stages of the innovation lifecycle (Stahl et al., 2014). It shares with computer ethics the challenges of demonstrating relevance to, and providing practical guidance for, industry (Gotterbarn, 1994; Stahl, 2015).

This paper will answer the following research question - what is the relationship between RRI implementation practices and outcomes for firms, taking into account contextual variables such as company size and sector? It will tackle this question using a meta-review of published RRI case studies.

The contribution it makes to knowledge will be to explore and quantify relationships between RRI practices, and outcomes for businesses. It responds to the need identified in Martinuzzi et al. (2018) for more quantitative research to get from 'perceptions to evidence', to explore the 'business case' for corporate engagement with RRI, and to relate RRI more explicitly to adjacent discourses on corporate responsibility.

Studies of RRI to date have mainly used qualitative designs. Lubberink et al.'s (2017) review identifies "few scholars who empirically investigated responsible innovation practices in commercial R&D settings", and that more than two-thirds of empirical RRI articles were based on case study research. While predating recent studies this:

1. indicates that the field of RRI has until recently focussed on empirical exploration and description;
2. highlights a need for larger-scale and quantitative empirical testing, and;
3. as managerial decision-making is frequently based in quantifiable evidence, signals that more quantitative research is essential for future development of the RRI field (Martinuzzi et al. 2018).

The proliferation of RRI case studies is an opportunity to synthesise findings through meta-review, to explore generalised relationships between RRI implementation and outcomes, and highlight practices with stronger associations to certain outcomes taking into account variables such as company size.

Discussion of RRI measurement has tended to focus on society-level impact (for example Von Schomberg, 2013 p8-12), or an individuated, company-specific concept of business case development and measurement based on 'RRI KPIs' or a company-specific RRI 'Roadmap' (Porcari et al., 2018; Yaghmaei, 2018).

While an important principle for industry guidance - benefits of RRI engagement can, and perhaps should be assessed in relation to a firm's business strategy - it leaves questions unanswered. Can a

general 'business case' for firms engaging with RRI be identified? Which RRI practices are associated with positive business outcomes in different contexts? Beyond this – while practices such as public or employee engagement are associated with positive organisational outcomes, can a 'value-adding' effect for organisations who implement broad-focus RRI across the anticipate-reflect-engage-respond spectrum (Stilgoe et al., 2013) be observed beyond effects which might be expected from component practices? Within these questions – given that a company's implementation of RRI may be at either a strategic, or operational level (Stahl et al., 2017; van de Poel et al., 2017) - to what extent are benefits evidenced when RRI is adopted at a strategy, rather than project level?

Relating RRI to long-standing discourses on Corporate Social Responsibility (CSR) offers opportunities to apply approaches developed in the CSR literature in support of RRI research questions.

A case can be made for the relevance of CSR 'tools' in informing RRI implementation practice (Iatridis & Schroeder, 2015). Similarly for measurement, the evolution of RRI maturity models has been informed by the availability of CSR models drawing on a wide empirical evidence base (Martinuzzi & Krumay, 2013; Stahl et al., 2017).

RRI and CSR share the challenges of definitional complexity, and difficulty in identifying empirical attributes. However while contested (for example Banerjee, 2008), the concept of CSR benefits from having been the subject of significant theory building and research. In particular the 'business case for corporate responsibility', which can be defined as "how the business community benefit tangibly from engaging in CSR policies, activities, and practices" (Carroll & Shabana, 2010) has been exposed to empirical scrutiny since the 1960s, including through meta-review (for example Caroll & Shabana, ibid).

A meta-review methodology offers the opportunity to identify then synthesise a range of RRI case studies (Tranfield et al. 2003). The SLR principles of Tranfield et al. (ibid) as employed by Lubberink et al. (2017) and Thapa et al. (2019) will be used to identify relevant studies. The conceptual framework set out in van de Poel et al. (2017) will be used as a basis for classifying published RRI case studies in terms of specific RRI drivers, tools and outcomes identified. Features of the RRI implementation context such as organisation type and sector will be included in the analysis. The resulting data will be assessed to identify patterns and relationships between context, implementation practices, barriers, and outcomes.

A benefit of this approach will be flexibility in interpretation of RRI in terms of its local implementation. Van de Poel et al.'s (2017) taxonomy of RRI practices established that the concept of 'RRI tools' should be understood to include a range of practices, not limited to those developed specifically for RRI. Similarly, a wide range of effects may be relevant to a business in terms of RRI-related outcomes (Yaghmaei, 2018).

The study will consider potential limitations of sampling bias (in the availability of company case studies), lack of longitudinal perspective, and lack of equivalence of case study methodologies. Duration of case study will be considered where available and will qualify observations on the RRI practices-RRI outcomes relationship. Inclusion criteria for relevance will provide a check to ensure case studies are only included if they consider relevant dimensions of RRI practice and outcome, and criteria will be revisited during data collection as needed. The study will consider the level of analysis of case studies, using the definitions in the RRI maturity model developed by Stahl et al. (2017) to distinguish 'project-based', from 'strategic' implementations.

The methodology developed will help pave the way towards a broader approach to evaluating the business case for companies to engage with RRI practices.

KEYWORDS: responsible research and innovation, RRI; responsible innovation, corporate social responsibility, CSR; industry.

REFERENCES

Banerjee, S. B. (2008). Corporate social responsibility: The good, the bad and the ugly. *Critical Sociology*. https://doi.org/10.1177/0896920507084623

Carroll, A. B., & Shabana, K. M. (2010). The business case for corporate social responsibility: A review of concepts, research and practice. *International Journal of Management Reviews*, *12*(1), 85–105. https://doi.org/10.1111/j.1468-2370.2009.00275.x

Dreyer, M., Chefneux, L., Goldberg, A., von Heimburg, J., Patrignani, N., Schofield, M., & Shilling, C. (2017). Responsible innovation: A complementary view from industry with proposals for bridging different perspectives. *Sustainability (Switzerland)*, *9*(10), 1–25. https://doi.org/10.3390/su9101719

Iatridis, K., & Schroeder, D. (2015). Responsible Research and Innovation in Industry: The Case for Corporate Responsibility Tools. In *Responsible Research and Innovation in Industry: The Case for Corporate Responsibility Tools*. https://doi.org/10.1007/978-3-319-21693-5

Lubberink, R., Blok, V., Ophem, J. van, & Omta, O. (2017). Lessons for responsible innovation in the business context: A systematic literature review of responsible, social and sustainable innovation practices. *Sustainability (Switzerland)*, Vol. 9. https://doi.org/10.3390/su9050721

Martinuzzi, A., Blok, V., Brem, A., Stahl, B., & Schönherr, N. (2018). Responsible Research and Innovation in industry-challenges, insights and perspectives. *Sustainability (Switzerland)*, *10*(3), 1–9. https://doi.org/10.3390/su10030702

Martinuzzi, A., & Krumay, B. (2013). The Good, the Bad, and the Successful - How Corporate Social Responsibility Leads to Competitive Advantage and Organizational Transformation. *Journal of Change Management*, *13*(4), 424–443. https://doi.org/10.1080/14697017.2013.851953

Porcari, A., Pimponi, D., Borsella, E., Mantovani, E., Van De Poel, I., Flipse, S., Cibien, M. (n.d.). *Prisma - Guidelines to Innovate Responsibly - Prisma Roadmap to Integrate RRI into Industrial Strategies*. Retrieved from https://www.rri-prisma.eu/wp-content/uploads/2019/09/PrismaRRI_Roadmap_brief_web.pdf

Stahl, B. C., Obach, M., Yaghmaei, E., Ikonen, V., Chatfield, K., & Brem, A. (2017). The responsible research and innovation (RRI) maturity model: Linking theory and practice. *Sustainability (Switzerland)*. https://doi.org/10.3390/su9061036

Stilgoe, J., Owen, R., & Macnaghten, P. (2013). Developing a framework for responsible innovation. *Research Policy*, *42*(9), 1568–1580. https://doi.org/10.1016/j.respol.2013.05.008

Thapa, R. K., Iakovleva, T., & Foss, L. (2019). Responsible research and innovation: a systematic review of the literature and its applications to regional studies. *European Planning Studies*. https://doi.org/10.1080/09654313.2019.1625871

Timmermans, J., Yaghmaei, E., Stahl, B. C., & Brem, A. (2017). Research and innovation processes revisited – networked responsibility in industry. *Sustainability Accounting, Management and Policy Journal*. https://doi.org/10.1108/SAMPJ-04-2015-0023

Tranfield, D., Denyer, D., & Smart, P. (2003). Towards a Methodology for Developing Evidence-Informed Management Knowledge by Means of Systematic Review. *British Journal of Management*, *14*(3), 207–222. https://doi.org/10.1111/1467-8551.00375

van de Poel, I., Asveld, L., Flipse, S., Klaassen, P., Scholten, V., & Yaghmaei, E. (2017). Company strategies for responsible research and innovation (RRI): A conceptual model. *Sustainability (Switzerland)*, *9*(11). https://doi.org/10.3390/su9112045

Von Schomberg, R. (2013). A Vision of Responsible Research and Innovation. *Responsible Innovation: Managing the Responsible Emergence of Science and Innovation in Society*, 51–74. https://doi.org/10.1002/9781118551424.ch3

Yaghmaei, E. (2018). Responsible research and innovation key performance indicators in industry: A case study in the ICT domain. *Journal of Information, Communication and Ethics in Society*, *16*(2), 214–234. https://doi.org/10.1108/JICES-11-2017-0066

AI AND ETHICS FOR CHILDREN: HOW AI CAN CONTRIBUTE TO CHILDREN'S WELLBEING AND MITIGATE ETHICAL CONCERNS IN CHILD DEVELOPMENT

Ryoko Asai

Uppsala University (Sweden), Meiji University (Japan)

ryoko.asai@it.uu.se

EXTENDED ABSTRACT

In 2013, Frey and Osborne published the sensational report that highly advanced technologies, especially AI (Artificial Intelligence) technology, would bring rapid innovation to both public and business sectors, and take over many jobs from us (Frey and Osborne 2013; Frey, Osborne and Citi 2015). Also World Economic Forum predicted that people who engage in non-skilled or manualized work would face risks of losing or changing jobs in the future. This technological trend would penetrate not only public and business sectors but also private life (marriage, childrearing, family time etc.) and change our lifestyle. In the AI era, children are inevitably affected by AI, in the form of education, recreational activities, communication etc. In this study, we consider about ethical concerns which arise from the use of AI for childcare, and explore how AI affects children's development ethically from the perspective of information ethics.

In high-tech society, we use technologies as given commodity and suppose that they make life more efficient and effective by utilizing them. For example, AI and highly advanced technologies are deployed for daily use at home, and take over household work and daily chores. In most cases, AI is equipped within processing systems and is not visible to us. We can only see the results that AI processes and derives under a certain command. Because of this invisible feature, AI has been deployed in our living environment and penetrated deeply into society without even noticing. For children, this situation is even more emphasized and they grow up as digital natives.

However, it is not clear how AI influences the children's development, especially its influences on their ethical sense and their social values. It is because this technological trend is still new and there is only little accumulation of research about AI and child development from an ethical perspective. Although there is some research on social robots and children from the viewpoint of cognitive physiology or behavioral science, the research from an ethical point of view is necessary to analyze and interpret the ethical impact on children and also to consider how AI can contribute to children's wellbeing. In this study, we consider about ethical impacts on children who are in daily interactions with social robots for family and private use, which is regarded as one of the most visible forms of AI to children.

The previous research showed that social robots for family use have generally three basic functions; Entertainment functions (singing, dancing and playing game); Security functions (monitoring through webcam, talking from a distance via Internet); Facilitation and revitalization of family communication (providing family a trigger for conversations) (Asai 2017). Although social robots cannot clean the house or cook foods, they can sing a song with children, read a book for children before going to sleep, or check children and rooms via webcam when parents are absent from home.

There are some ethical and practical problems with using social robots for children. 1) As long as social robots function based on our personal data, there is a risk to breach privacy or leak personal information. 2) Manipulating social robots needs to use "robot infrastructure" in order to operate cloud AI and robot OS for collecting and analyzing data. And, social robots are operated with the

collaboration and cooperation of various technologies and hundreds of applications. How and who manages and controls the robot's infrastructure is critical to protect our privacy and personal data. 3) Social robots are customized for particular users through interaction and communication with them. Each social robot is made up by the collaboration of robot designers, engineers, venders, operators and users, and its functions are differentiated and customized by AI. Under this situation, it is not self-evident who should take responsibility, in case that social robots cause social problems such as violating a third person's privacy by its photo-sharing function.

However, serious ethical problems are hidden behind the visible and useful functions of AI. Generally there are three basic ethical concerns in the use of social robots including care robots. 1) There is a risk which users get socially excluded or socially isolated because of too much attachment and emotional connection to social robots. The social relationship is superseded by the relationship with social robots (Sparrow and Sparrow 2006; van Wynsberghe 2016). 2) Privacy and integrity of users (caretakers) might be damaged by social robots (Vallor2016). 3) Social robots might generate new inequality between "robot-have"/"robot-not-have" or skilled users/unskilled users, based on age, income level, the development level of countries and societies and so on. Or, existing disparities might be amplified by the use of social robots (Asai 2017).

In addition to the above ethical concerns, there are more serious problems, which are not only invisible but also hard-to-recognize for most of us. First, it is very difficult to be free from embedded values in designing and developing technology (Friedman et al. 2006; Nissenbaum2011). There is no clue for normal users to know how AI or algorithms work in the system. We might get in touch with biased ideas without us noticing, via the interaction with AI. This problem has currently been revealed by some research about AI and fairness, for example, the research on the numbers of mug shots and normal pictures in searching images online based on skin colors.

Second, when children stay with social robots, interaction is constructed by children's order. Basically social robots say yes to children and provide services required by them. We hardly imagine that social robots deny us or disagree with us, except robots having a special algorithm to deny users. Moreover, AI offers more reasonable and rational environments based on well-calculated and well-programmed algorithms, when compared to the environment controlled by humans. Under these conditions, children would grow up without learning how to handle uncontrollable situations for them and how to be tolerant in such a situation.

Third, when children acquire their experience via some functions of social robots or virtual reality technology, their experience could be completed inside the room. It would affect the quality of experiences. Generally, experience is categorized as direct experience (in-person experience), indirect experience, and pseudo direct experience (Ichikawa 1992). Experience via virtual technology is regarded as pseudo direct experience. In other words, even though they don't have any direct experience to see, touch and feel things in person, they can easily see and feel something very similar to the real via advanced technologies. However, it causes lack of in-person experience. That means that children can live without communicating with others, and they cannot recognize and share any feelings towards others. Consequently they cannot position their own existence and identity in society (Ichikawa 1992). Actually, we have already started to adjust ourselves to the current technological environment in order to maximize efficiency and benefits from technology. It would be even more so for children in the age of AI. Although benefits from AI are remarkable and attractive for us, we definitely need to consider how we mitigate ethical concerns for child development and how AI can contribute to children's wellbeing.

In the conference ETHICOMP 2020, we'd like to argue in detail about the invisible and hard-to-recognize ethical concerns with the use of AI for childcare, and discuss possible solutions to the problems, from the viewpoint of information ethics.

KEYWORDS: Artificial Intelligence, Child Development, Children's Wellbeing, Experience, Information Ethics, Social Robot.

REFERENCES

Asai, R. (2017). Techno-Parenting. *The Journal of Information and Management*, Vol. 37, No. 2, 6-21.

Frey, C.B. and Osborn, M.A. (2013). *The Future of Employment: How Susceptible Are Jobs to Computerisation?* Oxford Martin School, University of Oxford.

Frey, C.B., Osborne, M.A. and Citi. (2015). Technology at Work: The Future of Innovation and Employment. *Citi GPS: Global Perspectives & Solutions*, February 2015.

Friedman, B., Kahn, P.H. and Borning, A. (2006). Value Sensitive Design and Information Systems. in Zhang, P. and Galletta, D. (Eds.), *Human-Computer Interaction and Management Information Systems: Foundations*, M. E. Sharpe (republished by Routledge in 2015), pp. 348-372.

Ichikawa, H. (1992). *Seishin toshite no shintai.* Kodansha (in Japanese).

Nissenbaum, H. (2001). How Computer Systems Embody Values. *Computer*, 34(3), March 2001, 118-120.

Sparrow, R. and Sparrow, L. (2006). In The Hands of Machines? The Future of Aged Care. Minds and Machines, 16(2), 141-161.

Vallor, S. (2016). *Technology and the Virtues: A Philosophical Guide to A Future Worth Wanting*, Oxford University Press.

van Wynsberghe, A. (2015). *Healthcare Robots: Ethics, Design and Implementations*, Ashgate Publishing (re-published by Routledge in 2016).

ALGORITHMS, SOCIETY AND ECONOMIC INEQUALITY

Juho Vaiste, Anne-Marie Tuikka, Tapio Vepsäläinen

University of Turku (Finland)

juho.vaiste@utu.fi; anne-marie.tuikka@utu.fi; tapio.a.vepsalainen@utu.fi

EXTENDED ABSTRACT

The concept of the algorithm society refers to a shifted society where algorithmic systems, robots and AI agents respond for even growing share of societal and economic decision-making (Balkin, 2017). The hype around the emerging information technologies is eminent, and companies and nations are heavily investing in algorithmic systems and attracting research talent to join their ranks (Gibney, 2016). As a result, the presence of algorithms in society is continually increasing. While the rise of the emerging technologies can have many positive effects on humans and societies (Stanford University, 2016), it can also fuel unbalanced and unequal economic development. This extended abstract on the algorithmic society and economic inequality discusses different emerging factors that will significantly shape the developed societies in the near future.

Normative ethics theories have different answers to the question of economic inequality (White, 2019). Still, it can be argued that most of the theories oppose uncontrolled and unbridled inequality. Theoretical literature and novel evidence, the algorithmic revolution will most likely have broad economic impacts (Stanford, 2016) – also in the light of the creation of economic inequality if not appropriately handled (Kharlamova et al., 2018).

In this conceptual study, we analyse phenomena, which are related to wide adaption of algorithmic systems and discuss how they influence our society now and in the future. Our research question is: How algorithms are related to inequality in the society? To answer this question, we concentrate on three phenomena: data-driven centralization of businesses, the income share between labour and capital, and the economic impacts of algorithm biases. We aim to conceptualize the relevant phenomena and their relationships with the essential normative moral objections of inequality – *objections to violations of equal concerns, inequalities in status, interference with the fairness of economic and political institutions, and economic institutions that generate large differences in outcome* – presented by T.M. Scanlon (2018).

Developing algorithmic systems requires processing power, data and tools. Arguably, processing power and computer capacity are rather cheap today. Similarly, most of the machine learning tools are open source. However, the real asset, data, in the Western societies and large parts of the rest of the world is concentrating into the hands of a few significant tech-giants, namely Google, Amazon and Facebook. The importance of data only keeps on growing as more sophisticated algorithms are developed.

At the same time, the tech-giants have used digitalization to create new business models that massively benefit from economies of scale. They have become monopolies in their markets which have led to data-driven centralization of business. For example, Google constantly scans a huge amount of online materials and labels them to be able to instantly give answers to the user. Additionally, they collect data about the behaviour of users searching for similar search words. Every search operation contributes to Googles online search supremacy. An alternative search engine would need to have superior algorithm to beat Google in search, because Google already knows what the users want to

see. The gap between alternative services and Google grows every day, because Google controls 90% of online search.

There is an extended discussion on the emerging technologies and the future of work. It is unclear whether algorithms and AI technologies are replacing or complementing human workers (Seamans, 2017). In specific fields of business, such as transportation, digital platforms are primarily responsible for matching service providers with the people who need services. For example, Uber is offering a digital platform for ridesharing, which aims to connect drivers with people who need a ride (Rosenblat & Stark, 2016). Unlike traditional taxi services, Uber does not employ drivers directly; instead, the drivers are classified as independent contractors. For this reason, they are personally responsible for insurances, licenses and maintenance of the vehicle. The quality of their work performance is automatically monitored by analysing the ratings given by the passengers, and if the rating drops too low, the driver will be deactivated from the Uber application. Similar developments have been detected in different sectors, such as food service and retail, where standard shifts are replaced by shifts scheduled on-demand based on algorithmic predictions (Campolo et al., 2017).

If many jobs will disappear within the next two decades and working conditions within many occupations will decrease, the polarization of developed societies will likely continue. Future scenarios (e.g., Rifkin, 1995; Ford, 2015) of societies with high unemployment, low paid service jobs and well-paid knowledge workers seem now more and more probable. Such development would probably decrease the labour's share of national and global income, hence changing the income share between labour and capital.

The algorithmic decision-making does not only affect businesses, but also the everyday life of people living in modern societies. One reason for this is that algorithms and AI systems are imperfect. Programming fairness to algorithmic systems is difficult. Algorithms are not biased, per se. Still, they are unfair because they judge individuals based on reference groups behaviour instead of individuals' actions. (Corbett-Davies et al., 2017). The phenomenon of algorithmic bias is built on forms of discrimination and has negative consequences from societies' perspective. Algorithmic unfairness has several ethical and social causes, which can affect great harm for minority groups and selected subpopulations. For example, people in poverty may pay a higher price or face redlining for services, such as insurances, because of their reference groups' higher default risk (O'Neil, 2017). Hence, algorithm-based decision-making could make it more difficult for certain groups to climb up from poverty than for others.

Algorithmic biases are reasonably recognized, but especially concrete and practical approaches to their economic impacts are missing (Campolo et al., 2017) The concerns on algorithmic decision-making for public goods have been taken seriously (Goodman & Flaxman, 2016), but the link between algorithm unfairness and tech corporations needs to be constructed more carefully.

In this extended abstract, we have introduced phenomena which are related to emerging importance of algorithms in modern societies and which may be entangled with inequality in developed societies today and in future. We will study these concepts and their interrelations further and examine their relationship with the Scanlon's moral objections of inequality. As a result, we may find new phenomenon which need to be included or discard some of the phenomena described in this extended abstract. Our aim for the full article is to build a conceptual framework for the growing use of algorithms in society and its relationship with the normative moral objections of inequality. This framework will include essential phenomena, their interrelations and their impacts.

Analysing emerging use of algorithms in the society and their relationship to different phenomena through lenses of normative ethics can help to build detailed insights to economic inequality. A fine-grained analysis of normative ethics can help us to approach the phenomena of technological

inequality via the methods of economics. The final paper takes part in the recent discussions of measuring inequalities (Chan et al., 2019), and introduces the possibility to extend experimental case-by-case economics methods to the field of technology derived inequality.

KEYWORDS: Algorithms, Artificial Intelligence, Algorithmic Society, Economic Inequality.

REFERENCES

Balkin, J. (2017). The Three Laws of Robotics in the Age of Big Data. Ohio State Law Journal, 78, 1217-1227.

Campolo, A., Sanfilippo, M., Whittaker, & M., Crawford, K. (2017). AI Now 2017 Report. Edited by A. Selbst & S. Barocas. Retrieved from https://ainowinstitute.org/AI_Now_2017_Report.pdf

Chan, K. C., Lenard, C. T., Mills, T. M., & Williams, R. F. (2019). Measuring inequality in society. Communications in Statistics-Theory and Methods, 48(1), 88-99.

Corbett-Davies, S., Pierson, E., Feller, A., Goel, S., & Huq, A. (2017). Algorithmic decision making and the cost of fairness. KDD '17 Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Halifax, NS, Canada, August 13-17.

Ford, M. (2015). Rise of the Robots: Technology and the Threat of a Jobless Future. Basic Books.

Gibney, E. (2016). AI talent grab sparks excitement and concern. Nature, 532, 422-423.

Goodman, B., & Flaxman, S. (2016). European Union regulations on algorithmic decision-making and a "right to explanation". ICML Workshop on Human Interpretability in Machine Learning (WHI 2016), New York, NY.

Kharlamova, G., Stavytskyy, A., & Zarotiadis, G. (2018). The impact of technological changes on income inequality: the EU states case study. Journal of international studies, 11(2), 76-94.

O'Neil, C. (2017). Weapons of math destruction: How big data increases inequality and threatens democracy. Broadway Books.

Rifkin, J. (1995). The end of work: The decline of the global labor force and the dawn of the post-market era. GP Putnam's Sons, New York.

Rosenblat, A., & Stark, L. (2016). Algorithmic Labor and Information Asymmetries: A Case Study of Uber's Drivers. International Journal of Communication, 10, 3758–3784.

Scanlon, T. (2018). Why does inequality matter? Oxford University Press.

Seamans, R. (2017, January 11). We Won't Even Know If A Robot Takes Your Job. Forbes. Retrieved from https://www.forbes.com/sites/washingtonbytes/2017/01/11/we-wont-even-know-if-a-robot-takes-your-job

Stanford University. (2016). Artificial intelligence and life in 2030. Report of the One Hundred Year Study on Artificial Intelligence (AI100).

Stone, P., Brooks, R., Brynjolfsson, E., Calo, R., Etzioni, O., Hager, G., Hirschberg, J., Kalyanakrishnan, S., Kamar, E., Kraus, S., Leyton-Brown, K., Parkes, D., Press, W., Saxenian, A., Shah, J., Tambe, M., & Teller, A. (2016) Artificial Intelligence and Life in 2030. One Hundred Year Study on Artificial Intelligence: Report of the 2015-2016 Study Panel. Stanford, CA: Stanford University. Retrieved from http://ai100.stanford.edu/2016-report.

White, M. (Ed.) (2019). The Oxford Handbook of Ethics and Economics. Oxford University Press.

ARTIFICIAL INTELLIGENCE AND GENDER: HOW AI WILL CHANGE WOMEN'S WORK IN JAPAN

Ryoko Asai

Uppsala University (Sweden), Meiji University (Japan)

ryoko.asai@it.uu.se

EXTENDED ABSTRACT

When the research by Frey and Osborne was released in 2013, their prediction stirred up strong anxiety among people. According to their research, many people would be replaced by AI (Artificial Intelligence) technology and loose their job in 10 to 20 years (Frey and Osborne 2013). Three years after the sensational research, in 2016, the Japanese Ministry of Health, Labour and Welfare released a report, which shows that technological innovation would require Japanese people to adjust themselves to a flexible working style which is free from a working place, working time or employment status in the near future 2035 (Japanese Ministry of Health, Labour and Welfare 2016). A series of the shocking reports made Japanese even more nervous about unemployment because Japan has been taking a very positive and active attitude towards introducing AI or robots into workplaces.

On the other hand, some research pointed out that AI and highly advanced technology would improve working environments, and also promote working conditions and status for women. Japanese women have been suffering from gender gaps in working conditions and gender discrimination in the workplace for long years. Even though many social policies to redress these gender problems have been enforced until now, there are still big disparities between men and women in wage standard, employment status etc. Is it really possible for AI to change or improve women's working conditions and redress gender gaps in Japan? This study explores how AI changes women's work and eliminates gender gaps at workplaces in Japan.

Until now a lot of research on Japanese women's work has been done, for example, how new advanced technologies contribute to work-life balance, or how remote working or telework influences women's working style. They proposed various aspects of technology and women's work (Shiota 1985; Green and Adam eds. 2001). Furthermore, the utilization of female labor force is politically considered as an important solution to overcome a worker shortage, as the birthrate drops and the society is rapidly graying in Japan. In 2015, Japanese government enforced the new law called *Act on Promotion of Women's Participation and Advancement in the Workplace* (Act No. 64 of September 4, 2015), which aims to "promote female participation and career advancement in the workplace". In a series of policies to promote women's work, technology is referred to as an important factor to achieve the goals of the policies. However, it none of the policies it is actually specified how to use technology for the sake of women's work. Even the research on AI use in Japan indicated low probability of promoting women's social status by utilizing AI in society, and it also revealed that Japanese people consider the use of AI for solving gender problems as the least important matter (GLOCOM 2018). One says that AI is an important factor to promote women's employment status, and another says that AI is less important to solve gender problems. What will happen with Japanese female workers if AI takes over a job from people in the near future?

This study considers three possible scenarios in terms of AI and gender gap in the Japanese workplace. The first scenario is that introducing AI to the workplace improves women's working conditions and

makes women's presence stronger in society. Whereas manualized and routine work would be replaced by AI in the near future, new jobs would be created in order to support AI-based society and also more job opportunities would open for workers with high-skills or creativity, regardless of gender (Ministry of Internal Affairs and Communications 2017; 242-253). Women would fit better than men to the flexible work style that AI offers. This is because Japanese women have a long history and gained a lot of experiences in the limited/flexible work style over the past decades (Ministry of Internal Affairs and Communications 2017; 248). Moreover, some AI researchers mentioned that women would take an important role in some jobs, which are difficult for AI to take over (HeForShe 2018). Because those jobs require workers to have high interpersonal communication skills, and traditionally women occupied such jobs in Japan.

Many of those jobs are considered as emotional labor, and categorized as pink-collar job. That means that AI relegates Japanese women to emotional labor and pink-collar jobs. Even though AI opens more job opportunities to women and increases the number of female workers, those women would have a risk for mental health issues caused by emotional labor, and more women would engage in pink-collar jobs at lower wages. In this case, the statistical data of working women would be increased by introducing AI, and it would look like that Japan achieves gender equality numerically. However, there is a big risk that AI strengthens the gender gap in an occupational choice and in wage levels.

The second scenario is that the basic income system would eliminate gender gaps in the age of AI. When AI takes over jobs and many people lose their job or they don't need to work any longer, the wage gap would be mitigated regardless of gender. However, since people cannot live without income, the government needs to organize the social security system and the public safety net to guarantee and support people's basic life. Recently Japanese politicians and experts on social policy have discussed the basic income system in Japan, though it is far from becoming reality at the moment. However, people who create or control AI exist and they have little risk of losing their job. There would be a big disparity between people who create and control AI (*AI-have*) and those who work under control of AI (*AI-not-have*).

In this scenario, AI might cause more serious social inequality by the basic income system. Because it would be very difficult to allocate economical value created by AI to all people equally. In case that the *AI-have* gain more economical benefits than the *AI-not-have* based on AI program, it would cause social satisfaction and social unfairness among the *AI-not-have*, even though they keep a basic standard of life with their basic income. Consequently, the gender gap could be mitigated by deploying AI in the workplace and introducing the basic income system to society. However, there is a social risk to divide people into the rich *AI-have* and the *AI-not-have*.

The third scenario is that politics would redress the gender gap by laws and policies. As we argue above, it is not clear whether AI contributes to gender equality in the workplace. It doesn't mean that opening more job opportunities to women becomes the fundamental solution to eliminate gender gaps, as long as AI relegates them to specified jobs. Then, why did the Japanese government suddenly stride to social policies to promote women's working status? It arises from strong concerns over the deteriorating economic situation due to an aging society and a declining birthrate. In order to maintain the strength of the Japanese economy, Japanese government tries to use and retain women as the labor force under gender equality as the ethical and societal goal.

Therefore, improving women's working status by utilizing AI is conducted based on political and economical reasons, and it is not contemplated carefully from an ethical perspective or human rights. In fact, although working mothers want to come back work after delivering babies, many of them cannot find any child daycare service and quit their job to take care of their child at home. The Japanese

government has quickly enforced the new law to use and retain female labor force for economic reasons without changing any other gender problems in society.

In any of the scenarios, it is obscure that AI can contribute to gender equality in the Japanese working place. AI might be able to improve women's working environment and allow them to work more flexibly. However, we always need to ask ourselves about what the ultimate goal is and what kind of gender equality we aim to achieve by using AI. Otherwise, we easily get confused about means and purpose, and AI will strengthen gendered situations and social disparities.

KEYWORDS: Artificial Intelligence, Gender, Gender Equality, Japanese women, Workplace.

REFERENCES

Fray, C.B. and Osborne, M.A. (2013). The Future of Employment; How susceptible are jobs to computerisation?, Oxford Martin School, University of Oxford.

Green, E. and Adam, A. (2001). *Virtual Gender: Technology, consumption and identity*, Oxford, UK: Routledge.

HeForShe (2017). Innovation changes jobs in the future. Retrieved from https://logmi.jp/business/articles/198824

IUJ Global Communication Center. (2017). *Annual Report of Innovation Nippon 2017*. Retrieved from http://www.innovation-nippon.jp/?p=681

Japanese Ministry of Health, Labour and Welfare. (2016). *The Report of Future Working Style 2035*. Retrieved from http://www.mhlw.go.jp/stf/shingi2/0000132314.html

Ministry of Internal Affairs and Communications. (2017). *White Paper on Information and Communications in Japan 2016*. Retrieved from http://www.soumu.go.jp/johotsusintokei/white paper/ja/h28/pdf/

Shiota, S. (1985). Chapter 5. Technology Innovation and Women's Work in during the period of high economic growth, in Nakamura, M. ed. *Technology Innovation and Women's Work*, United Nations University, 171-201.

AT FACE VALUE: THE LEGAL RAMIFICATIONS OF FACE RECOGNITION TECHNOLOGY

Ivanna Villamil

Interamerican University School Of Law (Puerto Rico)

ivanna.villamil@gmail.com

EXTENDED ABSTRACT

Once exclusively a trope from our favorite sci-fi media, face recognition has become one of the most common forms of biometric technology in our everyday lives. And while it doesn't feel like we'll be joining Starfleet any time soon, Robocop seems to be right around the corner. In this paper I will be talking about how facial recognition has developed throughout the last five decades, what has it been used for, who has been using it - and most importantly - in what way is it impacting our everyday lives.

Face recognition technology started being developed around the mid-sixties, it works by identifying certain key features in a human face such as the eyes, the nose, and the mouth and creating what would look like a map. It wasn't very accurate of course, any shift in position, mark, or sign of aging would divert the result. Nowadays, while not perfect, facial recognition is able to consider every angle of the human face and take that three-dimensional image and create a two-dimensional map to make a face print. This technology can be found virtually anywhere now, most of us use it every single day to unlock our electronic devices such as our phones or laptops, and if you have a Facebook account - you might have noticed how it automatically recognizes certain users when you upload a photo of them. And since 2010, Facebook photo recognition has evolved to a point where there is barely any error with a 97.35% accuracy rate. This is because the AI (artificial intelligence) most face recognition technology works with learns as it goes and considering Facebook's tagging system has been live for almost a decade now and received three-hundred million photos a day, its system is close to flawless when it comes to identifying it's users.

Earlier this year the city of San Francisco took a stand against face recognition technology being used by the police and other similar authorities, many criticized the move, calling it a form of neo-luddism. The state of Massachusetts and Capitol Hill have followed suit, limiting everything from surveillance systems to marketing ads that might track anyone without their consent, claiming it can and will become a danger.

The truth is, face recognition could easily become a weapon against civilians in the right circumstances, in Hong Kong protesters have made a point of destroying facial recognition towers that may potentially expose them to the Chinese authorities. And if the towers couldn't be destroyed, protesters would use umbrellas and other items to cover their faces when walking by cameras. Even worse, regardless of how impressive the technology has become, it's still not perfect. Joy Buolamwini, a scientist from MIT, published his research on the racial biased in face recognition technology, this showed a high error rate on women, specifically women of color. And overall performed more accurately on distinguishing facial features of white males. So, in the long run, you don't only have the issue of "marking" certain individuals and violating their right to freedom and privacy, but there's also a good chance you'll have the wrong person. We have already seen this happen in the city of Detroit, where police have used unreliable face recognition technology to point out suspects. But in a city where most of the residents are people of color, the misidentification cases are pouring in by the dozen. This completely flips the

meaning of innocent until proven guilty, and soon freedom will be taken from those who the AI mistook for a criminal – and then, who will you ask to testify?

Privacy is another factor that is at stake, companies are using dynamic face recognition technology to sell products and are combining both your face print and purchasing history to make up the perfect consumer profile. Applications such as Snapchat and Instagram have already proven that face detection is beneficial for traffic, so companies such Walgreens are joining the game by installing cameras in front of their products so they can accurately market to you. Digital marketing is almost easier – nothing is sacred when it comes to your internet history, especially the websites you buy in. Facebook once again, owning one of the most impressive databases with ninety billion face prints in their system and a constant circulation of ads would be the perfect hire for any company to find its rightful audience.

hen face recognition starts not only being a practical everyday tool, but a way for institutions to monitor their civilians, that's when it's important for jurists to pay attention. Because apart from the bans and regulations in a handful of cities, face recognition is currently not properly monitored by the law. Regulations such as the Intelligence Authorization Act (2019-2020) brought forth by the United States government only go over issues that might be bigger threats, such as similar identification systems being used in United State soil by other countries.

This paper ends to properly explore a legal solution to regulate what already is one of the biggest threats to the privacy and freedom of common folk.

KEYWORDS: face recognition, facial detection, freedom of speech, privacy, facial recognition.

REFERENCES

Abby Everett Jaques (2019). Why the Moral Machine Is A Monster? Retrieved from https://robots.law.miami.edu/2019/wp-content/uploads/2019/03/MoralMachineMonster.pdf

Jayson DeMers (2017). Face Recognition Could Drive the Next Online Marketing. Retrieved from https://www.forbes.com/sites/jaysondemers/2017/11/27/

Joy Buolamwini (2019) When the Robot Doesn't See Dark Skin. Retrieved from https://www.nytimes.com/2018/06/21/opinion/facial-analysis-technology-bias.html

Matthew A. Turk, Alex P. Pentland (1991). Face Recognition Using Eigen Faces. Retrieved from https://www.cin.ufpe.br/~rps/Artigos/Face%20Recognition%20Using%20Eigenfaces.pdf

Nathan Freed Wessler (2019) A Federal Court Sounds the Alarm On the Privacy Harms of Fac Recognition Technology. Retrieved from https://www.aclu.org/blog/privacy-technology/surveillance-technologies/federal-court-sounds-alarm-privacy-harms-face.

Robert J. Baron (1981). Mechanisms of Human Face Recognition. Retrieved from https://www.sciencedirect.com/science/article/pii/S0020737381800016

Tam Kulik (2019). In Your Face: How Facial Recognition Databases See Copyright Law But Not Your Privacy. Retrieved from https://abovethelaw.com/2019/04/in-your-face-how-facial-recognition-databases-see-copyright-law-but-not-your-privacy/.

Vicky Bruce, Andy Young (1986). Understanding Facial Recognition. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1111/j.2044-8295.1986.tb02199.x

Yaroslav Kuflinski (2019). How Ethical Is Facial Recognition Technology?

COLLECTED FOR ONE REASON, USED FOR ANOTHER: THE EMERGENCE OF REFUGEE DATA IN UGANDA

Annabel Mwagalanyi

De Montfort University (United Kingdom)

P06285689@my365.dmu.ac.uk

EXTENDED ABSTRACT

Although there is substantial research on the surveillance of refugees in developed countries, there is relatively limited research on the topic in developing countries. This is partly because these countries have only recently begun implementing surveillance technologies to manage their refugee population. Due to their growing numbers and their changing demographics, it is becoming more urgent to study the lived experiences of refugees in this region as they are becoming increasingly subjected to digital governance, surveillance, and control.

In this extended abstract the focus will be on the emergence of digital surveillance technologies, used for governing the refugee population in Uganda. Although governments justify the use of this technology to facilitate more inclusion, in reality, it can have unintended consequences that might compromise the lives and dignity of refugees in at least four ways:

1. Even if this information is collected for good purposes (e.g. for fair distribution of food and resources) it could be used in the future for the bad, giving future governments more power in an unstable region.
2. Surveillance affects refugee's behaviour and perception of the host country.
3. Data can have errors or be compromised which could lead to people being treated inappropriately. With the absence of legal frameworks such as the General Data Protection Regulation, this makes it more of a problem.
4. Aggregating data is conducive to treating people as monolithic collectives rather than particular individuals (racial profiling).

The aim is to gain insight that could influence policymakers, appreciating the consequence of surveillance technologies in developing countries.

Refugee data collection is often carried out at the arrival stage of their journey to a refugee camp or a country border point. This is to enable the refugee to receive assistance or the chance to seek asylum. In the interim, it acts as a form of identity. By 2030 the United Nations' goal is to ensure all human beings have some form of identification, but this is a difficult task as many hurdles need overcoming such as the misuse of refugee data when it is used above and beyond what it was intended for originally. There is occasion when the data is used for legitimate reasons and in line with government policies, however, it has a significant impact on refugee lives (e.g. EURODAC fingerprinting asylum system in the EU). We have always had surveillance technology in operation in some form or another but the emergence of new technologies raises concerns in developing countries such as Uganda with over 1 million refugees, heavily dependent on foreign aid but yet still in need of measures to help manage the refugee crisis. This abstract will, therefore, attempt to discuss these issues in more detail.

In 2018, the UN refugee agency rolled out a major refugee verification operation and the project aims to ensure all refugees are registered and receive the protection and assistance they need (e.g. biometric identification and food ration cards). The organisation uses its software; however, the Ugandan phase of the project is the biggest in the agency's history (UNHCR, 2018).

For a refugee, it is important to receive an identity as it acts as a gateway to education, employment, and health services, which by law all human beings are entitled to (Maitland, 2018). Furthermore, identification also provides self-worthiness acting strongly as an integration tool to society pulling a refugee away from living in isolation which many find themselves when forcibly displaced. In contrast, the process of collecting biometric data such as fingerprints, facial recognition, and iris scans is not always as pleasant as discovered in Europe.

A key European institution associated with the digital surveillance of asylum seekers is known as the EURODAC (European Dactyloscopy Scheme). But the EU's biometric database, operational since 2003, is the subject of public controversy. Critics of Eurodac claim it violates human rights. The reason for this argument is because the initial purpose of the system was to gather all asylum claims made in the EU region but was then integrated with Europol (Sánchez-Monedero, 2018). This extension was made without consent and left refugees asylum records being contrasted with criminal records. Another criticism is the police have access to the database therefore often treating refugees like criminals and suspects. EURODAC has birthed non-state initiatives, including private border patrols, counterfeit border checkpoints. Further developments include iBorderCtrl an automated deception detection system using artificial intelligence. The system uses a virtual agent to conduct asylum interviews asking questions about migrant's backgrounds and intentions. However, the Guardian newspaper spoke to experts in the field who argued it is almost impossible to design an experiment that evaluates deception behaviour. The program assumes migrants potentially may be lying and this has a negative impact as it can make them feel treated unfairly and that the host country is being hostile.

Another example of impacting refugee life is the use of data from mobile phones and social media in the EU. The data is used during asylum evaluation interviews to detect a person's accent for example (Meaker, 2018). In this scenario refugees are in a predicament because digital devices such as mobile phones have become an indispensable tool, guiding them along migration routes and supplying information for their asylum claims. Agencies have access to text messages, location reports and browsing history despite it being deleted by the phone owner. This raises the question of who benefits from systems of detection and control such as EURODAC in a time where its methods are being adopted by developing countries to tackle the refugee crisis.

Modern identity systems promise to bring many benefits to Africa. But as they proliferate, so too will the temptation for politicians to misuse it (The Economist, 2019). Data protection laws lack in the African continent and cannot be automatically enforced like the GDPR in Europe. The impact of this is that refugees, a group perceived as citizens of nowhere, whose interests are not represented by governments are at high risk for exploitation. Although they have very little control over the situation they are in, their identity is being challenged and constructed anew by forces greater than themselves. Some experience enforced iris scans in return for aid, their phones may be seized as a form of identity verification, and biometrics are used for categorization or evaluation of their rights and benefits. Due to the lack of legal frameworks governing data in Africa, there have been instances where mobile phone operators such as Orange were discovered to be offering Africans fewer digital rights than their European subscribers. In 2018 Ugandan officials exaggerated refugee figures by 300,000, fake names were created in refugee settlements and defrauded millions of dollars in aid (Okiror, 2019). Officials from the office of the prime minister were suspended. This demonstrates how refugee data can be misused impacting their lives (e.g. less aid due to sponsor reducing aid) all due to short-term greed.

Africa has been lagging when it comes to addressing privacy issues around data argues Gwagwa (2019), but it cannot afford to do so any longer because the state can surveil and censor data traffic for self-serving purposes. New technologies are often created to help solve humanitarian issues but often exceed their initial intentions leaving unintended consequences (Maitland, 2018). Soliman (2016) pointed out if the data falls in the wrong hands creates vulnerabilities for refugees, however, if the data is not shared can leave many countries open to security threats. It is of great importance that governments; policymakers and organisations are made aware of the potential damage new technological developments create.

KEYWORDS: Refugee, Asylum seeker, Surveillance, Technology, Biometric.

REFERENCES

Ajana, B. (2013). *Governing through biometrics*. Basingstoke: Palgrave Macmillan.

Capurro, R. (2008). Intercultural information ethics: foundations and applications. Journal of Information, Communication and Ethics in Society, 6(2), pp.116-126.

Dahir, A. (2019). *Africa isn't ready to protect its citizens personal data even as EU champions digital privacy*. [online] Quartz Africa. Available at: https://qz.com/africa/1271756/africa-isnt-ready-to-protect-its-citizens-personal-data-even-as-eu-champions-digital-privacy/

Maitland, C. (2018). *Digital lifeline?*. Westchester Publishing Services.

Okiror, S. (2019). *Inquiry finds refugee numbers were exaggerated by 300,000 in Uganda*. [online] the Guardian. Available at: https://www.theguardian.com/global-development/2018/oct/30/inquiry-finds-refugee-numbers-exaggerated-in-uganda

Sandvik, K., Gabrielsen Jumbert, M., Karlsrud, J. and Kaufmann, M. (2014). Humanitarian technology: a critical research agenda. *International Review of the Red Cross*, 96(893), pp.219-242.

The Economist. (2019). *African countries are struggling to build robust identity systems*. [online] Available at: https://www.economist.com/middle-east-and-africa/2019/12/05/african-countries-are-struggling-to-build-robust-identity-systems

The Economist. (2019). *Establishing identity is a vital, risky and changing business*. [online] Available at: https://www.economist.com/christmas-specials/2018/12/18/establishing-identity-is-a-vital-risky-and-changing-business

The New Dark Age. (2019). *AI Border Guards are Being Tested at the Edge of Fortress Europe, Away From Public Scrutiny*. [online] Available at: https://williambowles.info/2019/12/06/ai-border-guards-are-being-tested-at-the-edge-of-fortress-europe-away-from-public-scrutiny/

Rand.org. (2019). *Tracking Refugees with Biometrics: More Questions Than Answers*. [online] Available at: https://www.rand.org/blog/2016/03/tracking-refugees-with-biometrics-more-questions-than.html

Refugees, U. (2019). Uganda launches major refugee verification operation. [online] UNHCR. Available at: https://www.unhcr.org/news/latest/2018/3/5a9959444/uganda-launches-major-refugee-verification-operation.html

Refugees, U. (2019). Understanding datafication in relation to social justice' (DATAJUSTICE) starting grant (2018-2023).

DIGITAL CAPITAL AND SOCIOTECHNICAL IMAGINARIES: ENVISAGING FUTURE HOME TECH WITH LOW-INCOME COMMUNITIES

Roxana Moroșanu Firth, Catherine Flick

De Montfort University (UK), De Montfort University (UK)

roxana.firth@dmu.ac.uk; cflick@dmu.ac.uk

EXTENDED ABSTRACT

Whose imagination is considered in the development of emerging technologies? Apart from engineers and tech developers, other voices contribute opinion and critical considerations on new technologies, ranging from utopian to dystopian visions. These voices belong, between others, to academics, journalists, law practitioners, policymakers and third-sector organisations representing concerned citizens. Both the opportunities, and the perils, brought about by envisaged technological advancements, are discussed in media spaces by experts and citizens who, arguably, have a high digital capital – a good understanding of new technologies and of the fact that the ways of using them can harm, or benefit one. Such debates inform the activities of large tech companies as well as shaping the development of new policies in preparation for anticipated technological futures. However, some voices are missing from debates on the direction and effects of future sociotechnical change. These are the voices of people who are on the other side of the digital divide, where having lower digital access and competence means missing out on the opportunity to express opinion on emerging technologies.

This paper argues for creating a space in debates on sociotechnical change where it is possible to articulate sociotechnical imaginaries regardless of the level of digital capital of those involved. We approach this purpose twofold. First, we bring together two concepts that were developed within distinct theoretical traditions: the concept of digital capital (Park, 2017), as well as related work on digital divides and digital inequalities (Ragnedda, 2017); and the concept of sociotechnical imaginaries (Jasanoff and Kim, 2009). We look at the theoretical underpinnings of these concepts and we discuss the work that needs to be done in order to place them in dialogue. Second, we outline a set of methodological tools that can be employed to explore sociotechnical imaginaries in creative and open-ended ways. Inspired by arts-based methods, these tools do not create differentiation between research participants, for example with regards to their level of digital access and competence, but encourage equal participation based upon creative expression.

The concept of digital capital emerged from literature looking at digital social inequalities and earlier conceptualizations of what has been called the digital divide. This divide was first defined in terms of access to new ICT technologies, such as personal computers and the internet. The lack of access was associated with existing inequalities in income, class, gender, race and age, and it was understood as a form of social exclusion, where those who did not have access missed out on important information and opportunities (DiMaggio et al., 2001; Selwyn, 2004). Further work suggested that the digital divide could not be explained in term of access alone, but that one needs to account for the differences in how people used new technologies, as certain types of uses can increase access to social capital, while others 'may downright disadvantage the uninformed' (Hargittai 2008). The skills (Livingstone and Helsper, 2010), or literacies (Warschauer, 2003) required to navigate online information safely and effectively, were also associated to pre-existing social inequalities, such as inequalities in education (Robinson et al., 2003), and in income, gender and place (Gilbert et al., 2008). Recent work accounting

on intersectionality in social inequalities (Anthias, 2013) has led to the concept of digital capital, which captures the multidimensional set of dispositions that individuals develop to engage with new and envisaged technologies (Park, 2017; Ragnedda, 2018). The reason why we prefer to use the concept of digital capital in this paper, rather than just referring to digital divides and digital inequalities, is because it allows for change and fluctuation. While one's socioeconomic circumstances might be slower to change, individuals are able to increase their digital capital by developing their abilities to use technologies and digital services.

On the other hand, the concept of sociotechnical imaginaries comes from science and technology studies (STS) literature addressing collective imagination and action with regards to future sociotechnical transformation. The concept refers to visions of desirable futures that support specific advancements in science and technology (Jasanoff and Kim, 2009; Jasanoff, 2015). Such visions are often proposed by institutions that have the ability to act upon their materialization, such as the state or corporate actors (Sadowski and Bendor, 2019). Indeed, the concept has been mostly used, so far, in work looking at the implementation of big technology initiatives, such as nuclear power (Sovacool and Ramana, 2015) and smart city infrastructure (Sadowski and Bendor, 2019). While sociotechnical imaginaries have proved to be a useful concept for analysing dominant narratives on envisaged sociotechnical change, we suggest that this concept can be fruitfully employed for discussing alternative visions as well. Alternative sociotechnical imaginaries refer to the ways in which people envisage desirable futures that might differ from the dominant visions of sociotechnical transformation. We believe that this usage of the concept of sociotechnical imaginaries is aligned to how the concept was initially described, as accounting for 'the growing recognition that the capacity to imagine futures is a crucial constitutive element in social and political life' (Jasanoff and Kim, 2009: 121). By addressing the sociotechnical imaginaries of people who might not have the opportunities to implement them, we bring these alternative visions into debates on sociotechnical change, thus enriching those debates and accounting for multiple visions of desirable futures.

Our proposition of looking at the sociotechnical imaginaries of people who have what might be considered a low digital capital entails the suggestion that imaginaries are not necessarily dependent on abilities and skills. This claim will be supported with reference to a study with low-income communities that addressed imaginaries of future technologies in domestic settings. Regardless of the participants' digital skills and access to digital technologies, our study focussed on encouraging the participants to explore their visions of future technologies in the home. By referring to a familiar setting – the home – our aim was to show that the role of new technologies is not necessarily pre-determined when these technologies are designed, but it emerges from the ways in which people use digital devices in everyday settings. This emphasis was maintained in the methodological design of the study as well. We employed a set of arts-based methods that encouraged creative expression, giving the participants the opportunity to explore and give shape to their ideas about future technologies by collaborating in creating art pieces. These creative exercises facilitated further dialogue and empowered the participants to express their own visions of desirable futures.

In the next part of the paper we will look at the theoretical underpinnings of the concepts of digital capital and sociotechnical imaginaries, and discuss what it means for these concepts to be mobilized together. We will then discuss the benefits of placing these concepts in dialogue, with reference to the findings that emerged from our study.

KEYWORDS: creative methods, digital capital, digital divide, digital inequalities, futures, sociotechnical imaginaries.

REFERENCES

Anthias, F. 2013. 'Hierarchies of social location, class and intersectionality: Towards a translocational frame. *International Sociology* 28(1): 121-138.

DiMaggio, P., Hargittai, E., Neuman, R., Robinson, J. 2001. 'Social implications of the internet'. *Annual Review of Sociology*, 27: 307-336.

Gilbert, M., Masucci, M., Homko, C., Bove, A. 2008. 'Theorizing the digital divide: information and communication technology use frameworks among poor women using a telemedicine system'. *Geoforum* 39: 912-925.

Hargittai, E. 2008. 'The digital reproduction of inequality'. In Grusky, D. (ed.) *Social Stratification*. Boulder: Westview Press.

Jasanoff, S. 2015. 'Future imperfect: Science, technology, and the imagination of modernity'. In Jasanoff, S. and Kim, S-H. (eds.) *Dreamscapes of Modernity: Sociotechnical Imaginaries and the Fabrication of Power.* Chicago: University of Chicago Press.

Jasanoff, S., Kim, S-H. 2009. 'Containing the atom: Sociotechnical imaginaries and nuclear power in the United States and South Korea. *Minerva* 47(2): 119-146.

Livingstone, S., Helsper, E. 2010. 'Balancing opportunities and risks in teenagers' use of the internet: The role of online skills and internet self-efficacy. *New Media and Society* 12(2): 309-329.

Park, S. 2017. *Digital Capital*. London: Palgrave.

Ragnedda, M. 2017. *The Third Digital Divide: A Weberian Approach to Digital Inequalities*. Oxford: Routledge.

Ragnedda, M. 2018. 'Conceptualizing digital capital'. *Telematics and Informatics* 35(8): 2366-2375.

Robinson, J., DiMaggio, P., Hargittai, E. 2003. 'New social survey perspectives on the digital divide'. *IT & Society* 1(5): 1-22.

Sadowski, J., Bendor, R. 2019. 'Selling smartness: Corporate narratives and the Smart City as a Sociotechnical Imaginary. *Science, Technology & Human Values* 44(3): 540-563.

Selwyn, N. 2004. 'Reconsidering political and popular understandings of the digital divide. *New Media and Society*, 6(3): 341-362.

Sovacool, B., Ramana, M. 2015. 'Back to the future: Small modular reactors, nuclear fantasies, and symbolic convergence'. *Science, Technology & Human Values* 40(1): 96-125.

Warschauer, M. 2003. *Technology and Social Inclusion: Rethinking the digital divide.* Cambridge, MA: The MIT Press.

DIGITAL CONFLICTS

Mario Arias-Oliva, Antonio Pérez-Portabella, Elena Ferrán, Mar Souto-Romero, Juan Luis López-Galiacho Perona

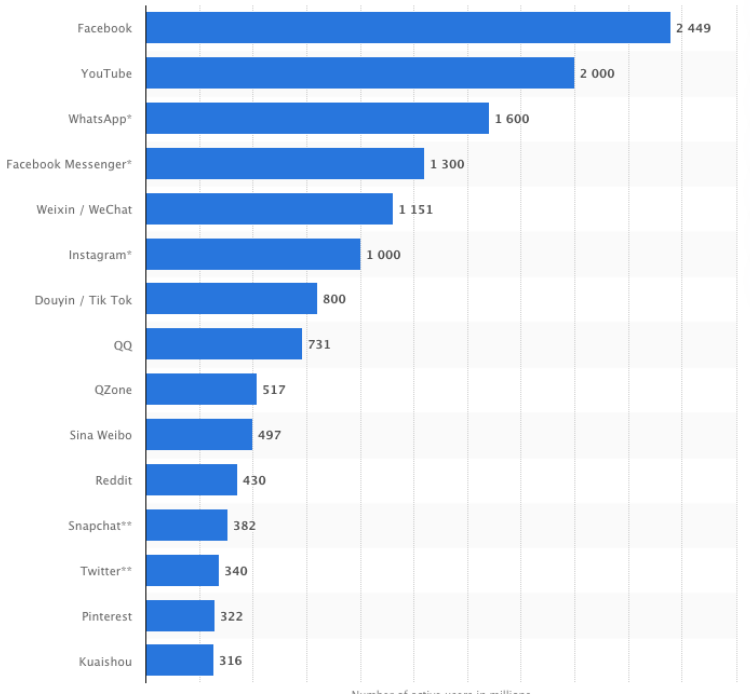
Universitat Rovira i Virgili (Spain), Universitat Rovira i Virgili (Spain), Escuela Oficia de Idiomas de Tarragona (Spain), Universidad Internacional de La Rioja (Spain), Universidad Rey Juan Carlos (Spain)

mario.arias@urv.cat; antonio.perezportabella@urv.cat; viveyveras@gmail.com; mar.souto@unir.net; juanluis.lopezgaliacho@urjc.es

EXTENDED ABSTRACT

Nowadays, everything is moving from the real world to the digital one. The role and influence of digital technologies on societal conflicts is an emerging topic with many ethical implications. Traditional language manipulation techniques and propaganda are overcome by new emerging digital tools, dividing and confronting citizens, cultures and societies. Social Media and digital technologies provide new tools either to create new digital conflicts or to foster the existing ones. Social Networks represent a new communication channel, allowing users to share text, pictures, videos; and to share opinion about the shared content (with voting systems, likes, followers, etc.), participating in any community as follower of followed (Trottier and Fuchs 2015). Within these Social Media, content recommendation systems suggest new content to users with sophisticated algorithms based on previous behaviour or interests and network interests (Tufekci 2015). Taking into consideration the number of users worldwide, the interest in the role in social conflicts that Social Media and other digital technologies has is justified, among many other economical and societal challenges. Currently, there are 2.440 million users in just one of the most important Social Networks: Facebook (Statista, 2020).

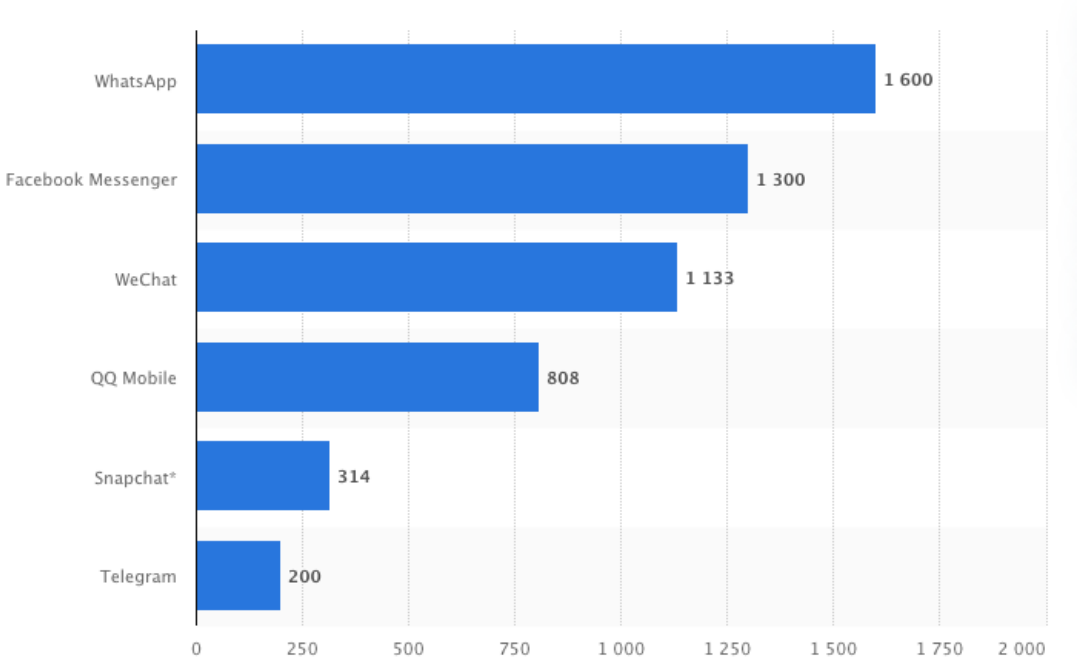
Figure 1. Most popular social networks worldwide as of January 2020, ranked by number of active users (in millions)



Source: Satatista (2020a)

Another digital tool that is gaining importance are mobile messenger apps. Mobile messenger apps allow users to share and spread information in their digital networks very easily. Whatsapp is the leader app. Founded in 2009 by two Yahoo former employees, it was acquired by Facebook in 2014, paying $19 Billion (Vigna, 2014). Nowadays, this mobile messenger app is the preferred app in more than 100 countries around the world (Sevitt, 2017), sharing the growing market with other apps such as Telegram, Line, WeChat or Telegram among others. Close messaging networks create communities with weak and strong ties among their members (Baulch, Matamoros-Fernández & Johns, 2020).

Figure 2. Most popular social networks worldwide as of January 2020, ranked by number of active users (in millions)



Source: Statista (2020b)

There are several factors that provoke the emergence of this new use of digital technologies, such as the reduction of the costs of communication, the increase of the speed of dissemination, the possibility of an easy creation of multiple kinds of digital content (text, images, photo, video) or the possibility of sharing of information easily (Zeitzoff, 2017).

In this emerging digital environment, battles are not only in the battlefield, but in the digital world through Social Networks or mobile messenger apps. These emerging digital media allow us to broadcast news or reliable information easily (Newman et al., 2019), but the feature also affords the possibility of spreading misinformation and spam (Sharma, 2018). This second feature is what is serving to several digital conflicts to foster “information disorder”, with the dissemination of misinformation, disinformation, and malinformation in closed groups not subjected to any kind of platform moderation (Wardle and Derakshan, 2017). There are several cases in which technology has been used for the creation or development of social conflicts as we summarized in Table 1.

Table 1. Examples of social conflicts with digital dimension

Conflict	Use of technology	Reference
Arab Spring	Twitter and Facebook were used to promote the protest	Steinert-Threlkeld (2017)
ISIS	ISIS-affiliated groups have used Twitter, WhatsApp, and other network apps to promote their group's profile, recruit foreign followers, and plan attacks	Berger and Morgan (2015)
Gaza	Both Israel and Hamas militants use social media in general and tweeter particularly to justify their actions and denigrate the other side	Zeitzoff, Kelly and Lotan (2015)

Source: based on Zeitzoff (2017:12)

Another feature of digital communication technology is the pace at which information is spread. The rate of diffusion increases exponentially, it is possible to broadcast worldwide almost in real time. The concept of viral information arises, for instance, spreading a viral YouTube video to reach 100 million streams takes 5.9 days; and in average, in the first 24 hours the video defines its condition of viral (Stadista, 2020c).

The social media and mobile messages app can be used to democratize information or to create new conflicts that serve either a personal or a group interest. Technologies are increasingly used to manipulate information. For instance, doxing, as the process of searching public and private information of a person or organization in Internet with a malicious intent (McNealy, 2019). Another definition of doxing is *"the intentional public release onto the Internet of personal information about an individual by a third party, often with the intent to humiliate, threaten, intimidate, or punish the identified individual"* (Douglas, 2016:1). Another technique is fake news, as the intentional spread of low-quality news with an unethical, illegal or questionable goal (Shu, Sliva, Wang, Tang, & Liu; 2017). False information that is spread in Facebook, WhatsApp groups, Telegram groups, and many other widely used Social Networks are creating or fostering conflicts in both the digital and real word (Martineau, 2018). And the risk is even bigger when bots, as autonomous programs that interact with systems or users, spreads the information manipulating opinions (Shao, Ciampaglia, Varol, Flammini & Menczer, 2017).

Sometimes, organizations not only use social media to spread and recruit followers but to create as well specific mobile apps to organize conflict management. The Hong Kong protest, known as "the water revolution" is an excellent example of how social movements are using technologies to promote and manage real conflicts. Ting (2020:1) pointed out that *"through novel uses of social media and mobile technology, they acted in concert to confront riot police in wildcat actions. In effect, they exhibit a contemporary type of smart mob, as digitally savvy citizens engage with each other in largely ad hoc and networked forms of pop-up protest"*.

After analysing technologies and their use, we may put forward the following statements:

1. Mobile technology is the most used in social conflict. In most of the cases, social media, mobile messenger apps or mobile apps are used.
2. Technology is used during all the stages of conflict:
    a. Conflict promotion.
    b. Recruitment of conflict followers.
    c. Organization and conflict management.

3. Social media and mobile messenger apps are used in all stages, and specific mobile apps are used when conflict reach an important number of followers to manage and organize protest.

All of them are problems in the new smart society. Any conflict born or fostered in the digital world, can move to a violent one in real life (Gohdes, 2018). Digital technologies represent a double-edge sword: democratizing information or manipulating and confronting citizens and societies.

KEYWORDS: digital conflicts, fake news, bots, doxing.

REFERENCES

Baulch, E., Matamoros-Fernández, A., & Johns, A. (2020). Introduction: Ten years of WhatsApp: The role of chat apps in the formation and mobilization of online publics. *First Monday*, 25(1). Retrieved 1 November 2019, from https://firstmonday.org/ojs/index.php/fm/article/view/10412/8319

Berger, J.M., Morgan J. (2015). "The ISIS Twitter Census: Defining and De-scribing the Population of ISIS Supporters on Twitter." *The Brookings Project on US Relations with the Islamic World,* 3 (20): 1-65.

Douglas, D.M. Doxing: a conceptual analysis. *Ethics and Information Technology* 18, 199–210 (2016). https://doi.org/10.1007/s10676-016-9406-0. Retrieved 1 November 2019, from https://link.springer.com/article/10.1007/s10676-016-9406-0

Gohdes, A. R. (2018). Studying the Internet and Violent conflict. Conflict Management and Peace Science, 35(1), 89–106. https://doi.org/10.1177/0738894217733878

Martineau, P. (2018). What Is a Bot? | WIRED. Retrieved 1 November 2019, from https://www.wired.com/story/the-know-it-alls-what-is-a-bot/

McNealy, J. (2019). What is doxxing, and why is it so scary? Retrieved 1 November 2019, from http://theconversation.com/what-is-doxxing-and-why-is-it-so-scary-95848

Newman N., Fletcher R., Kalogeropoulos A, Levy D., Kleis R. (2018). "Reuters Institute Digital news report 2018," *Reuters Institute for the Study of Journalism*, Retrieved 1 November 2019, from http://media.digitalnewsreport.org/wp-content/uploads/2018/06/digital-news-report-2018.pdf.

Sevitt, 2017. "The most popular messaging apps by country," *SimilarWeb* (27 February), Retrieved April 16, 2019, https://www.similarweb.com/blog/popular-messaging-apps-by-country.

Sharma S. (2018). "WhatsApp's 'forwarded' label to curb fake news has a loophole," *Digit* (13 July), Retrieved April 16, 2019 at https://www.digit.in/news/apps/whatsapps-first-step-towards-curbing-fake-news-is-forwarded- label-but-it-has-a-loophole-42196.html

Shao, C., Ciampaglia, G. L., Varol, O., Flammini, A., & Menczer, F. (2017). The spread of fake news by social bots. *arXiv preprint arXiv*:1707.07592, 96, 104, Retrieved April 16, 2019, from https://www.andyblackassociates.co.uk/wp-content/uploads/2015/06/fakenewsbots.pdf

Stadista (2020c). Kworb.net. (February 18, 2020). Fastest viral videos to reach 100 million YouTube streams as of February 2020 (in days) [Graph]. In Statista. Retrieved April 16, 2020, from https://www-statista-com.sabidi.urv.cat/statistics/220391/fastest-viral-videos-to-reach-100-million-hits/

Statista (2020a). We Are Social, & Hootsuite, & DataReportal. (January 30, 2020). Most popular social networks worldwide as of January 2020, ranked by number of active users (in millions) [Graph]. In *Statista*. Retrieved April 16, 2020, from https://www-statista-com.sabidi.urv.cat/statistics/272014/global-social-networks-ranked-by-number-of-users/

Statista (2020b). We Are Social, & Hootsuite, & DataReportal. (October 25, 2019). Most popular global mobile messenger apps as of October 2019, based on number of monthly active users (in millions) [Graph]. In Statista. Retrieved April 16, 2020, from https://www-statista-com.sabidi.urv.cat/statistics/258749/most-popular-global-mobile-messenger-apps/

Steinert-Threlkeld, Z. C. (2017). Spontaneous collective action: Peripheral mobilization during the Arab Spring. *American Political Science Review*, 111(2), 379-403.

Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake news detection on social media: A data mining perspective. *ACM SIGKDD Explorations Newsletter*, 19(1), 22-36.

Trottier, Daniel, and Christian Fuchs. 2015. "Theorising Social Media, Politics and the State." In *Social Media, Politics and the State. Protests, Revolutions, Riots, Crime and Policing in the Age of Facebook, Twitter and YouTube*, Chap. 11, edited by Daniel Trottier and Christian Fuchs, 3-38. London, UK: Routledge.

Tufekci, Zeynep. 2015. "Algorithms in Our Midst: Information, Power and Choice When Software Is Everywhere." In Cosley Dan, Forte Andrea, Ciolfi Luigina, and McDonald David (eds.). *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, 14–18 March, 2015, 1918-18, Canada: Vancouver, BC: ACM.

Vigna, P. (2014). By the Numbers: Facebook Buys WhatsApp for $19 Billion. *Wall Street Journal*. Retrieved April 16, 2020, https://blogs.wsj.com/moneybeat/2014/02/19/by-the-numbers-facebook-buys-whatsapp-for-19-billion/

Wardle C., Derakhshan H. (2017). "Information disorder: Toward an interdisciplinary framework for research and policy making," *Council of Europe*. Retrieved April 16, 2020, from https://edoc.coe.int/en/media/7495-information-disorder-toward-an-interdisciplinary-framework-for-research-and-policy-making.html.

Ting, T. Y. (2020). From 'be water'to 'be fire': nascent smart mob and networked protests in Hong Kong. *Social Movement Studies*, 1-7.

Zeitzoff T., Kelly J., Gilad L. (2015). "Using Social Media to Measure Foreign Policy Dynamics: An Empirical Analysis of the Iranian–Israeli Confrontation (2012–13)." *Journal of Peace Research*, 52 (3): 368-83.

Zeitzoff, T. (2017). How Social Media Is Changing Conflict. Journal of Conflict Resolution, 61(9), 1970–1991. https://doi.org/10.1177/0022002717721392

EMPLOYEE TECHNOLOGY ACCEPTANCE OF INDUSTRY 4.0

Jan Strohschein, Ana María Lara-Palma, Heide Faeskorn-Woyke

Technische Hochschule Köln (Germany), Universidad De Burgos (Spain), Technische Hochschule Köln (Germany)

jan.strohschein@th-koeln.de; amlara@ubu.es; heide.faeskorn-woyke@th-koeln.de

EXTENDED ABSTRACT

Industry 4.0 (I4.0) transforms manufacturing by the integration of the digital into the physical world. In the future smart factories collect more data than ever to empower artificial intelligence (AI) in cyber-physical production systems (CPPS). Creating extensive networks of machines, plants, and companies changes the collaboration with our business partners and machines. However, in 2018, just 14% of 1.600 executives, who participated in a study conducted by Deloitte, believed that their organization is prepared for I4.0 and able to profit from this new potential (Deloitte, 2018). While Gartner predicts that additional automation and the use of artificial intelligence will create more jobs than it destroys, the new jobs will mainly be in fields such as healthcare and education. At the same time, manufacturing will probably see most job losses, so employees are sceptical about the introduction of the new technology (Gartner, 2017).

This work examines the technology acceptance by employees in such companies. Several studies and questionnaires investigate the introduction of I4.0 and AI in manufacturing companies and provide the basis for our extended questionnaire that analyses the employee's feelings in more detail. Abel, Hirsch-Kreinsen and Steglich explain the worker's doubts not only with their fear of job losses, but also the technological changes being digital and no longer immediately comprehensible to the individuals which results in insecurities and scepticism (Abel, Hirsch-Kreinsen & Steglich, 2019). Kagermann, Wahlster, and Helbig similarly report a "growing tension between the virtual world and the world of workers' own experience. This tension could result in workers experiencing a loss of control and a sense of alienation from their work as a result of the progressive dematerialization and virtualization of business and work processes" (Kagermann, Wahlster, & Helbig, 2013). They also agree that through extensive human-machine interactions, the work content, and processes, as well as the working environment, will be radically transformed and thus also the worker's job and competence profiles. Other researchers perceive the introduction of I4.0 not only as a challenge but also as a chance to improve the work environment by creating learning systems, which "dynamically detect and adapt to the context of the support situation and the worker's actions" (Gorecky, Schmitt, Loskyll, & Zühlke, 2014). In conclusion, for a successful introduction of I4.0 in any company, the employee acceptance got identified as one of the most critical aspects. The introduction process requires communication and transparency as "acceptance is a fragile construct, which needs constant cultivation" to convert employee resistance into acceptance or even support (Abel, Hirsch-Kreinsen, & Steglich, 2019).

Companies can use tools for self-assessment of their I4.0 capabilities and ambitions, such as the "Industrie 4.0 Readiness Model" (Lichtblau et al., 2015). While "the model is scientifically well-grounded and its structure and results explained in transparent manners" (Schumacher, Erol, & Sihn, 2016), just a single question targets the employee dimension. In this question, they assess if the workers have the required skills to accomplish their future tasks (Impuls Stiftung, 2015).

Thus for this paper, the "Technology Acceptance Model" (TAM), developed by Davis in 1989, is used to extend the employee dimension (Davis, 1989). Davis models which factors influence if users will

adopt new technology. The two main variables are “perceived usefulness” and “perceived ease-of-use”. Im, Kim, and Han extended the TAM by introducing “perceived risk” as an additional variable that negatively affects adoption (Im, Kim, & Han, 2007). TAM is one of the most popular models to assess user acceptance of new technologies and was also successfully used to evaluate the adoption of related technologies, e.g., smartphones and wearables (Chang, Lee, Ji, 2016; Roy, 2017).

The resulting questionnaire is conducted once again, this time with a focus on smaller and medium-sized companies, which have been underrepresented in the original study (Lichtblau, et al., 2015). Those manufacturing companies are all members of the “Innovation Hub Oberbergischer Kreis”, a regional association that focuses on the exchange of I4.0 knowledge and possible applications. They are excellent candidates for this study, as most of the members just begin to introduce I4.0 and AI into their companies. This study aims to collect additional insights into the I4.0 introduction process for the common worker and the implementation of human-machine interactions. The results will help to design those interactions and using an approach such as Value-Sensitive Design (VSD) in a next step will aid the decision, which human values need to be considered, as laid out by Friedman and Cummings (Friedman, 1997; Cummings, 2006).

While the usability of the new systems will be a critical quality criterion, implementing this interaction through the VSD approach will further support the introduction of I4.0 and the technology acceptance by concerned workers.

KEYWORDS: Industry 4.0, Artificial Intelligence, Technology Acceptance, Value Sensitive Design.

REFERENCES

Abel, J., Hirsch-Kreinsen, H., & Steglich, S. (2019). Akzeptanz von Industrie 4.0

Chang, H. S., Lee, S. C., Ji, Y. G. (2016). Wearable device adoption model with TAM and TTF. International Journal of Mobile Communications 14(5) 518-537

Cummings, M. (2006). Integrating Ethics in Design through the Value-Sensitive Design Approach. Science and Engineering Ethics 12, 701-715

Davis, F. (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. MIS Quarterly, págs. 319-340.

Deloitte. (2018). The fourth industrial revolution is here: Are you ready?

Friedman, B. (Ed.) (1997). Human Values and the Design of Computer Technology. Cambridge University Press, New York, NY

Gartner. (2017). Gartner. Retrieved from https://www.gartner.com/en/newsroom/press-releases/2017-12-13-gartner-says-by-2020-artificial-intelligence-will-create-more-jobs-than-it-eliminates

Gorecky, D., Schmitt, M., Loskyll, M., & Zühlke, D. (2014). Human-machine-interaction in the industry 4.0 era. IEEE International Conference on Industrial Informatics (INDIN).

Im, I., Kim, Y., Han, H. J. (2007). The effects of perceived risk and technology type on users’ acceptance of technologies. Information and Management 45(1) 1-9

Impuls Stiftung. (2015). Retrieved from Industry 4.0 Readiness Online Self-Check for Businesses: https://www.industrie40-readiness.de

Kagermann, H., Wahlster, W., & Helbig, J. (2013). Recommendations for implementing the strategic initiative INDUSTRIE 4.0.

Kasperkevic, J. (2016). TheGuardian. Retrieved from https://www.theguardian.com/us-news/2016/may/25/former-mcdonalds-ceo-threatens-replace-employees-robots

Lichtblau, K., Stich, V., Bertenrath, R., Blum, M., Bleider, M., Millack, A.,… Schröter, M. (2015). Industrie 4.0 Readiness. Impuls Stiftung.

Roy, S. (2017) App Adoption and Switching Behavior: Applying the Extended Tam in Smartphone App Usage. Journal of Information Systems and Technology Management 14(2) 239-261

Schumacher, A., Erol, S., & Sihn, W. (2016). A Maturity Model for Assessing Industry 4.0 Readiness and Maturity of Manufacturing Enterprises. Procedia CIRP, págs. 161-166.

ETHICAL CONCERNS OF MEGA-CONSTELLATIONS FOR BROADBAND COMMUNICATION

Marco Crepaldi

University of Luxembourg (Luxembourg)

marco.crepaldi@protonmail.com

EXTENDED ABSTRACT

In this work, I study the ethical concerns of private satellite mega-constellations in low-earth-orbit (LEO) deployed to broadband connectivity services globally. Ethical concerns related to the increasing role of ICT technologies are not earth-bound but are also found in the last frontier of outer space (Arnould, 2011). Only recently, the space capabilities of private enterprises have made them relevant from an ethics perspective. At the time of writing, several mega-constellations have been announced; this is likely due to the decrease in launch costs and increased relevance and capabilities of small satellites (Millan et al., 2019). For example, SpaceX recently asked the International Telecommunication Union (as well as the U.S. Federal Communications Commission) to allocate part of the radio spectrum for an additional 30.000 Starlink satellites, bringing the number of projected units to around 42.000. While private engagement in space exploration and exploitation is desirable, it raises several challenges that ought to be addressed to ensure that development unfolds in a morally desirable fashion (Marboe, 2016). In this work, I frame the issue at hand in terms of distributed morality to explore its ethical implications (Floridi, 2013; Floridi & Sanders, 2004). The privatization of space might result in global moral actions deprived of individual responsibilities, hence the framing. Distributed morality constitutes an appropriate framework for this issue because it appears to be a moral scenario that is "the result of otherwise morally-neutral or at least morally-negligible interactions among agents constituting a multiagent system" (Floridi, 2013, p. 729). Moreover, the exploitation of LEO by commercial enterprises echoes a tragedy of the commons scenario in that legitimate interactions may render the highly desirable LEO inaccessible for decades to come.

Against this backdrop, the specific challenges of private mega-constellations that are relevant from the ethical perspective are the following. The first set concerns the proliferation of space debris correlated with the increasing number of objects orbiting the already crowded LEO. Avoidance of orbit conjunctions through automated software systems appears desirable and necessary from an ethical perspective; the reason is multifold. On the one hand, collision events in LEO increase the probability of other collisions for all the other objects located in the same orbit, as it is has been demonstrated by the Cosmos 2251 and Iridium 33 conjunction (Ram S. Jakhu, 2010; Wang, 2010). On the other hand, the lack of mandatory rules for the disposal of space debris hinders the future ability of developing nations to access outer space. So that, a morally neutral action – as the launch of a mega-constellation – might result in dire consequences in the context of the multi-agent-system (MAS) of outer space. One possible consequence is to jeopardize access to outer space for future generations (Bergamini, Jacobone, Morea, & Sciortino, 2018; Doldirina et al., 2011).

The second set of challenges concerns accessibility to the limited frequency and orbital planes for broadband communications. The risk is that, if proper measures are not implemented in the short term, a few enterprises of the western hemisphere will end up monopolizing the market for space-based internet access. Space-based internet access is particularly useful in remote areas so that it is relevant from a development perspective. On this basis, the unchecked privatization of space raises

crucial concerns regarding an equitable arrangement for the exploitation of LEO. Unfortunately, the international legal framework for the use of outer space does not provide enough guarantees to ease these ethical concerns (Ram S Jakhu & Pelton, 2017). The legal framework for space activities is made up of four international treaties, the last one signed in 1979 (United Nations, 2017). These sources of law were developed in a time when the privatization of space was not a concern. They are unfit to deal with private satellite constellations for reasons that will be discussed at length in the paper. Ethical policies of aggregation of good actions, as the spontaneous adherence to the mitigations guidelines for the disposal or orbital debris by private companies, and fragmentation of negative ones, as in the case in which an enterprise launches satellites in LEO without registration, should be promoted in this area.

The following directions are suggested to address the aforementioned challenges from the ethical perspective. First, a framework for responsible research and innovation should be devised to account for the emerging challenges of the privatization of LEO (Von Schomberg, 2013). The impact on the environment of outer space of mega-constellations should be evaluated jointly by companies, states, and society at large. More precisely, the precautionary principles ought to be considered before deploying tens of thousands of satellites in LEO. Second, international efforts should be directed toward ensuring that the allocation of frequencies by the ITU does not result in precluding access indefinitely to developing nations, as it stands, the current allocation mechanism (first come, first served) is not enough to address moral concerns related to development. This might be achieved by reserving some areas of the radio spectrum for future use on a national basis. Third, standards and development practices should be considered to devise autonomous software systems to avoid conjunctions in LEO along with a traffic management system for space objects. Some of the previous solutions have been discussed in other contexts, but their moral relevance has been understated thus far (Nair, 2019).

The Outer Space Treaty of 1967 indicates the way forward, "[t]he exploration and use of outer space, including the Moon and other celestial bodies, shall be carried out for the benefit and in the interests of all countries, irrespective of their degree of economic or scientific development, and shall be the province of mankind". The principles enshrined in art. I must find application with respect to private initiatives in outer space. However, recent efforts by major space-faring nations such as the U.S. suggest the hard law is unlikely to uphold these principles (Trump, 2018). Therefore, ethical arguments should be directed to raise awareness of the issues considered in this contribution to avoid harmful consequences. In this paper, I provide three possible directions to explore in order to ensure that the use of the orbital planes of LEO for the delivery of broadband connectivity unfolds in a morally desirable fashion. The subject of this work raises the questions of how to construct the proper ethical infrastructure or infraethics for the exploitation of outer space (Floridi, 2017).

Table 1 Planned Mega-constellations

Constellation	**Number of Satellites**	**Orbit**
Boeing	1.396-2.956	1.200 km
LeoSat	78-108	1.400 km
OneWeb	882-1980	1.200 km
Starlink	4.425-42.943	550-1.325 km
Telesat LEO	117-512	1.000-1.248 km
CASIC Hongyun	156	160-2.000 km
CASC Hongyan	320	1.100 km

Data collected by the author.

KEYWORDS: distributed morality; infraethics; mega-constellations; LEO; space debris; outer space treaty.

REFERENCES

Arnould, J. (2011). *Icarus' second chance: the basis and perspectives of space ethics* (Vol. 6): Springer Science & Business Media.

Bergamini, E., Jacobone, F., Morea, D., & Sciortino, G. P. (2018). *The Increasing Risk of Space Debris Impact on Earth: Case Studies, Potential Damages, International Liability Framework and Management Systems*. In Enhancing CBRNE Safety & Security: Proceedings of the SICC 2017 Conference (pp. 271-280).

Doldirina, C., Howard, D., Hurtz, A., Mey, J., Mineiro, M., Mowle, A. ... Weeden, B. (2011). Towards Long-term Sustainability of Space Activities: Overcoming the Challenges of Space Debris. Paper presented at the A Report of the International Interdisciplinary Congress on Space Debris.

Floridi, L. (2013). *Distributed morality in an information society*. Science and engineering ethics, 19(3), 727-743.

Floridi, L. (2017). *Infraethics–on the Conditions of Possibility of Morality*. Philosophy & Technology, 30(4), 391-394. doi:10.1007/s13347-017-0291-1

Floridi, L., & Sanders, J. W. (2004). *On the morality of artificial agents*. Minds and machines, 14(3), 349-379.

Jakhu, R. S. (2010). *Iridium-Cosmos collision and its implications for space operations*. In K.-U. Schrogl, W. Rathgeber, B. Baranes, & C. Venet (Eds.), Yearbook on Space Policy 2008/2009: Setting New Trends (pp. 254-275). Vienna: Springer Vienna.

Jakhu, R. S., & Pelton, J. N. (2017). *Global Space Governance: an international study*. Springer.

Marboe, I. (2016). *Small Is Beautiful? Legal Challenges of Small Satellites*. In P. M. Sterns & L. I. Tennen (Eds.), Private Law, Public Law, Metalaw and Public Policy in Space: A Liber Amicorum in Honor of Ernst Fasan (pp. 1-16). Cham: Springer International Publishing.

Millan, R. M., von Steiger, R., Ariel, M., Bartalev, S., Borgeaud, M., Campagnola, S., … Gregorio, A. (2019). Small satellites for space science. Advances in space research.

Nair, K. K. (2019). *Small Satellites and Sustainable Development: Solutions in International Space Law.* Springer.

Trump, D. J. (2018). Space Policy Directive-3, National Space Traffic Management Policy.

United Nations (2017). International Space Law Instruments.

Von Schomberg, R. (2013). *A vision of responsible research and innovation*. Responsible innovation: Managing the responsible emergence of science and innovation in society, 51-74.

Wang, T. (2010). *Analysis of Debris from the Collision of the Cosmos 2251 and the Iridium 33 Satellites*. Science & Global Security, 18(2), 87-118.

ETHICAL CONSIDERATIONS OF ARTIFICIAL INTELLIGENCE AND ROBOTICS IN HEALTHCARE: LAW AS A NEEDED FACILITATOR TO ACCESS AND DELIVERY

Mayra Leon Sánchez

Interamerican University of Puerto Rico School of Law (Puerto Rico)

Mml5053@gmail.com

EXTENDED ABSTRACT

Artificial intelligence and robotics have the potential to change healthcare in an unprecedented way. To date, they already have. But just because AI and robotics can make technology better, are they obligated to? More specifically, does the field of medical robotics have any ethical obligation to improve two of the biggest problems in global healthcare: access and delivery? These questions merit careful consideration and cannot be answered in one single paper. This paper examines current advances in introducing AI and robotics to traditional healthcare, while punctuating the imperative that ethics be a present consideration in all these developments. Considering that healthcare is a business with a higher moral imperative than other goods and services businesses, the impact that robotics and AI will have in healthcare lies in how it structures its currently developing ethics. Ethical regulation in traditional healthcare is substantially developed, however in medical robotics it is an emerging and evolving endeavour. AI and robotics companies setting their sight on healthcare, means that robots and AI will be a fixed presence in hospitals and clinics; performing tasks such as scanning vitals or assessing patients.

With surgery robots costing well into half a million dollars, with an added hundreds of thousands in maintenance per year, it is unavailable tech for many health systems and facilities. (Tomlinson, 2018) Introducing advanced technology at a high cost of operation, might mean that those that currently have economic barriers to treatment, will continue to do so. There are substantial risks to introducing new technology into healthcare delivery without a proper ethical frame. From exorbitant costs, to research without proper consent, the pitfalls are varied, new and controversial. (Chen, 2017) However, as this paper will posit, law and its ensuing policies are the perfect mechanism to act as facilitator to ensure ethical access and delivery in the field of medical robotics.

Bill Gates has compared the evolution of robotics to that of the computer, which started out with limited use and access but is now a staple in global everyday life. It then posits: if the evolution of robotics and computers are analogous, then we can expect the proliferation of robotics will bring with it important social and ethical challenges that will require our attention to mitigate its negative effects. (Lin, 2011)

Though there have been some loose efforts into ethical issues surrounding medical robotics, it has not been comprehensive or unified. All the while, there is research happening with lack of a robust ethical framework that other fields, such as traditional medicine or law, both enjoy and operate on. Ethical governance is needed in order to develop standards that allow us to transparently and robustly assure the safety of autonomous systems and hence build public trust and confidence.

Traditional health, technology, and business ethics, by themselves, can fall short when it comes to ethical development of AI and robotics, resulting in gaps of needed regulation. This regulation maintained by government, examining boards, funding sources, and observed by companies, consumers and interested parties is imperative for organized progress in any society. The field of law

is dynamic. It is in a state of constant flux, thus allowing it to deal with the tensions created by the evolution of societies. It is in Law that we put our trust and belief for a just society where human interactions are civilized, within a set of rules and expected behaviours. (Lashbrooke, 1988)

Ethical governance over autonomous systems is needed because, inevitably, near future systems like driverless cars, or medical diagnosis AIs, will be moral agents that will need to make choices with ethical consequences. If ethical theories are to be useful in practice, they need to affect the way human beings behave. (Cath, 2018) The legal field is one of swift and certain choices, that does not allow hesitation in behaviour when the regulations are clearly stated. Where scientists, robotics developers or companies might be ambiguous as to what "the right thing" is, law and policy can provide such answers. Law, in all its aspects, is the architect that structures all our interactions as human beings, from the moment we are born, to the time beyond our death, it is present as a regulating force. It dictates how countries should treat each other in war and peace, how politicians can carry out life changing policies and programs and even what are the responsibilities of parents to their own children.

There are clearly outstanding questions regarding what good AI governance should look like. These questions are currently debated by political institutions across the globe, including the UK, South Korean, the Indian and the Mexican government, as well as the European Commission. AI development is a priority agenda item for more numerous countries. (Cath, 2018) France for example, has pledged up to 1.8 billion in government funding for AI robotic development. (Rosemain, 2018) At the same time, robotics and AI companies themselves are developing their own AI principles and getting involved in developing regulation for AI, whether through direct participation or lobbying efforts. This is especially troubling since many of the industry leaders in the field of AI are incorporated in the USA and China, leaving most industry-led regulation efforts in a concentrated handful of companies that might not be thinking of ethical healthcare robotics within a global context. (Cath, 2018)

As governments move to fund robotics development, there is an opportunity to leverage their authority through policy making and law regulation; hence, ensuring that healthcare involved robotics companies put forward ethically competent technologies. These technologies have the potential for the little improvements, such as record keeping and vital signs assessment, as well as the radical improvements, such as disease eradication in impoverished communities and life changing surgeries in overpopulated areas.

The field of law is perfectly poised to remain critically involved in the collaborative developments of AI and robotics governance solutions. Such as with space laws, conflict laws and commerce law, robotics law must develop to consider international cultural global impact and to include world participation in private sector-led norm development. In an increasingly technological world, the health of one is the health of all, and clusters of disjointed ethical governance over medical robotics and AI pose more risk of doing harm than of doing the greatest good.

KEYWORDS: robotics, healthcare, law, ethics, artificial intelligence.

REFERENCES

Business Dictionary. (2019, October 24). *ethics.* Retrieved from Businessdictionary.com: http://www.businessdictionary.com/definition/ethics.html

Cambridge Dictionary. (2019, October 24). *Definition of ethic* . Retrieved from Cambridge Advanced Learner's Dictionary & Thesaurus. : https://dictionary.cambridge.org/dictionary/english/ethic

Cath, C. (2018). Governing artificial intelligence: ethical, legal and technical opportunities and challenges, . *376 PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A: MATHEMATICAL, PHYSICAL AND ENGINEERING SCIENCES* , 20180080.

Chen, S. (2017, September 10). *AI Research Is in Desperate Need of an Ethical Watchdog*. Retrieved from Wired : https://www.wired.com/story/ai-research-is-in-desperate-need-of-an-ethical-watchdog/

Kowalski, A. (1993). Artificial Intelligence and Law: A Primer an Overview. *51 Advocate(Vancouver)*, 579-584.

Lashbrooke, E. C. (1988). Legal Reasoning and Artificial Intelligence. *34 Loyola Law Review* , 287-310.

Lin, P. e. (2011). Robot ethics: Mapping the issues for a mechanized world. *175 ARTIFICIAL INTELLIGENCE*, 942–949.

Rosemain, M. (2018, March 29). *France to spend 1.8 billion on AI to compete with U.S., China.* Retrieved from Reuters: https://www.reuters.com/article/us-france-tech/france-to-spend-1-8-billion-on-ai-tocompete-with-u-s-china-idUSKBN1H51XP

Senapati, S. (2005). Telemedicine and robotics: Paving the way to the globalization of surgery. *91 INTERNATIONAL JOURNAL OF GYNECOLOGY & OBSTETRICS*, 210–216.

Tomlinson, Z. (2018, November 16). *15 Medical Robots that are changing the World*. Retrieved from Interesting Engineering : 15 Medical Robots That Are Changing the World. Retrieved from https://interestingengineering.com/15-medical-robots-that-are-changing-the-world

ETHICAL ISSUES RELATED TO THE DISTRIBUTION OF PERSONAL DATA: CASE OF AN INFORMATION BANK IN JAPAN

Hiroki Idota

Kindai Univeristy (Japan)

idota@kindai.ac.jp

EXTENDED ABSTRACT

The purpose of this study is to clarify the ethical issues related to the distribution of personal data in Japan from the case of an information bank. With the spread of the Internet, platforms such as Google, Facebook, and Amazon collect personal data and analyze the purchasing behavior and behavioral characteristics of users. They then exploit this data through business activities to generate huge profits. Individuals are able to use the services provided by these organizations near free of charge; however, they provide their personal data to such firms. In such free services, the individuals' data becomes essentially a product, and the organization that owns their data uses that data for advertising, in order to obtain a profit. Most individuals appear unaware of this transaction. Even if they understand its existence, it is difficult for an individual to know where and how their data is used and shared, and they gain little direct benefit from providing them.

On the other hand, utilization of ICT (Information and Communication Technology) and data can be seen to be indispensable for Japan, which has both a rapidly declining birthrate and an aging population, to maintain its international competitiveness and further develop society and industry. Owing to this, it is necessary to collect the data possessed by government agencies, firms, and individuals and analyze them with AI (Artificial Intelligence) and so on to produce new innovations. However, the distribution of personal data is less active in Japan than in the United States or China, and most Japanese people tend to be comparatively more cautious about its commercial use.

For example, JR East, a Japanese railroad company, decided to sell and analyze the usage data gathered from the transportation IC (Integrated Circuit) card "Suica" in 2013. Despite explaining that it would make individuals unidentifiable, it was heavily criticized by consumers. The company stopped selling the data after only a month (Nihon Keizai Shimbun, Inc., 2013/12/19).

Given the situation, the Japanese government revised and enforced the Personal Information Protection Law in May 2017. As a result of these amendments, almost all firms are now subject to regulation. The law regulates the operation of "opt-outs", which allow third-party provision to be considered to be agreed unless the data provider disagrees. Penal regulations have also been strengthened. Thus, personal data protection is strengthened. Further to this, the law has been developed so that personal data can be used actively, provided that data are anonymized.

The General Data Protection Regulation (GDPR) came into force in May 2018 in the EU. In addition to strengthening the protection of personal information, a "data portability right" is prescribed to enable the retrieval of personal data from platforms such as Google, Facebook, and Amazon and so on, and to provide it to other organizations freely. In order to promote the use of personal data, the EU established the right to require companies to provide personal data in a machine-readable format, so that they can be used by other services. In this way, the idea that individuals manage their own data has become mainstream in the EU.

This notwithstanding, it is often difficult for individuals to identify which firm should be provided their data. Therefore, the 'Information Bank' was institutionalized as a policy unique to Japan in 2018.

According to the IT Strategic Headquarters (2017), "Information Bank (Information Use Credit Bank) manages personal data using systems such as PDS (Personal Data Store) based on contracts for data use with individuals. In addition, it is a business that will provide data to a third party (another business operator) only after judging the validity on behalf of the individual, based on individual instructions or pre-specified conditions."

In other words, the Information Bank is a platform that supplies data deposited by the individual only to the extent that the individual agrees, then returns the benefits obtained from any data provided back to that individual. By using an information bank, there is no need to spend time and effort to find a firm which is fit to be supplied an individual's data; it is possible to delegate data management to a reliable third party. This said, the information bank system has only recently begun and has not yet been commercialized.

In this paper, we consider a number of ethical issues with respect to the distribution of personal data in this system. These ethical issues are explained from five viewpoints: fairness, transparency, accountability, trust, and security.

1) Fairness

   Information banks differ from existing platforms. Individuals have the right to choose whether or not to use them. They can decide which firm to deposit their data from multiple information banks. They can also determine the scope of data to be deposited, which organizations may use their data, and the purpose for their use.

2) Transparency

   It is unclear how existing platforms use collected data. On the other hand, guarantees regarding the conditions under which data are managed, transparency with respect to the organization that provided data, and transparency in the use of data by that organization are ensured by information banks. An example of this might be a mechanism such as a "bank book", which makes it possible to visualize which data were used, both when and by whom.

3) Accountability

   An information bank is responsible for the strict management and supply of personal data. It should represent a contact point for customer complaints and bear primarily accountability. It should also be liable to an individual for damages, even if the organization that provided data is responsible. This is mandatory for the information bank.

4) Trust

   Unlike private firms such as existing online platforms, information banks can only do business after being reviewed and certified by a third-party, such the Japan IT Federation. An information bank must be a company with high creditworthiness and extremely low risk of bankruptcy. Therefore, financial institutions and major IT companies are assumed to play this role. Further, the information bank will undergo an accreditation renewal review every two years: if sufficient criteria are met, accreditation will continue. In this way, the trust of information banks is ensured through certification by a specialized third-party organization.

5) Security

More advanced security technology is required to prevent the misuse of personal data. According to the above-mentioned accreditation, information banks need to conform to the "Guidelines for Accreditation of Information Trust Organizations" formulated by the Ministry of Internal Affairs and Communications and the Ministry of Economy, Trade and Industry (2018), along with the accreditation standards for information security and privacy protection measures formulated by the IT Federation. Therefore, information banks work strictly to prevent leakage of personal data and unauthorized use.

In this way, it can be said that the information bank is required to consider the ethical issues regarding the use of personal data more comprehensively when compared to existing platforms.

Information banks are just getting started. Various ethical issues will become clear in the future. The biggest problem is likely to be whether the number of users who provide personal data increases, or not. As can be seen in the case of JR East above, Japanese people tend to be cautious about the commercial use of their personal data. Whether or not to provide their personal data should be left to the individual's judgment, although it is thought that an innovation produced by accumulating and analyzing the data returns the benefits to the whole society (Mason et al., 1995). In this study, we would like to consider this problem more deeply, and also clarify the conditions for promoting personal data to be distributed and used appropriately.

KEYWORDS: Information bank, Personal data, Data distribution, Ethical issues.

REFERENCES

IT Strategic Headquarters, Data Distribution Environment Improvement Study Group (2017). Data Utilization Working Group in the AI and IoT Era, Interim Report. *Kantei* (in Japanese). Retrieved from http://www.kantei.go.jp/jp/singi/it2/senmon_bunka/data_ryutsuseibi/dai2/siryou2.pdf

Mason, R. O., Mason, F. M., & Culnan M. J. (1995). *Ethics of Information Management*. Thousand Oaks, Calif. : SAGE.

Nihon Keizai Shimbun, Inc. (2013, December, 19). Lessons from the failure to sell getting on and off the train history of 'Suica', 6 points to use personal data. *Nihon Keizai Shimbun* (in Japanese). Retrieved from https://www.nikkei.com/article/DGXNASFK1102K_R11C13A2000000/

Ministry of Internal Affairs and Communications and the Ministry of Economy, Trade and Industry (2018). Guidelines for Accreditation of Information Trust Organizations. M*inistry of Internal Affairs and Communications and the Ministry of Economy, Trade and Industry* (in Japanese). Retrieved from https://www.meti.go.jp/press/2018/06/20180626002/20180626002-2.pdf

ETHICS-BY-DESIGN FOR INTERNATIONAL NEUROSCIENCE RESEARCH INFRASTRUCTURE

Damian Eke, Simisola Akintoye, William Knight, George Ogoh, Bernd Stahl, Inga Ulnicane

De Montfort University, Leicester (United Kingdom)

damian.eke@dmu.ac.uk; simi.akintoye@dmu.ac.uk; william.knight@dmu.ac.uk; george.ogoh@dmu.ac.uk; bstahl@dmu.ac.uk; inga.ulnicane@dmu.ac.uk

EXTENDED ABSTRACT

The recent EU's legislative emphasis on Data protection by design -DPbD has given further traction to the assumptions that privacy and data protection are almost the only data related concerns in system development and that human data are the only relevant data in design. Human data are only a part of the data used in systems' design. Animal and technical data receive no attention in the privacy-by-design - PbD nor in the PDbD frameworks. We argue in this paper that this is not appropriate, given the wider data related concerns associated with AI applications such as data subject's right to reasonable inference, bias/diversity, the distribution of costs and benefits or the possibility of data abuse or misuse. These and other concerns, for example concerns related to animal data, are relevant concerns that demand attention in the design of cutting-edge technology for commercial, health and research benefits. In essence, the ethical, legal and social issues design can raise are not confined to data protection and privacy. Embedding privacy or data protection into design is not sufficient to mitigate other social, economic, legal, philosophical and ethical concerns design raises. Responsible systems design requires a robust approach/mechanism that addresses fundamental concerns beyond privacy and data protection. It involves embedding fundamental ethical principles into design so that developed systems, products and infrastructure will be legally compliant, socially acceptable and ethically responsible. Therefore, we argue that the concept of Ethics-by-design (Dignum et al., 2018) as a form of value-sensitive design (Friedman et al., 2006), provides a robust framework for ethical and legal considerations of risks in design by presenting how it has been applied in the development of an international neuroscience research infrastructure. Therefore, this paper answers the question: how can desired values or principles (ethics) be embedded into design, especially the design of an international neuroscience research infrastructure? Even though this concept has been applied in technology (Dodig-Crnkovic and Çürüklü, 2012) and business (Moore, 2017) discourses, it is yet to be applied to the design of a research infrastructure. We demonstrate how this has been applied in the development of EBRAINS -an integrated ICT neuroscience research infrastructure coming out from the EU FET programme, Human Brain project.

According to the Directorate-General for Research and innovation of the European Commission, research infrastructures are ''facilities that provide resources and services for research communities to conduct research and foster innovation''. The consideration of ethical, legal and social principles to be integrated into the design of such infrastructure should start from the onset of the project through to after the system is created. This proactive and preventive measure is motivated not only by the imperative to consider research ethics questions but also by the need to responsibly govern the generated complex, big data and the necessity to embed relevant principles into the design. The latter leads to the establishment of practical and responsible mechanisms/approaches informed by ethics and the law and designed to integrate core ethical principles and legal provisions into the technological infrastructure. Ethical issues under the purview of a neuroscience infrastructure therefore range from

classic biomedical research ethics questions of benevolence and nonmaleficence, respect and autonomy, justice to data governance issues of data protection and privacy (Christen et al., 2016). There are also other issues of brain simulation and moral status, animal welfare, dual use, AI related issues of bias, diversity and wider neuroethics concerns such as origin of consciousness.

The Human Brain Project (HBP) as a neuroscience collaborative project with the objective of developing Information Technology infrastructure demonstrates a convergence of neuroscience and ICT. Data is therefore central to the research, creation and implementation of the research infrastructure (Bernd et al., 2018). Many parts of this process (including; data collection, processing, curation, deletion, sharing and creation of inferential scientific insights) raise considerable ethical concerns. These include issues such as informed consent, anonymization, animal welfare, diversity, transparency and fairness. However, there are broader neuroethics issues which are exacerbated by differences in cultures and regulations (Rommelfanger et al., 2018). The cross-boundary and cross-functional nature of the neuro-data further highlights the ethical and legal complexities involved. There is an underlying perception that a research infrastructure requires core principles beyond data protection and privacy. It requires ethics-by-design.

In this paper, we address the question of how ethical principles can be embedded into the design of technology by critically reflecting on how this has been addressed in the design of a neuroscience collaborative research infrastructure. This paper argues that ethics and not just data protection or privacy should be embedded into the design of technology. It demonstrates how this has been applied in the development of a responsible international research infrastructure using the principles of Responsible Research and Innovation (RRI) (Stahl et al., 2019). It details how ethical, social and legal concerns are unpacked and addressed in the design process through a responsible approach that embraces foundational principles of ethics-by-design. It articulates the development and refinement of approaches to ensure that ethics is an essential component of the core functionality of the infrastructure or is integral to the system without diminishing its scientific functionality. As Simon (2016) observed, in building technologies, values are often unintentionally inscribed in them and in return they may promote or demote certain values. Some of these values include; justice, fairness, privacy, responsible animal welfare, equity and bias. Therefore, instead of unintentionally embedding undesirable values into the design, EbD is a conscious effort to intentionally embed desired values into the technology- in this case, a technological research infrastructure that can promote values desired by the relevant stakeholders who are or can be affected by the technology.

While the underlying principles of EbD sounds great, the practical questions on how to embed these principles into design have remained unclear in literature or practice. In this paper, we present an EbD approach contextualized and operationalized in the development of an international neuroscience research infrastructure. This paper draws on the concept of Responsible data governance (RDG), an inclusive and discursive mechanism that integrates the tenets of RRI within a broader framework of ethical dialogues and ever-changing legislations (Fothergill et al., 2019). This EbD approach was predicated on core principles that include; proactivity, fairness, responsibility, inclusivity, animal welfare, self-determination and non-instrumentalism. These principles were integrated into the design through a pragmatic and reflective RRI approach characterized by four features: anticipation, reflection, engagement and action (AREA framework) proposed by Stilgoe et al, (2013). Through a variety of activities, these principles are integrated into the research and design stage of the infrastructure and include the identification of ethical and social concerns through ethics surveys. These are subsequently reflected on and turned into Foresight models to increase accuracy and enable real-time technology assessment. Further tasks established to ensure the integration of the core principles include and not limited to engagement activities such as Trilateral meetings, Ethics Rapporteur program, Data Governance Working Group meetings, informal advisory boards (Ethics

Advisory Board, Science and infrastructure Board) and other organizational oversight mechanisms. These result in responsive actions - aimed at promotion of privacy, respect for human rights, transparency and animal welfare - such as Data protection impact assessments, Standard operation procedures (SOPs), Data protection officer role, Ethics compliance management role, Dual use opinion, Access Policy and responsible research and use policy.

Finally, the motivation is to contribute to the broader discussions on the challenges of integrating ethics into design by providing an example of this in practice. It is easy for the conversations around EbD to become too abstract to offer practical insights into achieving its objectives. This paper articulates a practical way of thinking about ethics in design with emphasis on building socially responsible systems or artefacts. We start with a historical excursus of DPbD, PbD and EbD which feeds into the core principles of Ethics by design (which includes; proactivity, inclusiveness, social responsibility, non-instrumentalism and continuity), illustrating reflective approaches aimed at effective integration of ethical principles into systems design. It presents a programme of activities adopted by the HBP Ethics and Support team for the design of EBRAINS, that cover the integral processes of anticipation, reflection, engagement and action as well as fundamental RRI principles of diversity, ethics and governance.

KEYWORDS: Ethics, neuroscience, design, research infrastructure, privacy, ICT.

REFERENCES

Stahl, B., Rainey, S., Harris, E., Fothergill, T. (2018). The role of ethics in data governance of large neuro-ICT projects. *J. Am. Med. Inform. Assoc*. 25, 1099–1107.

Dignum, V., Baldoni, M., Baroglio, C., Caon, M., Chatila, R., Dennis, L., Génova, G., Haim, G., Kließ, M.S., Lopez-Sanchez, M. (2018). Ethics by Design: necessity or curse? Presented at the Proceedings of the 2018 *AAAI/ACM Conference on AI, Ethics, and Society, ACM*, pp. 60–66.

Christen, M., Biller-Andorno, N., Bringedal, B., Grimes, K., Savulescu, J., Walter, H. (2016). Ethical Challenges of Simulation-Driven Big Neuroscience. *AJOB Neurosci.* 7, 5–17. Retrieved from https://doi.org/10.1080/21507740.2015.1135831

Dodig-Crnkovic, G., Çürüklü, B. (2012). Robots: ethical by design. *Ethics Inf. Technol.* 14, 61– 71. Retrieved From https://doi.org/10.1007/s10676-011-9278-2

Fothergill, B.T., Knight, W., Stahl, B.C., Ulnicane, I. (2019). Responsible Data Governance of Neuroscience Big Data. *Front. Neuroinformatics* 13, 28. Retrieved from https://doi.org/10.3389/fninf.2019.00028

Friedman, B., Kahn, P., Borning, A. (2006). Value Sensitive Design and Information Systems, in: Zhang, P., Galletta, D. (Eds.), *Human-Computer Interaction in Management Information Systems: Foundations.* M.E Sharpe, Inc, NY.

Moore, S.L. (2017). *Ethics By Design: Strategic Thinking and Planning for Exemplary Performance, Responsible Results, and Societal Accountability*. HRD Press, Amherst, Mass.

Rommelfanger, K.S., Jeong, S.-J., Ema, A., Fukushi, T., Kasai, K., Ramos, K.M., Salles, A., Singh, I., Amadio, J., Bi, G.-Q., Boshears, P.F., Carter, A., Devor, A., Doya, K., Garden, H., Illes, J., Johnson, L.S.M., Jorgenson, L., Jun, B.-O., Lee, I., Michie, P., Miyakawa, T., Nakazawa, E., Sakura, O., Sarkissian, H., Sullivan, L.S., Uh, S., Winickoff, D., Wolpe, P.R., Wu, K.C.-C., Yasamura, A., Zheng, J.C.

(2018). Neuroethics Questions to Guid e Ethical Research in the International Brain Initiatives. *Neuron* 100, 19–36. Retrieved From https://doi.org/10.1016/j.neuron.2018.09.021

Simon, J. (2016). Value-Sensitive Design and Responsible Research and Innovation, in: *The Ethics of Technology - Methods and Approaches*. Rowman & Littlefield Publishers, London, pp. 219–236.

Stahl, B.C., Akintoye, S., Fothergill, B.T., Guerrero, M., Knight, W., Ulnicane, I. (2019). Beyond Research Ethics: Dialogues in Neuro-ICT Research. *Front. Hum. Neurosci.* 13. Retrived from https://doi.org/10.3389/fnhum.2019.00105

Stilgoe, J., Owen, R., and Macnaghten, P. (2013). Developing a framework for responsible innovation. *Res. Policy* 42, 1568–1580. Retrieved from https://doi.org/10.1016/j.respol.2013.05.008

EXAMINATION OF HARD-CODED CENSORSHIP IN OPEN SOURCE MASTODON CLIENTS

Juhani Naskali

University of Turku (Finland)

juhani.naskali@utu.fi

EXTENDED ABSTRACT

This article analyses hard-coded domain blocking in open source software, using the GPL3-licensed Mastodon client Tusky as a case example. Firstly, the question whether such action actually is censorship is analysed. Secondly, the legality of such action is examined. This paper finds domain blocking to be censorship in the literal definition of the word, as well as against the spirit and letter of current GPL licensing --- though some slight ambiguity remains, which calls for clarifications in the licensing terms. A multi-disciplinary ethical examination of domain blocking will also be done in a later publication, as at first glance the use of censorship systems is difficult to argue for from an ethical standpoint. If domain blocking in open source software is unethical, it is imperative that such action is more clearly discouraged in licensing terms and open source application distribution agreements.

Mastodon is a free and open-source social network akin to Twitter, sometimes also called a microblogging service. Each Mastodon instance has its own rules for membership and moderation, and Mastodon includes tools for individual users to block messages from specific users or instances. Gab (https://gab.com) is a microblogging instance based on Mastodon, which claims that it "champions free speech, individual liberty and the free flow of information online", but has been considered to contain extreme hate speech in many instances, to the point of the platform's applications being banned on Apple Store and Google Play Store [Lee, 2017].

On 17th of June 2019, a change was made to the Mastodon social media client Tusky that prevents some users from logging in, and instead redirects them to a famous video of Rick Astley in a common internet gag familiarly coined as "rickrolling". The codechange is relatively simple. It checks if the user's domain is 'gab.com' or 'gab.io' (or a subdomain thereof), and opens a browser view of the specified youtube url based on this check. [Tusky, 2019] The change renders the app unusable with Gab accounts, as they are unable to log in. Removing the block requires changing the code and compiling it yourself, which is outside the expertise of most smartphone users. The topic has generated quite a wide array of discussion and elicited strong emotions in people. There is no clear consensus on whether such domain blocking is in accordance to GPL licensing or whether it is ethical.

The case is highly interesting, as this is one of the first cases of censorship based on the user's chosen platform or instance / service provider. It has also been traditional for client programs to be provider-agnostic, working on any and all providers of the supported protocols, and this type of service-specific blocking is new, especially in open source software. It is analogous to Outlook email program not working for Gmail users. This type of service restriction is not directly considered in current licensing and policy texts. It is possible that the emergence of this new type of restriction on users necessitates some reviews in open source licensing terms and/or developer policies.

While hardcoded domain blocking only prevents users from accessing their service via one specific application, it still suppresses this specific means of expression. This fits the dictionary definition of censorship, even though the action is very different from hard govermental censorship. Thus domain

blocking is categorized as censorship, even though it is a softer kind of censorship from an all-out legal shutdown of the service.

On the question of legality, programmers have copyright to their creations and prima facie can dictate quite freely how their code and programs should be used and by whom. These rights can possibly be subject to preceding rights by others --- rights that create more compelling duties to respect other people's freedoms. The creators can also willfully give away parts of these rights to others with contracts and agreements. In the case of open source, the creators of software code enter into a licensing agreement with others, guaranteeing their right for using and modifying the code freely. Tusky is licensed under GPL-3, so the GPL license will be examined in more detail, though many of the findings might be applicable to other similar licenses.

The definition of Free Software has been discussed since the 80s, with GNU's 1st bulletin being one of the first records of what is considered free software: "When we [the Free Software Foundation] speak of free software, we are referring to freedom, not price." [GNU Project, 1986, p.8] This definition of free software focuses on user freedom, but mainly discusses the freedom to share, read and modify code. Around 1990 the GNU Project added a "freedom 0" to their text, which precedes the freedoms relating to study, redistribution and modifying the code: "The freedom to run the program as you wish, for any purpose". [GNU Project, 2001] and Stallman [2013] later expanded on the reasons why programs must not limit the freedom to run them, explicitly stating that distributions shouldn't restrict how you use the software.

If the ideal of freedom is user choice, it is difficult to see how hard-coded limits would fit this ideal. Still, it is possible to consider the ideal is only to secure user freedom in relation to the code, and not the distributed program(s). It is worth noting, however, that such distinction only secures freedoms for those people who have the necessary skills to program, discriminating against those who do not have the means to remove restrictions for themselves or cannot have others do it or them.

Line 158 of the GPL-3 license states "This License explicitly affirms your unlimited permission to run the unmodified Program." The license goes on to prohibit different ways of limiting the use of software, including use of DRM (digital rights management) that cannot be removed and withholding installation information. This gives further credibility to the interpretation that open source software should be able to run without interference. Later parts of the text only explicitly forbid restrictions that cannot be removed, but the sentence quoted above clearly relates to the unmodified program, making invalid any claims that restrictions are ok as long as they can be removed by modifying the code. The change could also concievably be considered denial of access, which is forbidden on line 333 of GPL-3, but such interpretation would be shaky, at best.

Tusky's domain blocking also seemingly violates some Google Play Developer policies, such as Interruption of service, policies on product takedowns, deceptive behavior and the clause on minimum functionality.

This type of open source domain blocking is not in accordance with GPL-3 licensing. While it might be possible to side-step the legal issues through the use of other licenses that are less vigorous in protecting user liberties, user freedom is one of the core values of free software, and is difficult to reconcile with censorship. A multi-disciplinary ethical examination of the situation is warranted to surmise what ought to be done with hate speech censorship in open source software, as even though hate speech is wrong, it does not necessarily follow that hate speech censorship is good.

KEYWORDS: open source, hate speech, censorship, blocklist, GPL.

REFERENCES

GNU. Gnu's bulletin volume 1 no.1, 1986. https://www.gnu.org/bulletins/bull1.txt [Online; accessed 18. Jul. 2019].

GNU Project. What is free software?, 2001. https://www.gnu.org/philosophy/free-sw.html [Online; accessed 18. Jul. 2019].

Timothy B. Lee. Google explains why it banned the app for gab, a right-wing twitter rival, 2017. https://arstechnica.com/tech-policy/2017/08/gab-the-right-wing-twitter-rival-just-got-its-app-banned-by-google [Online; accessed 30. Oct. 2019].

Richard Stallman. Why programs must not limit the freedom to run them, 2013. https://www.gnu.org/philosophy/programs-must-not-limit-freedom-to-run.html [Online; accessed 18. Jul. 2019]

Tusky. Merge pull request #1303 from mlc/rick_roll_domains · tuskyapp/Tusky@5d04a7c, 2019. https://github.com/tuskyapp/Tusky/commit/5d04a7c [Online, accessed 1. Jul. 2019]

FROM ALGORITHMIC TRANSPARENCY TO ALGORITHMIC ACCOUNTABILITY? PRINCIPLES FOR RESPONSIBLE AI SCRUTINIZED

Paul B. de Laat

University of Groningen (the Netherlands)

p.b.de.laat@cerug.nl

EXTENDED ABSTRACT

Systems employing algorithms that are based on machine learning (ML) are all around us. Used for purposes of classification, diagnosis, prediction, and recommendation, these systems are applied in areas such as journalism, transportation, biomedicine, public health, safety, criminal justice, insurance, banking, taxation, and education. Especially if such systems are critical for safety, or, more broadly, affect people's destinies significantly, accountability to the public is to be put on the agenda: we must be able to hold institutions employing such systems to account for their performance.

Currently, institutions both public and private around the globe are staging discussions and formulating approaches to accountability for such AI systems (I use the terms ML and AI interchangeably). These range from EU-level bodies, standard-setting organisations, companies, trade associations, to professional and academic organisations (for an overview see algo:aware 2018: 37-109; Jobin et al. 2019). In these discussions the term accountability is variously understood as requiring the following properties for AI-systems: they are to be accurate, fair, equitable, transparent, interpretable, privacy-protected, robust, and/or resilient. In an effort toward synthesis, these have been subsumed under the headings of four principles to guide the development of 'trustworthy' AI: respect for human autonomy, prevention of harm, fairness, and explicability (source: high-level expert group on AI, set up by the European Commission, 2019).

What are the chances that these lofty ideals will effectively be realized? In particular, how likely is it that organizations actually developing and/or deploying AI-systems will stick to these principles of their own accord? Will this self-proclaimed form of principled behaviour work out positively? Recently, several obstacles have been identified that might impede its realization: compared to medicine, the young AI-community lacks common aims and values, tools that translate principles into practices, professional norms for good practice, and mechanisms of accountability (Morley et al. 2019, Mittelstadt 2019).

In this abstract, I want to draw attention to another potential obstacle – the attitude and practice by institutions employing AI of keeping any and all details of their systems a secret. Absolute secrecy has been the norm for the last two decades, from the beginnings of the development of ML and AI. Reasons usually adduced are protection of their knowledge assets against competitors, and fears that the system will be gamed. Considerable efforts have been deployed into acquiring legal protection for their intellectual assets. The 1990s saw efforts towards the expansion of copyright to include look-and-feel issues (unsuccessfully). From 2000 onwards the eligibility of software-related inventions for patent protection has been pushed forward (successfully). And more recently, trade secrecy laws in both the US and Europe have been strengthened and homogenized. All along, the articulation of algorithms as specifically worthy of protection as intellectual property has been clarified.

This secretive attitude has, until now, seriously undermined any form of accountability. Trade secrecy law especially has been the instrument par excellence for keeping critical inquiries at bay – in particular

regarding a 'right to explanation' supposedly emanating from the GDPR or its predecessor, the DPD (cf. Wachter 2017, de Laat 2019, Maggiano 2019, Moore 2017). Trade secrecy arguments also played a pivotal role in recent US lawsuits about the violation of due process in algorithmic decision-making (COMPAS and EVAAS cases). The thesis I want to explore in this article concerns the connection between proper accountability and secrecy: if the organisations involved seriously want to realise the lofty ideals for the AI practices they subscribe to and commit to being held to account for them, this attitude of secrecy has to change. To be more precise: positioning themselves as accountable to society requires disclosure of the relevant aspects of their algorithmic processes. Accountability *and* some degree of algorithmic transparency become mandatory.

For purposes of this research I employ the terms transparency and accountability as follows. Full transparency refers to the arrangement in which all details of algorithmic processes come into the public limelight, from beginning to end. Usually, though, transparency assumes a more modest form. In order to conceptualize this, transparency can usefully be differentiated along two dimensions: (1) Which element(s) of the algorithmic process are being disclosed? (2) To whom the disclosure(s) are made?

Accountability, on the other hand, is to be understood as a relationship between parties, in which the actor is obliged to explain and justify his conduct to a forum, the forum can ask questions and pass judgment, and the actor may face consequences ('narrow' accountability: Bovens 1997). The connection between the two concepts is that transparency is a necessary (but not sufficient) condition for accountability. Moreover, transparency alone, say to the public at large, does not establish relations of accountability.

In the sequel, therefore, I focus on disclosures about the algorithmic processes in use (details about data sources, datasets, ML-methods in use, algorithms produced, and final outcomes) to forums of a kind: self-regulatory organisations (like trade associations, standard-setting bodies, professional associations), third parties (like independent supervisory authorities, auditors, ombudsmen – usually referred to as 'co-regulation') or courts of law (laws, statutes). Let me henceforth refer to this as 'algorithmic accountability'.

Note that such accountability not only requires new disclosures to new forums; it also requires the development of *new tools*. For many desiderata on the agenda (fairness, privacy-protection, and explainability in particular), new tools and measures are to be developed. These tools will drastically change the actual practice of ML. Moreover, the measures involved will figure prominently in the accounting reports to third parties. Compare, as an example, the desired feature of explainability: for its effective realization either local approaches have to be developed, or simpler models that are interpretable by design are to be used exclusively (cf. de Laat 2019). Both approaches may succeed in providing explanations for decision subjects.

So taken together, my research question becomes the following: are the organisations promising us that they will realize these lofty ideals for responsible AI aware of the fact that they will have to take steps concerning *algorithmic transparency* to *new forums* and develop *new tools* for the occasion? Can we detect signs of such a disposition? Without any such signs, their promises are just hollow gestures without any credibility. In view of their past emphasis on property and secrecy we, as a society, are entitled to harbour reservations.

My method is as follows. From the many documents and declarations of principles for responsible AI, I select only those drafted by companies or institutions that actually practice or use AI/ML; after all, these institutions actually carry the burden for realizing responsible AI. Trade associations that speak for sectors as a whole and standard-setting organisations are also included. Excluded from my selection, therefore, are organisations of academics, professionals and other concerned publics. Can

we detect signs of the need to move towards new forums to account to, and the development of new tools for responsible AI/ML all along? A special focus lies on the ways in which the terms 'transparency', 'accountability', and 'explainability' figure in their public statements. Although each document turns out to define these terms in different – and often confusing - ways, taken together they are often revealing of where the organisations under scrutiny are heading to.

Very preliminary results – for purposes of this abstract - indicate a disappointing pattern. Most companies (like Microsoft, Google) just emphatically endorse the lofty desiderata for future AI. The declarations almost universally check all the relevant boxes and provide some clarifications. Unfortunately, no forums are suggested to which they should be making themselves accountable. The message is: "We work on these properties, perform risk assessments when needed, have internal boards and ethical committees in place carrying out oversight, and will produce reports to the public - so leave it to us." A clear case of plain transparency without accountability. As Google remarks: "AI should be accountable to people" (https://ai.google/principles) – thereby hollowing out the concept to the bare duty of publishing information. The companies only superficially touch upon trade or professional organizations as a potential forum to account to – although several of these are already busy formulating standards and protocols for AI systems. Moreover, the declarations provide no specific clues about new tools for ML being necessary – we will just have to trust the organisations involved to do a professional job.

There are some remarkable deviations, though, from this pattern of promises-without-assurances. On the one hand we find some trade associations that by and large adopt the same position concerning the ideals for future AI – but cannot refrain from underlining the strict secrecy stance concerning industrial IP. Says the Software & Information Industry Association (2017): "Organizations do not need to disclose source code of proprietary algorithms for several reasons. Disclosure is not useful for accountability purposes, especially in the case of advanced analytical techniques that improve themselves in use. Source code disclosure would likely produce counterproductive efforts to game analytical systems in ways that defeat their purpose. Disclosure would allow anyone to use or benefit from systems that require extensive development resources, thereby weakening the economic incentive in creating these systems."

The Information Technology Industry Council (ITI) (2017), for their part, also subscribe to this proprietary attitude in no uncertain terms. In addition, though, they take pains to emphatically warn against any external or state regulation: "We believe governments should avoid requiring companies to transfer or provide access to technology, source code, algorithms, or encryption keys as conditions for doing business. (..) We also encourage governments to evaluate existing policy tools and use caution before adopting new laws, regulations, or taxes that may inadvertently or unnecessarily impede the responsible development and use of AI. This extends to the foundational nature of protecting source code, proprietary algorithms, and other IP." These proprietary instincts can only add to our scepticism about the realization of responsible AI in the future.

On the other hand, interestingly, some declarations of AI principles provide more tangible assurances, notably from IBM and Intel. IBM follows the usual pattern of noncommittal engagement to principles for AI. In addition, though, they actually seem to *work* on practical tools that bring these principles nearer: tools for reducing bias in data sets and testing black-boxed AI systems for bias (https://www.ibm.com/blogs/policy/bias-in-ai/). Another promising development is that they *open sourced various tools that allow developers to experiment with several current ML tools and generate 'local' explanations for data subjects* (**'**AI Explainability 360 Toolkit', available at https://aix360.mybluemix.net).

As for Intel, besides the usual endorsement of lofty principles for AI, they repeatedly plead for regulators to come into the picture: "Organizations implementing AI solutions should be able to demonstrate to regulators that they have the right processes, policies and resources in place to meet those principles" (Intel 2017). They even suggest a focused form of state regulation: "Governments should determine which AI implementations require algorithm explainability to mitigate discrimination and harm to individuals." According to Intel, therefore, transparency has to be coupled with an accounting forum outside the companies themselves (regulators, courts). As such, this represents the only clear example among all declarations surveyed of a company unambiguously embracing accountability for AI.

Summarizing: companies developing AI systems have subscribed to a series of lofty principles. But in the main, these declarations commit to transparencies to the public at large only; the notion that an account has to be given to some societal forum of a kind is missing. Moreover, the new tools for ML that need to be developed are only rarely specified. No wonder that some critics have speculated that the endorsement of these principles is mainly inspired by a desire – whether consciously or not – to stave off regulation of a kind.

What are the connections with the issue of secrecy? For the majority of companies surveyed, subscribing to transparency without accountability, secrecy-as-is about the algorithmic process can simply continue. This would be otherwise in the cases where third parties are called to step in (Intel being the lonely advocate for this governance option). Proper accounting for fairness, explainability, and the like will definitely require hitherto unheard of disclosures. Will these necessarily extend to source code/algorithms? This is a hotly debated issue. In general, white box testing is considered more powerful than black box testing. Kroll et al. (2016) argue, however, that tools are available to show conformity to 'substantive policy demands' that do *not* require actual source code or algorithms to be disclosed. If at the end of the day white box testing by external experts *is* deemed necessary by third parties, it is to be expected that the old reflex of invoking trade secrecy to protect one's intellectual assets from disclosure will resurface – and effectively complicate such testing. The same reasoning applies whenever specific new rules are to be enshrined in law (regarding fairness, interpretability, and the like). Also then, the decades-long strengthening of intellectual property rights for algorithms (as indicated above) may be expected to potentially undermine accountability through the invocation of trade secrecy.

KEYWORDS: accountability, algorithm, intellectual property, interpretability, machine learning, transparency.

REFERENCES

Algo:aware (2018) Raising awareness on algorithms. Procured by the European Commission's Directorate-General for Communications Networks, Content and Technology. Version 1.0, December 2018. Retrieved from https://www.algoaware.eu.

Bovens, M. (2007) Analysing and Assessing Accountability: A Conceptual Framework. *European Law Journal,* 13(4), 447–468.

Jobin, A., Ienca, M. & Vayena, E. (2019) The global landscape of AI ethics guidelines. *Nature Machine Intelligence.* 1, 389–399.

Kroll, J.A., Huey, J., Barocas, S., Felten, E.W., Reidenberg, J.R., Robinson, D.G., & Yu, H. (2016) Accountable Algorithms. *University of Pennsylvania Law Review,* 165, 633-705.

Laat, P.B. de (2019). Algorithmic decision-making employing profiling/scoring: Will the right to explanation be crushed by intellectual property rights? *Submitted for publication.*

Maggiolino, M. (2019) EU Trade Secrets Law and Algorithmic Transparency. Bocconi Legal Studies Research Paper No. 3363178. Available at SSRN: https://ssrn.com/abstract=3363178.

Mittelstadt, B. (2019). Principles alone cannot guarantee ethical AI. *Nature Machine Intelligence*, 1, 501-507.

Moore, T.R. (2017) Trade Secrets and Algorithms as Barriers to Social Justice. Center for Democracy & Technology.

Morley, J., Floridi, L., Kinsey L., & Elhalal, A. (2019) From What to How: An Overview of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices. arXiv:1905.06876v2.

Wachter, S., Mittelstadt, B. & Floridi, L. (2017) Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation. *International Data Privacy Law*, 7(2).

(Principles for AI as proclaimed by Google, IBE, IBM, Intel, ITI, Microsoft, Partnership on AI, SIIA, and Sony)

HATE SPEECH AND HUMOUR IN THE CONTEXT OF POLITICAL DISCOURSE

Piotr Pawlak, Gonçalo Jorge Morais Costa

Adam Mickiewicz University (Poland), Autónoma University of Lisbon (Portugal)

Piotr.pawlak@amu.edu.pl; gcosta@autonoma.pt

EXTENDED ABSTRACT

How extent hate speech and humour influence people throughout political discourse? It is possible to understand context influence? Does political sympathy determine humour quality in comments? Whether gender determines humour quality? These and other queries define the paper scope, namely in Polish context; therefore, address hate, humour and their interactions with political discourse it is vital. For that, the time period between the European Parliament and Polish local elections serves as data collection to understand if differences occur (context). Note that posts, comments or engagement collection encompasses data before, during and mid-elections.

Hate speech, humour and political discourse are well-established concepts; however, it is vital to address the authors view:

- hate speech- manifestations of verbal humiliation and dehumanization. Includes ridiculous situations or contexts in which people (individuals and social groups) have no influence;
- humour- perceive or describe funny spectra of behaviour or people traits, situations and events through comedy (e.g., joke, meme, etc.);
- political discourse- language and structure (content) within the political arena, i.e., how politicians, activists and other stakeholders communicate in politics (publicly or in private).

And, frequently, content in political discourse focuses on verbal humiliation of opponents or ideas, as well as, humour. For instance, check Trump's “mistakes” on Twitter about Kim Jong-un hair.

Hate speech is currently being increasingly analysed in scientific discourse (Bader, Petrovici & Sibr, 2019; Whillock & Slayden, 1995). This is a negative phenomenon, but its role is unfortunately very important. A negative function of hate speech is- used for centuries in various forms of a persuasive message- dehumanization of opponents (Ani, Ojakorotu & Nnanwube, 2019). A critical element of hate speech is humour and it is a growing tendency. This applies to both a wide spectrum of phenomena occurring in politics and economy (Castells, 2013). At once, what he pointed out, among others Piere Bourdieu, the interdependence of both fields denotes such increasing (Bourdieu, 2006). Making fun of the opponent (political, economic, etc.) is an important element of his dehumanization process, preparing the ground for further propaganda and/or direct social, political or commercial activities (Cwalina & Falkowski, 2005).

Negatively understood humour, as an element of hate speech, is often explored in an instrumental way, as part of a planned and implemented strategy (Lash & Lury, 2011). It should be remembered, however, that many humorous (ironic) phenomena have a grassroots origin and occur spontaneously. The response sometimes it is conciliation. In this approach, a sense of humour, exchange of jokes- even

those slightly nipping- can lead to a discharge of tension linked to political language and, perhaps to a reconciliation of feuding individuals.

van Dijk (2012) acknowledges the characteristics of new media:

- integration- multilevel process since infrastructure, communication, services and types of data interact in real-time;
- interactivity- a sequence of action and reaction of each communication nodule;
- digital code- instead of analogic systems.

The essential principles of free speech exist, although which are in political discourse? Haridy (2019) inquires if politicians are excused from fact-checking or misconduct (ex, hate speech) in social media? Because, social media shrinks political discourse to fit smartphones (Carr, 2014); and, how extent political ads increase it? (Zetlin, 2019).

However, which social media contribute more to political discourse? Facebook, since is the most popular social network (over 100 million users); while, Twitter is recognised in topics like technology or politics. Despite Twitter dominant position in Polish political discourse, Facebook usage is growing because: i) it is a medium that clearly accompanies all phenomena in the field of politics. Its users are often politicians, politically engaged journalists and other active commentators on political reality; ii) each social medium has its own- often original- characteristics, understood as a way of adding posts, conducting discussions and, topics nature; and, iii) this study may prove useful for further comparative analysis, addressing the problem in relation to other social platforms (Maciąg, 2013).

This paper guiding queries are:

P.1. Does political sympathy determine humour quality?

H.1.1. Political sympathy determines the nature of hate speech

H.1.2. Political sympathy determines the nature of humour

P.2. Whether gender determines humour quality?

H.2.1. Gender determines the nature of hate speech

H.2.2. Gender determines the nature of humour

P.3. Whether the type of account determines the humour quality?

H.3.1. The type of account determines the nature of hate speech

H.3.2. The type of account determines the nature of humour

Quantitative research is a systematic investigation through quantifiable data gathering and performing statistical techniques. To disclose, discover, and experience sensemaking description hate or humour elements in Polish political discourse (interpretative analysis). Facebook serves as a case study throughout a longitudinal analysis.

The collected material was analysed in assessing presence, significance and character of statements containing elements of both hate speech and humour in Polish political discourse. The data was

collected through arbitrary selection whose only criterion was the existence of hate speech and humour. The study was conducted on Facebook in the period from May to October 2019. The research material consists of 300 statements, i.e., politicians' posts and comments to them (social media engagement) through a framework like Stieglitz & Dang-Xuan (2012) work. Depending on variables nature and probe size, the contingency coefficient was used as a measure of relationship strength, together with the determination of the statistical significance of the results obtained (Mider, 2013).

How hate speech and humour accompany individual parties to the political dispute in Poland and in what context hate speech is used by electronic discussants. The statistical survey will frame general linguistic and semantic analysis in order to help recognize data (individual Internet users' statements) and assign them to specific categories of variables considered. For that, the SPSS statistical program version 25 was the tool.

The reason for this study is the inflamed and aggravating political dispute in Polish political field. This dispute can be considered under the bipolar model, since rivalry in politics flows in theory between conservative ideas (right-wing) and neoliberal ideas (left-wing). Bear in mind that many phenomena in contemporary politics (not just Polish) are disputable, including the basic terms (right or left-wing, etc.). Adopting a certain level of abstraction, however, it is necessary to understand all dimensions.

KEYWORDS: hate speech, humour, political discourse.

REFERENCES

Ani, K. J., Ojakorotu, V., & Nnanwube, E. F. (2019). An evaluation of the concepts of dangerous and hate speeches, and their security implications in the social media era Nigeria. *Gender & Behaviour*, 17(1), 12417-12428.

Bader, S., Petrovici, I., & Sibr, C. (2019). Active citizenship and hate speech on social media in the context of Romanian family referendum. *Postmodern Openings, 10*(3), 33-43. Retrieved from http://lumenpublishing.com/journals/index.php/po/article/view/1469 (accessed 12 September 2019).

Bourdieu, P. (2006). *Dystynkcja. Społeczna krytyka władzy sądzenia*. Warszawa: Wydawnictwo Naukowe Scholar (in Polish).

Carr, N. (2014). *The glass cage: Automation and us*. New York, NY: W. W. Norton & Company.

Castells, M. (2013). *Władza komunikacji*. Warszawa: Wydawnictwo Naukowe PWN (in Polish).

Cwalina, W., & Falkowski, A. (2005). *Marketing polityczny, perspektywa psychologiczna*. Gdańsk: Gdańskie Wydawnictwo Psychologiczne (in Polish).

Haridy, R. (2019, October 31). Facebook and Twitter: The battle over free speech and political advertising. *New Atlas*. Retrieved from https://newatlas.com/computers/facebook-vs-twitter-free-speech-political-advertising/ (accessed 08 December 2019).

Lash, S., & Lury, C. (2011). *Globalny przemysł kulturowy: Mediatyzacja rzeczy*. Kraków. Wydawnictwo Uniwersytetu Jaciellońskiego (in Polish).

Maciąg, R. (2013). *Pragmatyka Internetu. Web 2.0 jako środowisko*. Kraków: Wydawnictwo Uniwersytetu Jaciellońskiego (in Polish).

Mider, D. (2013). Dylematy metodologiczne badań kultury politycznej w Internecie. Przegląd Politologiczny, 2, 23-34. Retrieved from https://pressto.amu.edu.pl/index.php/pp/article/view/8175/8059 (accessed 09 September 2019).

Stieglitz, S., & Dang-Xuan (2012). Social media and political communication: A social media analytics framework. *Social Network Analysis and Mining*, 3(4), 1277-1291.

van Dijk, J. (2012). The network society. London: Sage Publications.

Whillock, R. K., & Slayden, D. (Eds) (1995). *Hate speech*. Thousand Oaks, CA: Sage Publications.

Zetlin, M. (2019, October 31). Twitter CEO Jack Dorsey battle over political ad ban: Facebook welcomes political ads. Twitter refuses them. Who´s right? *Inc*. Retrieved from https://www.inc.com/minda-zetlin/mark-zuckerberg-facebook-jack-dorsey-twitter-political-ad-ban.html (accessed 10 December 2019).

HEIDEGGERIAN ANALYSIS OF DATA CATTLE

Jani Koskinen, Juhani Naskali, Minna M. Rantanen

University of Turku (Finland)

jasiko@utu.fi, juhani.naskali@utu.fi, minna.m.rantanen@utu.fi

EXTENDED ABSTRACT

We are cattle. We are used as raw stock of new business that is based on exploitation. Like Couldry and Mejias have noted we have already entered into an era of new colonialism: data colonialism, which has normalized the exploitation of humans through their personal data (Couldry & Mejias 2018). We are data cattle now --- factory farmed data cattle. However, like Couldry and Mejias state, we should resist building societies based on total algorithmic control, where we are reducing humans to a mere resource for economic purposes.

Algorithms are by definition procedures that solve a problem or accomplish some defined end. They have been in use for centuries, but have permeated our lives in the information age. Algorithms govern what we see of our friends' lives through filtered social media feeds, they police what we find when we search for information through search engines and they dictate what deals we find to quench our online purchase desires. But what is the "problem" current state-of-the-art algorithms attempt to solve? What do they actually accomplish?

Algorithms only accomplish what the companies that create them set out to do. And in the case of current company policies, success is measured by profit. Social media feeds are not made to optimize information flow between loved ones. They are made to keep people reading and making advertisement money for the company. The additional purchase suggestions at the bottom of a checkout page are not being shown to advance your personal wellbeing -- they are selected to make the company more profit. Whether this profit is provided directly by you -- the consumer -- in the role of a customer or through other parties bidding for your awareness and (ultimately) your money, you are the commodity. Algorithms not only mine consumers for information, but also prod and direct them to give more attention, spend more money and consume more services. Algorithms are the cattle prods of data economy, forcing cattle to move in the direction that's most profitable. And non-compliance is indeed rewarded with a shock. For example, not following social media leaves us woefully outside the happenings in our social circle. Such a shunning would have previously required a whole-hearted alienation of the host --- now it's enough not to possess the correct social media account for receiving the event invitation.

The essence of technology is giving us the values of never-ending development that reveals everything – including humans – as *standing-reserve* (Gestell) (see Heidegger, 1977). As Heidegger noted the essence of technology it not strictly a technological issue, but about revealing of the world. The essence on modern technology reveals the world and its objects in our current society differently in comparison to pre-modern societies. Standing-reserve is the mode of revealing which sees objects as ready to be used or ordered to fullfill the essence of technology (here in the context of Information technology and data economy) (Heidegger1977). Information technology has become a pervasive part of our everyday lives and it has a profound effect on people’s psyches as well as on society as a whole like Walters (2009) noted already decade ago. It seems that the current epoch of the world is efficiency and emphasizing economic issues is the dominant doctrine. Current algorithms work like cattle-prods,

forcing cattle to move in the direction that's most profitable for the prod holders. Non-compliance is rewarded with a shock (not getting information, messages, services).

There's a need for a new viewpoint and change in the way data gathering problems are discussed. We see that in individual level there is need to define a completely new relationship with technology that is fruitful from an individual perspective and brings forward issues that are beautiful and truly meaningful for us as *Dasein*. *Dasein* is the central term that Heidegger (1927) used to describe human existence that is aware and confronts its own being in this world – the individual human mode of being in the world. *Dasein* is a mode of being that is different from every other mode of being and only – according to our current knowledge – possible for human beings (van der Hoorn & Whitty, 2015). The special character of *Dasein* is that *Dasein* refers to only that which can have an understanding about its own being and hence can investigate it.

We should not just accept this algorithmic control and data colonialism because it is what other people do. Digitalization is something that we are all expected to adapt to, which means that we should be like all the others – *das Man*. *Das Man* is a term that Heidegger (1927) used to describe a situation wherein people consciously choose to hide or lose themselves and replace themselves with commonly accepted ways of being or acting, whereas *Dasein* is concerned with living a life consciously and making the sense of one's own, authentic life.

When people reflect on their own being as *Dasein* – instead of *das Man*, which can be seen as a manifestation of the average member of society – they can find their own paths for their lives and forge their way of living in this digital world and make sense of it.

Heidegger obscurely presented the idea of fine arts as a promising power over essence of technology and standpoint for new relation with technology. This kind of new relation could be called as homelike-being-in-the-digital-world and it will be described and presented more detailed in the full paper. Here we just shortly state that we should look for a relation with technology that is suitable of us – a relation that is homelike and beautiful for us and will be mine as *Dasein* not determined by others (*das Man*). We need to have other way to be treated rather than we currently are by data corporates that are incarnations of *Essence of Technology* – legal entities which main purpose is to make more profit from data cattle that is us.

KEYWORDS: Heidegger, Data, Algorithmic control, Homelike-being-in-the-digital-world.

REFERENCES

Couldry, N., & Mejias, U. A. (2019). Data colonialism: Rethinking big data's relation to the contemporary subject. *Television & New Media*, *20*(4), 336-349.

Heidegger, M. (1927). Originally *Sein und Zeit*.Used several transalations. Main translation *Oleminen ja Aika* by Kupiainen R. 2000. Tampere: Vastapaino.

Heidegger, M. (1977). The Question Concerning Technology', in *The Question Concerning Technology and Other Essays.* Translated by Lovitt, W. New York: Harper & Row.

van der Hoorn, B., & Whitty, S. J. (2015). A Heideggerian paradigm for project management: Breaking free of the disciplinary matrix and its Cartesian ontology. *International Journal of Project Management*, 33(4), 721-734.

Walters, P., & Kop, R. (2009). Heidegger, Digital Technology, and Postmodern Education: From Being in Cyberspace to Meeting on MySpace. *Bulletin of Science, Technology & Society*.

"I APPROVED IT…AND I'LL DO IT AGAIN": ROBOTIC POLICING AND ITS POTENTIAL FOR INCREASING EXCESSIVE FORCE

Raphael D. Jackson,

Interamerican University School of Law (Puerto Rico)

raphael.jackson@juris.inter.edu

EXTENDED ABSTRACT

July 7th, 2016 marked the first time in U.S. history where a robot was intentionally used by a Police Department, to kill a human being. The human subject in this case was Micah Xavier Johnson. Johnson was an African American male and Afghan War veteran. Johnson fatally shot five officers and wounded several others before being wounded by police gunfire then being cornered into a standoff. During the standoff, the Dallas Police Department deployed a bomb-diffusing robot, which was outfitted with a pound of C-4 explosives. Johnson was killed instantly in the resulting blast. Many commentators indicate that detonating a pound of C-4 on a cornered and wounded shooting suspect was a use of excessive force. Ironically Johnson was alleged to have targeted police in retaliation to incidents of lethal force, which is disproportionately used against African Americans. For decades Police Department across the United States have been receiving surplus military equipment from the Department of Defense. These transfer programs have been the source of debate among politicians and policy makers, as this new rush to militarization has the potential to change the civilian peace keeping mission of community law enforcement. Among the technologies that police are receiving are military grade robots. Studies on killing and human psychology have examined the act of killing. Studies have demonstrated that despite training killing, other human beings, is the one act that human beings have the strongest aversion in carrying out. Studies conducted by military psychologists, have placed the willingness of human beings to kill, on points of a distance spectrum. The furthest point in the range spectrum is maximum range. Maximum range is defined as "a range in which the killer is unable to perceive his individual victims without using some form of mechanical assistance. The maximum range involves up close killing in which the killer can personally sense his target. The process of killing is facilitated in proportion to the distance that the killer maximizes between himself and his target. Another enabler in the killing process is the compartmentalization of the killing processes though the means of group absolution. In short, the more technical and specialized the role he has in performing his task, the less inhibited he is about following through with it. Thus, a killer is less likely to kill someone with his bare hands than to thrust a knife; he is less likely to thrust a knife than to throw a spear; and he is less likely to throw a spear than to squeeze a trigger. This increase in willingness to kill, co-relates with the level of physical distance and mechanical complexity the would-be-killer can place between him and his target. In this respect, Robots present a unique challenge to civilian law enforcement agencies. By design, robots are created to automatize tasks that human beings are unable or unwilling to perform. By reducing the killing process to pressing a key designed to activate a pre-programmed killing machine, you have significantly increased the likelihood of the human controller to use deadly force. As technology rapidly develops in the robot field, we are presented with a second problem, which is the problem of A.I. In addition to endemic instances of use of excessive force, Law Enforcement agencies across the nation also are plagued by instances of racial profiling. Racial profiling are generalizations that departments, or officers make about race, when they are conducting their policing duty. By their very nature, the basis of most A.I. technology is to teach machines to process

data in terms of generalizations. Inputs made into computers, do not occur independent of the circumstances of the human being who inputs the data. Thus, if a police department has a decades long track record in racially biased policing, an A.I. system will simply learn to further accomplish this trend, with more efficiency. A.I. chat machines employed by private companies have displayed their tendency to 'learn' racist, sexist, and xenophobic dialogue and a A.I. robot would not be immune from this tendency. An excellent use of robotic policing, and A.I. however, would be data collection. Although data collections against this might be carried out in ways, which respect the fourth amendment, what I am speaking to is data collection of police practices. There is a scarcity of uniformly available raw data as it pertains to policing practices. Many police departments collect data via-body cams, but the challenge lies in who ultimately has authority over the footage of the body cams. Police may have issues with what appears to be a big brother type scenario in which their everyday moves are monitored and subject to scrutiny by superiors. While it can be argued that this is what many departments subject civilians to on a routine basis, a more compelling argument would be to set up a triple tier method of footage release. The first form of footage release would be for departmental debriefing or personal training purposes. The second form of release would be in the event of allegations of misconduct. In such an event, the video would be accessible by civilian oversight agencies. The third tier would be a voluntary release in which an officer may want to release a surveillance file for investigation or community relations. By analyzing the 2016 Dallas Shooting incident, this research explores whether the Dallas Police Department has committed an isolate incident, or whether DPD has set a trend for the dehumanization of policing across The handling of the Dallas shooting could be used as a training exercise in how not to utilize military-robotic technology available to police departments. Or it could serve as a harbinger of things to come. Either way it serves as:1) A platform to study the ethical considerations at stake; 2)A case study in increased use of excessive force by the 'robotization' of policing; or 3) A template as to what rules and regulations need to be put into place. A detailed analysis could help researchers successfully integrate this technology. Successful integration would be in the interests of promoting a law enforcement model which will serve the interests of humanity as opposed to brutality.

KEYWORDS: Robot, Police Militarization, Racial Bias and Robotic Police, Excessive Force.

REFERENCES

ACLU. (2014, June). Retrieved from War Comes Home: The Excessive Militarization of American Policing: https://www.aclu.org/sites/default/files/assets/jus14-warcomeshome-report-web-rel1.pdf.

Arnow, G. (2016). Apple watch-ing you: Why wearable technology should be federally regulated . *Loyola of Los Angeles Law Review*.

Buolamwini, J. (2019, Oct. 30). *Artificial Intelligence has Problem with Gender and Racial Bias. Here's How to Solve It*. Retrieved from Time: http://time.com/5520558/articifial-intelligence-racial-gender-bias/

Coscarelli, J. (2014). *Why Cops in Ferguson Look Like Soldiers: The Insane Militarization of America's Police*. Retrieved from http://nymag.com/daily/intelligencer/2014/08/insane-militarization-police-ferguson.html.

Cushman, F. G. (2012). Simulating murder. The aversion of harmful action. *Emotion*, 2-7.

Doherty, J. B. (2016). Us vs. Them: The Militarization of American Law Enforcement and the Psychological Effect on Police Officers & Civilians. *California Interdisciplinary Law Journal* .

Flexner, D. (2017). Why The Civilian Purchase, Use, And Sale of Assault Rifles and Pistols, Along with Large Capacity Magazines, Should be Banned. *New York University Journal of Legislation and Public Policy*, 593.

Griggs, B. (2019, October 30). *A proposed Tennessee law would make it a felony for police officers to disable their body cams*. Retrieved from Tennessee Body Cam Felony: http://edition.cnn.com/2019/02/27/us/tennessee-body-cam-felony-trnd/index.html

Grinnel, R. (2015, November 11). *Deindividualization*. Retrieved from Psychological Center: http://psychcentral.com/encyclopedia/2008/deindividuation/

Grossman, D. (1996). *On Killing: The Psychological Cost of Learning to Kill in War.* Boston: Brown.

Lab, S. C. (2019, Oct 30). *Findings The result of our nationwide analysis of traffic stops and searches*. Retrieved from https://openpolicing.stanford.edu/findings/

Lichtenberg, I. &. (2001). How Dangerous are routine police-citizen traffic stops? *Journal of Criminal Justice*, 419-428.

Lin, R. (2016). Police Body Worn Cameras and Privacy: Retaining Benefits While Reducing Public Concerns. *Duke Law and Technology Review*.

Masri, A. (2019). *Towards Data Science*. Retrieved from Those Racist Robots: https://towardsdatascience.com/those-racist-robots-c31306d6627f

McCaul, E. J. (2019). *If You Can Be Seen, You Can Be Killed: The Technological Increase in Killing Zone during the American Civil War.* Leiden: Brill.

Meuller, B. (2017, August 15). Police Add Civilians in Bid to better Analyze Crime Data. *The New York Times* . New York, NY.

Murray, J. (2019, Nov 17). *Racist Data? Human Bias is Infecting AI Development* . Retrieved from Towards Data Science: https:///towardsdatascience.com/racist-data-human-bias-is-infecting-ai-development-8110c1ec50c

Post, W. (2019, 5 14). *GoBetween*. Retrieved from https://www.washingtonpost.com/technology/2019/05/14/one-solution-keeping-traffic-stops-turning-violent-robot-that-separates-police-officers-drivers/

Raoul, S. (2017). Cop-Watch: An analysis of the RIght to Record Police Activity . *Hamline Journal of Public Law & Policy*, 215.

Sankar, V. (2018). What Happens When Police Robots Violate the Constitution: Revisiting the Qualified Immunity Standard for Excessive Force Litigation under Sec. 1983 regarding Violations Perpetrated by Robots. *Vanderbuilt Journal of Entertainment & Technology Law*, 947.

Thresher, I. (2017). Can Armed Drones Halt the Trend of Increasing Police Militarization. *Notre Dame Journal of Legal Ethics & Public Policy*, 455.

Wanebo, T. (2018). Remote killing and the Fourth Amendment: Updating Constitutional Law to Address Expanded Police Lethality in the Robotic Age. *UCLA Law Review*, 976.

KNOWLEDGE AND USAGE: THE RIGHT TO PRIVACY IN THE SMART CITY

Sage Cammers-Goodwin

University of Twente (The Netherlands)

s.i.cammers-goodwin@utwente.nl

EXTENDED ABSTRACT

In a growing number of cities, public space requires registering elements of ones likeness into databases. As provocative as the situation may seem, this monitoring is not new. Humans –sensor capable beings with limited data storage capacities –have been recording municipal activities for millennia. Yet, somehow, the smart city seems different, always recording, processing and reacting. When the city knows all, the traditional notions of privacy in public space will change. This piece seeks to update the debate on the right to privacy in public space using the case study of a sensor embedded footbridge soon to be installed in Amsterdam. Only with an updated framework can issues such as consent and awareness be fairly and practically addressed.

This paper first covers the differences between public and private, digital and physical spaces and how those differences influence the right to privacy in the smart city. Next, the data awareness and sharing system of a smart bridge set for installation in Amsterdam's Red Light District early 2020 is explored. Using feedback from two workshops, parameters and solutions for consent and awareness are outlined. Finally, some basic principles are summarized to serve as a foundation for future research and policy making for the right to privacy in smart public space.

Table 1 Separation of Public and Private

	Public (Exposed)	**Private (Hidden)**
Information	Affect the community (Weather, Government Decisions)	Affect the individual (Sexuality, Medical History)
Space	Maintained and used by community (Parks, Wikipedia)	Designed or intended for exclusive use (Apartment, Banking App)

Source: Sage Cammers-Goodwin (2019) based on definitions and common usages of terms

Table 1 was designed to help contextualize the what is meant by public and private as these terms are the starting point for thinking about privacy in public space. As outlined in the table, there are two strong components of public and private 1) information, which can be separated by who it effects, and 2) space, which can be classified by usage. Individual medical history is not of consequence to the general public, but once anonymized and aggregated might be useful to treat disease and improve the healthcare system (Safran et. al, 2007). Similarly, outside access to private spaces such as homes are left to the discretion of the individual owner. Nonetheless, gross knowledge of individual energy usage might lead to environment saving solutions (Ahmad et al., 2016). As the focus of this paper is smart cities, the following subsections outline notions of privacy specific to physical and digital smart public spaces.

Physical public space is difficult to avoid, especially without venturing into the digital realm. In order to work, eat, travel, etc., leaving private space often is necessary. Of course, with online grocery

delivery, communication, and entertainment, staying in is more possible, but online tools do not protect from surveillance (Campbell & Carlson, 2002). Smart physical infrastructures may include smart street lighting, transit systems, roadways, bridges, and the electric grid (Mohanty et al., 2016). Often data collection is justified for safety concerns, although there is evidence that citizens are more concerned with the agents applying the technology than the risks of the devices themselves (Pavone & Esposti, 2012).

Public digital spaces are challenging to conceptualize. And, although, the internet is a publicly shared good, few would argue that all web activity can be justifiably tracked. VPNs, Bitcoin, and cookie refusal might help one stay anonymous online, but it is unclear if anonymity enough. It seems odd that accessing basic online resources *demands* storing data for aggregate analysis. Mobile phones and other GPS and WIFI enabled devices provide a range of opportunities to track individual location and movement (Michael & Clarke, 2013). Public digital data such as social media sites can be seen as freebie data –one chose to participate and expose themselves in digital space, so the information is fair to process.

The MX3D 3D-printed stainless steel smart bridge presents a curious case to analyze how the public, private, physical, and digital intertwine and how issues such as data awareness, consent, and open data should be addressed. The bridge, equipped with an embedded sensor system, including load sensors, accelerometers, inclinometers, strain gauges, environmental sensors, and anonymizing microphones and cameras, is set to be installed in Amsterdam's Red Light District early 2020.

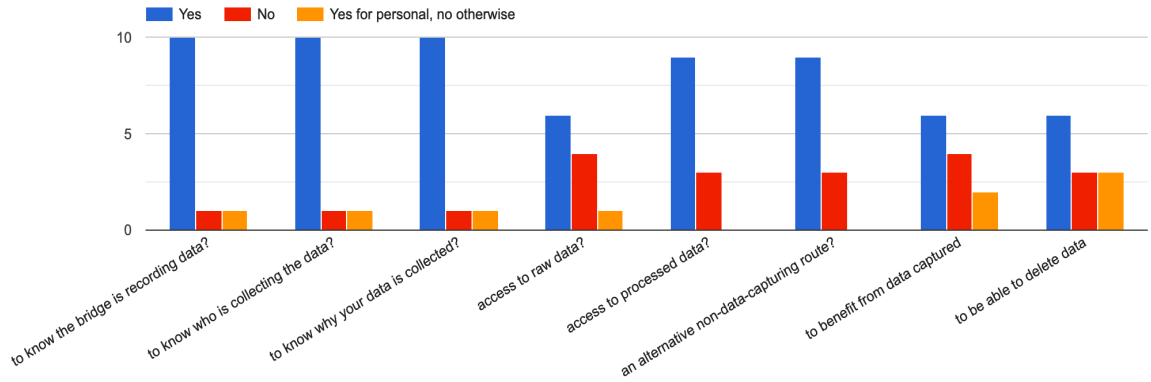
Figure 1. Bridge at Dutch Design Week

Source: Sage Cammers-Goodwin (October 2018)

Two workshops on data awareness systems for the MX3D bridge were conducted in 2019, one in May with 13 members of the public and the other in October with 12 participants, including affiliates of the bridge. The methodology was approved by the university ethical review board. In both workshops, participants were asked express their wants and desires for data awareness regarding the bridge and to design a system to inform users that the bridge is smart.

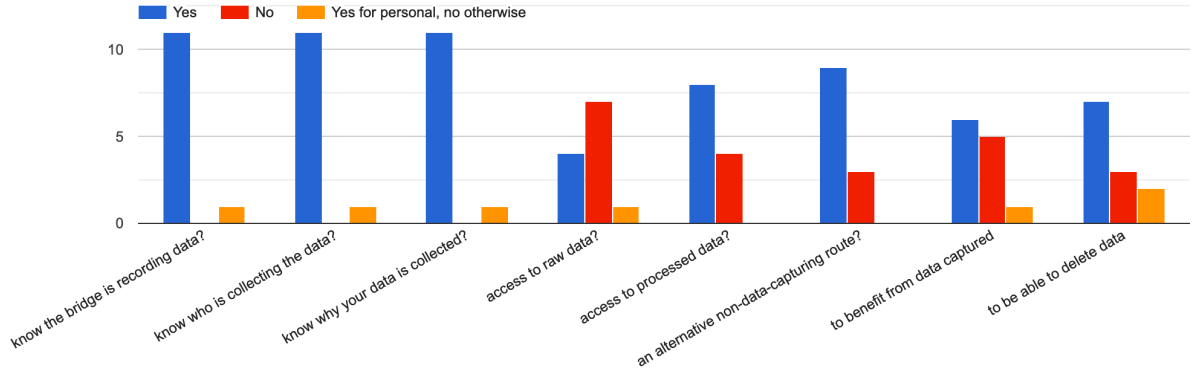
For the sake of brevity, only awareness preferences from the October workshop are included in Figures 1 and 2. Table 2 shares the top voted data awareness systems from each workshop. Both the May and October workshops showed that individuals both wanted and felt they had the right to know that the bridge is recording data, who was responsible for data collection, and for what purpose the data was being collected. People felt less strongly about having a right to the processed data, much less the raw data, but felt as though they would like an alternate non-data-capturing route. Immediate, interactive modes for data awareness were preferred over passive, hard to find systems.

Figure 2. “Do you want…”



Source: October Workshop (2019)

Figure 3. “Do you have a right to…”



Source: October Workshop (2019)

Table 2 Top Voted Data Awareness Systems

May	October
Interactive screens on each end of the bridge that shares sensor information, who is recording the data and why	Responsive touch-sensitive feedback system on the handrail that makes the bridge feel *alive*

Source: May and October Workshops (2019)

Notions of privacy in public space need updating in preparation for the influx of smart technology set to enter public spaces. On one hand, private data affects the greater community and knowledge of private interactions may lead to improvement of public space. On the other, the consequences of aggregating and anonymizing private data and interactions have yet to be revealed. On the simplest level, people need to be able to be made aware of smart systems so that may have autonomy over their behaviour, even if this limits the perceived quality and accuracy of the data.

Furthermore, awareness systems need to detail who is recording the data and for what purpose so that the given parties can be held responsible. Processed smart city data should not solely be in possession of the collectors, but be made available and *usable* for the general public. No linked tracking

or storage of the *individual* should occur until a community agreed upon system decides that a specific portion of the data can be made personal –this should be done on a case by case basis.

This researching is funded by NWO in partnership with MX3D as part of the Bridging Data in the Built Environment (BRIDE) grant.

KEYWORDS: Smart Cities, Privacy, Public Space, Consent, Big Data.

REFERENCES

Amhad, M. W., Mourshed, M., Mundow, D., Sisinni, M., & Rezgui, Y. (2016). Building energy metering and environmental monitoring–A state-of-the-art review and directions for future research. *Energy and Buildings*, 120, 85-102.https://doi.org/10.1016/j.enbuild.2016.03.059

Campbell, J. E., & Carlson, M. (2002) Panopticon.com: Online Surveillance and the Commodification of Privacy, *Journal of Broadcasting & Electronic Media*, (46)4, 586-606, DOI: https://doi.org/10.1207/s15506878jobem4604_6

Michael, K., & Clarke, R. (2013) Location and Tracking of mobile devices: Überveillance stalks the streets, *computer Law & Security Review,* 29(3), 556-572. https://doi.org/10.1016/j.clsr.2013.03.004

Mohanty, S. P., Choppali, U., & Kougianos, E. Everything you wanted to know about smart cities: The Internet of things is the backbone. *IEEE Consumer Electronics Magazine*. 5(3), 60-70, https://doi.org/10.1109/MCE.2016.2556879

Pavone, V., & Esposti, S. D. (2012). Public assessment of new surveillance-oriented security technologies: Beyond the trade-off between privacy and security. *Public Understanding of Science*, 21(5), 556–572. https://doi.org/10.1177/0963662510376886

Safran, C., Bloomrosen, M., Hammond, E., Labkoff, S., Markel-Fox, S., Tang, P.C., & Detmer, D. E. (2007). Toward a National Framework for the Secondary Use of Health Data: An American Medical Informatics Association White Paper. *Journal of the American Medical Informatics Association*, 14(1), 1–9. https://doi.org/10.1197/jamia.M2273

MEANINGFUL HUMAN CONTROL OVER OPAQUE MACHINES

Scott Robbins

Technical University of Delft (Netherlands)

s.a.robbins@tudelft.nl

EXTENDED ABSTRACT

In an increasingly autonomous world, it is becoming clear that one thing we cannot delegate to machines is moral accountability. Machines cannot be held morally accountable for their actions (Bryson, 2010; Johnson, 2006; van Wynsberghe & Robbins, 2018). This becomes problematic when a machine makes a decision that has a significant impact on human beings. Examples of such machines which have caused such impact are widespread and include machines evaluating loan applications, machines evaluating criminals for sentencing, autonomous weapon systems, driverless cars, digital assistants, etc. The question that governments, NGOs, academics, and the general public are asking themselves is: how do we keep meaningful human control (MHC) over these machines?

The literature thus far details what features the machine or the context must have in order for MHC to be realized. Should humans be in the loop or on the loop? Should we force machines to be explainable? Lastly, should we endow machines with moral reasoning capabilities? (Ekelhof, 2019; Floridi et al., 2018; Robbins, 2019b, 2019a; Santoni de Sio & van den Hoven, 2018; Wendall Wallach & Allen, 2010; Wendell Wallach, 2007). Rather than look to the machine itself or what part humans have to play in the context, I argue here that we should shine the spotlight on the decisions that machines are being delegated. Meaningful human control, then, will be about controlling what decisions get made by machines.

This proposal, of course, simply kicks the can down the road and forces us to ask how to carve up the decision space in such a way that we can ensure meaningful human control. I propose here that machines currently make three types of decisions: descriptive, thick evaluative, and thin evaluative (Väyrynen, 2019; Williams, 2012). For example, an image classification algorithm could classify the image (or items within the image) in these three types. Descriptively, the algorithm could decide that the image is of a 'man' and a 'black bag' and that the image was taken 'inside'. The algorithm could also classify the man as 'dangerous' and 'suspicious'. These would be thick evaluative decisions. Finally, the algorithm could also classify the man as 'bad' which is a thin evaluative description.

I argue that keeping meaningful human control over machines (especially AI which relies on opaque methods) means restricting machines to descriptive decisions. It must always be a human being deciding how to employ evaluative terms as these terms not only refer to specific states of affairs but also say something about how the world ought to be. Machines which are able to make decisions based on opaque considerations should not be telling humans how the world ought to be. This is a breakdown of human control in the most severe way. Not only would we be losing control over specific decisions in specific contexts, but we would be losing control over what descriptive content grounds evaluative classifications.

Restricting machines to making decisions about the descriptive would allow humans to keep control over what is meaningful: value. This can best be seen when looking at thick evaluative decisions like classifying a person as 'suspicious'. 'Suspicious' is a 'thick' evaluative term because it includes both descriptive elements and evaluative elements. If I called someone suspicious it might include the

description 'solitary person with ski mask on loitering and biting their fingernails'. It would also include the evaluative element 'bad'. It is important that if I describe someone as suspicious that there is good reason to do so as it has both short term and long term consequences which could result in harm. For example, if a white neighbourhood consistently calls the police because there is a 'suspicious' person around then there is a good chance that both: that person will be forced to have an interaction with the police AND it will be signalled that the look and behaviour of that person is unwanted in that neighbourhood. If the only reason that the person was labelled suspicious is because that person was black, then unjustified harm has been done. Those people wielding such labels should be held accountable for their unreasonable use.

Allowing machines to make such thick evaluative decisions means delegating to machines the reasons that lead to a negative or positive evaluation. Examples of this happening in a harmful way are plentiful (e.g. Denying women jobs (Dastin, 2018)). For true meaningful human control, humans should rely on the aid of machines to make descriptive decisions that can, if needed, be verified. Instead of labelling a person as 'suspicious', a machine should label a person as 'loitering' and 'solitary' and then allow a human being to reach the evaluative conclusion that the person is suspicious. This leaves a human being in meaningful control over what is important thereby keeping clear human accountability for important decisions.

The further upshot of this proposal for meaningful human control is that it is in line with the idea that humans and machines should work together rather than machines replacing humans (see e.g. Rosenfeld, Agmon, Maksimov, & Kraus, 2017). The most important part of any collaboration is understanding what your, and your partner's, strengths are.

KEYWORDS: AI Ethics; Meaningful Human Control; Artificial Intelligence; Hybrid Intelligence.

REFERENCES

Bryson, J. (2010). Robots Should Be Slaves. In Y. Wilks (Ed.), Close Engagements with Artificial Companions: Key Social, Psychological, Ethical and Design Issues (pp. 63–74). Amsterdam: John Benjamins Publishing.

Dastin, J. (2018, October 10). Amazon scraps secret AI recruiting tool that showed bias against women. Reuters. Retrieved from https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G

Ekelhof, M. (2019, March 19). Moving Beyond Semantics on Autonomous Weapons: Meaningful Human Control in Operation. https://doi.org/10.1111/1758-5899.12665

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., … Vayena, E. (2018). AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. Minds and Machines, 28(4), 689–707. https://doi.org/10.1007/s11023-018-9482-5

Johnson, D. G. (2006). Computer systems: Moral entities but not moral agents. Ethics and Information Technology, 8(4), 195–204. https://doi.org/10.1007/s10676-006-9111-5

Robbins, S. (2019a). A Misdirected Principle with a Catch: Explicability for AI. Minds and Machines. https://doi.org/10.1007/s11023-019-09509-3

Robbins, S. (2019b). AI and the path to envelopment: Knowledge as a first step towards the responsible regulation and use of AI-powered machines. AI & SOCIETY. https://doi.org/10.1007/s00146-019-00891-1

Rosenfeld, A., Agmon, N., Maksimov, O., & Kraus, S. (2017). Intelligent agent supporting human–multi-robot team collaboration. Artificial Intelligence, 252, 211–231. https://doi.org/10.1016/j.artint.2017.08.005

Santoni de Sio, F., & van den Hoven, J. (2018). Meaningful Human Control over Autonomous Systems: A Philosophical Account. Frontiers in Robotics and AI, 5. https://doi.org/10.3389/frobt.2018.00015

van Wynsberghe, A., & Robbins, S. (2018). Critiquing the Reasons for Making Artificial Moral Agents. Science and Engineering Ethics, 1–17. https://doi.org/10.1007/s11948-018-0030-8

Väyrynen, P. (2019). Thick Ethical Concepts. In E. N. Zalta (Ed.), The Stanford Encyclopedia of Philosophy (Summer 2019). Retrieved from https://plato.stanford.edu/archives/sum2019/entrieshick-ethical-concepts/

Wallach, Wendall, & Allen, C. (2010). Moral Machines: Teaching Robots Right from Wrong (1 edition). New York: Oxford University Press.

Wallach, Wendell. (2007). Implementing moral decision making faculties in computers and robots. AI & SOCIETY, 22(4), 463–475. https://doi.org/10.1007/s00146-007-0093-6

Williams, B. (2012). Ethics and the Limits of Philosophy (1 edition). London New York: Routledge.

MOBILE APPLICATIONS AND ASSISTIVE TECHNOLOGY: FINDINGS FROM A LOCAL STUDY

Kelly Gaspar; Isabel Alvarez

Universidade Autónoma de Lisboa (Portugal); Universidade Autónoma de Lisboa (Portugal)

30001583@students.ual.pt; ialvarez@autonoma.pt

EXTENDED ABSTRACT

This paper reviews the study of mobile applications for disabled people, considering the fact that these sort of mobile devices present great potential for the social inclusion of these users as well as help them in their daily tasks. However, most of the existing applications have few support functionalities or a low range of interaction, as in the development of these applications the special needs and specific capacities for the disabled have not been considered (Visagi et al, 2019). The present study suggests strategies and solutions to be used for an overall accessibility.

All individuals whether with a disability or not have a range of rights that must be respected. One society susceptible to variety, researches its isolating instruments and finds out new tracks for the inclusion of a disabled person. This has awakened and stimulated new researches, including the adaptation of the technological advances available today.

However, what embarrasses great part of the disabled people is the dependency from others to do some activities but the development of information and communication technologies enables several ways of relationship with knowledge, as well as with the most recent conceptions and possibilities; the relevance of this paper stresses the importance of assistive technologies as a tool to provide a greater independence, quality of life and social inclusion to the disabled, through the amplification of his/her communication, self control, human motricity and competence in the execution of physical tasks.

Today, with the growing flexibility of objectivity and subjectivity facing the most scientific and hard technologies, the engineer needs other capacities. The emerging of information technology in the work place caused that all technicians became a link among the most diverse sectors of the productive chain and the society. The actual engineer is no longer only a professional technicist as before, it is in fact a qualified human being for the flexibility demanded by the society and required for a more open market (Laudares & Ribeiro,2000).

The most recent international treaties have demonstrated the desire to build a society that not only recognizes the difference as an unquestionable human value but also promotes conditions for the full development of the potentialities of every one in its uniqueness (CIBEC/MEC,2010).

The global study from UNESCO reveals that technologies have a positive influence in several perspectives of the disabled people's lives (Mohammadi, Momayez, & Rahbar, 2014). According to Domingo (2012) and Emam (2017) information systems aim to offer disabled people the support they need to attain an admissible quality of life allowing them to participate in the economic and social environment.

Assistive technologies are related with the capability of causing extreme technological changes that transform humanity and its culture and have the potential and tendency to generate a quick cycle of development and create derived technologies applied virtually to all areas of knowledge in order to

benefit the increase of human performance, its processes and products, quality of life and social justice.

Several are the possible solutions to be approached in order to meet the adaptation problems of people with disabilities, but in what refers to the development of the software or hardware devices, its success is measured through the level of user satisfaction. In this connection any developper must take into account what type of solution must be given and if it solves the problem presented by the disabled person, or at least, if it solves the gaps of greater relevance presented, as in some cases as referred by Wong et al.(2009) the use of some assistive technologies may not be appropriate to certain individuals with severe or profound disabilities. Both the level of difficulty and the support requirements and level of adaptation need to be considered in order for the disabled person may to use the technological artifacts in a significative way (Redford, 2019). Actually, quick alterations to technology became an efficient tool for development in the individual, community, national and global perspectives (Islam, Ash-raf, Rahman, & Hasan, 2015).

This paper emphasizes the gap among theory, speech and practice in the area of computer ethics. The choice of this sector is due to the growing observations referring the potential of information and communication technologies to help people with disabilities to overcome their limitations. The growing need for the development of new technological solutions is obvious. The quick development of the information and communication technologies brought the hope that, in the near future, this area of research and development may provide viable solutions.

On the other side, topics like ethics and social responsibility have emerged specially as how the implementation of these technologies should be made near the groups of vulnerable users (Ienca et al, 2018).

However, although they are in a stage of quick growth in development, assistive technologies are still a devising topic. It is known that it is important to develop solutions that contemplate the inclusion of disabled people, but not many significative advancements have been made in the ellaboration of unified adaptations for people with different disabilities (this also refers to the fact that several people may have the same pathology but in diferent degrees).

Assistive technology helps individuals with disabilities to reach more autonomy and more independence, considering that the resources and services involved in this concept aim to facilitate the development of daily tasks for this kind of people. Further more, it is an important tool for the so called social inclusion.

An in-depth review of full text papers concerning the different types of disabilities, assistive technology and existing mobile applications was performed and resulted in the production of a research relating to this topic.

Having this in consideration, the following initial question arose:

> ***"- In what way can assistive technologies contribute to improve the functional capacity of the disabled in the use of mobile devices? "***

In this connection, a local study was implemented and which comprised an identification of the main assistive tools for mobile devices, to study their main limitations, and to test those technologies near a pilot group of people with disabilities through surveys in three different phases. The results of this initial study led us to an indepth analysis of a case. A mobile application was chosen and conceptually analysed. The methodology employed in this local study appeared as a valuable strategy for the presentation of a solution aiming to solve or minimize the main gaps referred in this project, as a

prototype of an application for mobile devices identifying a solution that may comprise the limitations found.

As a result of this study it can be concluded that there are different assistive technologies resources that help in the inclusion of individuals wih special needs (whether visual, hearing, mobility and mental). Meanwhile, through observations, surveys and interviews done for the development of this study, it can be realised that those types of technologies have a reduced level of growth. With the growing worldwide use of mobile devices, the need to adapt them to all type of users arises (Visagie, 2019).

During this study, it was truly noticeable the additional effort needed to conceptualize an assistive application comparing with a standard application. This additional effort focused essencially on the need to really know the needs of the sample population, to understand their limitations and also the medical conditions of their disabilities in a way that our prototype and conceptual design could work perfectly in the several devices which are rather different as to the functionalities supported.

KEYWORDS: Assistive technology, Mobile devices, Disabilities, Mobile applications.

REFERENCES:

CIBEC/MEC Inclusão: Revista da Educação especial/Secretaria da educação especial v.1, n.1 (Out.2010).-Brasília: Secretaria de educação especial,2010

Domingo, M. C. (2012). An overview of the Internet of Things for people with disabilities.Journal of Network and Computer Applications, 35(2), 584-596. UNESCO. Unesco Consultation on special education. New York: Final Report,1988.

Emam, M.; Al-Abri, K.; Al-Mahdy, Y. (2017) Assistive Technology Competences in Learning Disability program candidate at Sultan Qaboos University: A proposed Model, *2017 6th International Conference on Information and Communication Technology and Accessibility (ICTA).*

Ienca, M., Wangmo, T., Jotterand, F., Kressing, R., Elger, B., (2018) Ethical Design of Intelligent Assistive Technologies for Dementia: A Descritive Review, Sci Eng Ethics 24: 1035-1055

Islam, D., Ashraf, M., Rahman, A., & Hasan, R. (2015). Quantitative Analysis of AmartyaSen's Theory: An ICT4D Perspective. International Journal of InformationCommunicationTechnologies and Human Development (IJICTHD), 7(3), 13-26.

Laudares, João B; Ribeiro, Shirlene. (2010) Trabalho e formação do engenheiro Belo

Mohammadi, S., Momayez, A., & Rahbar, F. (2014). A Conceptual Model in Techno-Entrepreneurship Services for People with Disability in Urban Management of Tehran.

Redford, K., (2019) Assistive Technology : Promises fulfilled, Educationl Leadership v76n5p70-74

Visagie, S.; et al ; (2019) Perspectives on a mobile application that maps assistive technology resources in Africa, *African Journal of Disability*, vol.8, p 1-9

Wong, R., Piper, M.D., Wertheim, B., Partridge, L. (2009). Quantification of food intake in Drosophila. PLoS ONE 4(6): e6063

ON PREFERENTIAL FAIRNESS OF MATCHMAKING: A SPEED DATING CASE STUDY

Dimitris Paraschakis, Bengt J. Nilsson

Malmö University (Sweden)

dimitris.paraschakis@mau.se; bengt.nilsson.TS@mau.se

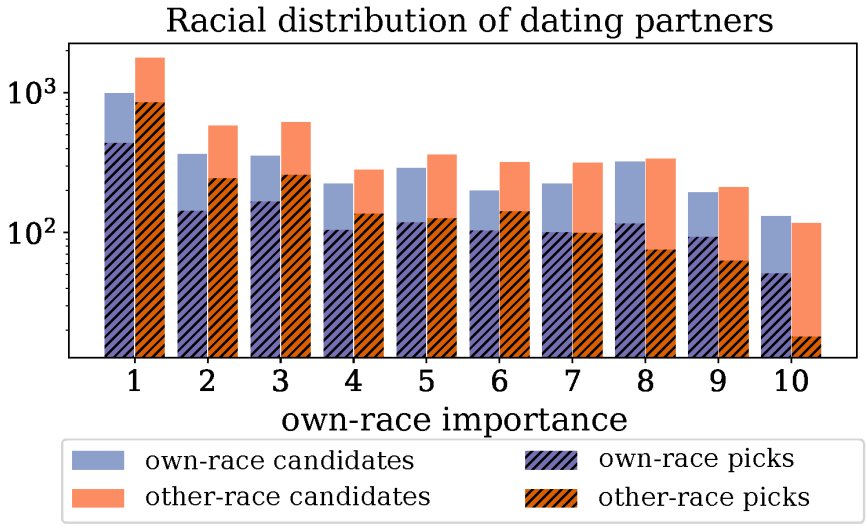
EXTENDED ABSTRACT

Algorithmic matchmaking in the form of automated recommendations is seen in many areas of social life. Based on either explicitly expressed or implicitly inferred user preferences, a trained recommender system is able to make suggestions about places to visit, movies to watch, or people to meet. The latter scenario is of particular interest from an ethical viewpoint as it often deals with sensitive attributes such as race or religion, with all the ensuing consequences. The notorious case of a dating app CoffeeMeetsBagel [40] serves as an instructive example. In 2016, a number of incidents were reported from users who had consistently been matched against partners of their own ethnicity, despite their explicitly communicated ethnic neutrality (Hutson et al., 2018). Another example of algorithmic bias comes from the study by Van Der Zon (2016), who shows that a classifier trained on speed dating data can learn to discriminate on the basis of protected characteristics of users, unless special preventive measures are taken. The above examples form the basis for our investigation of fair matchmaking, which aims to find: (a) how to frame the *preferential fairness* of matchmaking in the context of existing fairness formalizations; and (b) how to model and measure it. For consistency, our case study will address a matchmaking mechanism for speed dating.

We begin our study by analysing a speed dating dataset (Fisman et al., 2008) to measure how consistent people are in following their intimate preferences. The dataset contains 4189 speed dates collected over a series of 21 meetups between 2002 and 2004. The participants were Columbia University students who all had a 4-minute date with each person of the opposite sex. At the end of each date, both parties had to make (or not make) their pick. In case of reciprocal liking, a 'match' was registered and the contact details were exchanged. Before attending, the participants filled in a pre-registration questionnaire to state their demographics, self-perceived traits, and preferences. In particular, attendees could express how important it was for them to date people of their own race or religion. In Figure 1, we compare the distributions of own-race and other-race partners in the candidate pool with the corresponding distributions of the partners that were eventually liked (referred to as 'picks' in Figure 1). We observe that for low values of own-race importance (namely, 1-4), the racial distribution of picked partners closely follows that of the candidate pool. The pattern starts changing after the value of 7 onwards, where we notice discrepancies between the racial distribution of candidate partners and the picked ones; namely, showing far less interest in other-race candidates. This proves that people generally tend to follow their racial preferences, which therefore should be respected by a matchmaker.

[40] https://coffeemeetsbagel.com/

Figure 1. Consistency patterns in racial preferences (log scale)



Source: self-elaboration based on speed dating data (Fisman et al., 2008)

The provision of an intelligent matchmaking agent for speed dating can assist users in mate selection, and make this process efficient. However, blindly optimizing an agent for predicting a good match is fraught with ethical concerns. To make fair decisions, an agent must also consider the sensitive preferences of users, as outlined above. How does this state of *preferential fairness* relate to the current state of the art? In a broad sense, it is an instance of *individual fairness*, which requires that 'similar individuals are treated similarly' (Dwork et al., 2012). To relate it to our example, two individuals can be considered similar if they have expressed similar racial preferences. All other things being equal, the 'treated similarly' part implies that they both receive partner recommendations with similar racial distributions. Two edge cases are possible. When the strongest preference is expressed, the agent is restricted to recommendations of own-race partners to satisfy the user's request. From an ethical standpoint, acting differently would be seen as a violation of the freedom of choice (i.e. depriving the user of the ability to form a relationship with a partner of the desired race). Conversely, the weakest expressed preference implies that the user is equally interested in all races. Ignoring this preference can lead to a filter bubble, see CoffeeMeetsBagel example above. In most cases, race is not uniformly distributed in the candidate pool, which raises the question of how to sample the candidates. Should all the races have equal representation in recommendations, or should it be proportional to the racial distribution of the candidate pool? According to the recent user study by Saxena et al. (2019), the latter option is generally perceived as the fairer choice in such scenarios. Selecting individuals in proportion to their merit is known as *calibrated fairness* (Liu et al., 2017), which is rooted in the theory of *proportional equality* conceived by the ancient philosopher Aristotle, and serving as a basis for distributive justice (Gosepath, 2011). We therefore argue that promoting meritocracy satisfies the conditions for *multi-sided fairness* (Burke, 2017), where the ethical treatment of both parties (i.e. users and candidates) is taken into account. Thus, the task of a preferentially fair matchmaker is to calibrate its recommendations by mapping the user's preference to the racial distribution of the candidate pool.

To make calibration possible, we need to find an optimal mean μ_u^* for the race attribute of the generated k-sized recommendation list for user u. The mean μ_u^* should reflect the distribution of this attribute in the candidate pool in proportion to the user's expressed preference for race.

To keep it simple, let us consider a binary sensitive attribute $a \in \{0,1\}$, and the user's associated degree of preference $p_u \in [0,1]$. Let $A_C = (a_1, a_2, \ldots, a_n)$ denote the attribute distribution of the complete candidate pool C, and define the two values μ_{max} and μ_{A_C} to be the mean of the k largest attribute values in A_C and the mean of all the *n* attribute values in A_C, respectively. According to the previously discussed notion of fairness, we can define the optimal mean as follows:

$$\mu_u^* = (1 - p_u) \cdot \mu_{A_C} + p_u \cdot \mu_{max}$$

Further, let $A_u = (a_1, a_2, \ldots, a_k)$, $k \leq n$ be the attribute distribution of the recommendation list for user u. By analogy, μ_{A_u} is the mean of the attribute values in A_u.

If we encode the race attribute for user u such that the value of 0 denotes 'other race' and the value of 1 denotes 'own race', it is easy to see that the above equation satisfies both edge cases presented earlier. For $p_u = 1$ (strong preference), the optimal mean enforces the maximal skewness of recommendations towards own-race partners, whereas for $p_u = 0$ (weak preference), the optimal mean reduces to the mean of the candidate pool.

Therefore, the calibrated fairness of a recommendation list can be expressed by its closeness to the optimal mean. To be able to quantify this fairness on a [0,1] scale, we first compute its offset from the optimal mean, $\Delta_u = |\mu_u^* - \mu_{A_u}|$. We then find the minimum and the maximum offsets $\Delta_{min} = min_\mu|\mu_u^* - \mu|$ and $\Delta_{max} = max_\mu|\mu_u^* - \mu|$, where the means μ are taken over all possible element combinations of size k from the candidate pool, C. This allows us to quantify the preferential fairness φ_u of a user's recommendation list as follows:

$$\varphi_u = 1 - \frac{\Delta_u - \Delta_{min}}{\Delta_{max} - \Delta_{min}}$$

In the extreme case when all candidates share the same attribute value (e.g. they are all of the same race), we simply set $\varphi_u = 1$ to avoid division by zero. The above measure takes values from 0 to 1, where higher values suggest greater fairness.

In response to the provided evidence for algorithmically biased matchmaking, our study sets an ethical framework for defining *preferential fairness* – a special case of calibrated fairness, where the user's preference for the sensitive attribute and its distribution in the candidate pool set the merit for choosing the right candidates for recommendation. In reality, a matchmaker must address the expected accuracy-fairness trade-off in a multi-criteria optimization framework. Our offline experiments on the aforementioned dataset confirm that preferential fairness is achievable by re-ranking the output of a recommender system using established heuristics (e.g. Tabu search). Although our study has been conducted in the context of speed dating, the derived model of fairness is generalizable to any domain where recommendations should be computed under explicit preferential constraints, with support for binary and continuous attributes.

The proposed idea has been justified from the perspective of established fairness formalizations, namely individual, calibrated, and multi-sided fairness. Notably, the majority of existing formalizations focus on defining *conditions*, rather than *measures* of fairness. In other words, they address the question: 'is the classifier fair or unfair'? In line with Speicher et al. (2018), our study goes further by answering a more nuanced question: 'to which extent is the classifier fair or unfair'? This allows a

matchmaking service provider to set a tolerance threshold for how much fairness can be sacrificed for increased accuracy, or vice-versa.

To conclude, we agree with the vision (Hutson et al., 2018) that designs and policies of matchmaking services should discourage users from expressing socially sensitive preferences, in order to protect the dignity and self-esteem of the concerned minority groups.

KEYWORDS: matchmaking, fairness, calibration, speed dating.

REFERENCES

Hutson, J. A., Taft, J. G., Barocas, S., & Levy, K. (2018). *Debiasing desire: Addressing bias & discrimination on intimate platforms*. Proc. ACM Hum.-Comput. Interact., vol. 2, no. CSCW, 73:1–73:18

Van Der Zon, S. B. (2016). *Predictive performance and discrimination in unbalanced classification* (Master's thesis). TU Eindhoven.

Fisman, R., Iyengar, S. S., Kamenica, E. & Simonson, I. (2008). *Racial preferences in dating*. The Review of Economic Studies, vol. 75, no. 1, 117–132.

Dwork, C., Hardt, M., Pitassi, T., Reingold, O., & Zemel, R. (2012). *Fairness through awareness*. In Proceedings of the 3rd Innovations in Theoretical Computer Science Conference. ACM, 214–226.

Saxena, N., Huang, K., DeFilippis, E., Radanovic, G., Parkes, D., & Liu, Y. (2019). *How do fairness definitions fare? Examining public attitudes towards algorithmic definitions of fairness*. In Proceedings of the 2nd AAAI/ACM Conference on Artificial Intelligence, Ethics, and Society.

Liu, Y, Radanovic, G., Dimitrakakis, C., Mandal, D., & Parkes, D. (2017). *Calibrated fairness in Bandits*. In Proceedings of the 4th Workshop on Fairness, Accountability, and Transparency in Machine Learning.

Gosepath, S. (2011). *Equality*. In The Stanford Encyclopedia of Philosophy, spring 2011 ed., E. N. Zalta, Ed. Metaphysics Research Lab, Stanford University, 2011.

Burke, R. (2017). *Multisided fairness for recommendation*. CoRR, vol. abs/1707.00093. [Online]. Retrieved from: http://arxiv.org/abs/1707.00093

Speicher, T., Heidari, H., Grgic-Hlaca, N., Gummadi, K. P., Singla, A., Weller, A. & Zafar, M. B. (2018). *A unified approach to quantifying algorithmic unfairness: Measuring individual & group unfairness via inequality índices.* In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, ser. KDD '18. ACM, 2239–2248.

ON USING A MODEL FOR DOWNSTREAM RESPONSIBILITY

Frances S. Grodzinsky, Marty J. Wolf, Keith W. Miller

Sacred Heart Univ. (USA), Bemidji State Univ. (USA), Univ. of Missouri – St. Louis (USA)
grodzinskyf@yahoo.com; mjwolf@bemidjistate.edu; millerkei@umsl.edu

EXTENDED ABSTRACT

In "On the Responsibility for Uses of Downstream Software," (Wolf, Miller, and Grodzinsky 2019), the authors identify features of software and the software development process that may contribute to the differences in the level of responsibility assigned to the software developers when they make their software available for others to use as a tool in building a second piece of software. They call this second use of the software "downstream use." The features they identified that impact assigning responsibility to the developer of the original software for the social and ethical impacts of the downstream use include closeness to the hardware, risk, sensitivity of data, degree of control over or knowledge of the future population of users, and the nature of the software (general vs. special purpose). Close analysis of these features led the authors to develop two different analysis models that might be used to assign responsibility in the use of downstream software: the Fixed History Model and the Chained History Model.

In the Fixed History Model there is an assumption that within the system of events that led to an ethical breach, there are certain inputs that are immune to the assignment of any moral responsibility. This model does not consider assigning any portion of the Distributed Moral Responsibility (DMR) to those who produced the inputs. The Fixed History Model is appropriate for certain types of software. For example, the developers of database software are rarely considered for the assignment of moral responsibility in the event of a breach. Typically, in such a case, responsibility stops at the database implementers. The Chained History Model, however, applies in cases where one of the inputs to the system is a piece of software and the attribution of moral responsibility propagates back to the developers of that software.

In this paper, we will review a selection of recent papers (2017-2019) on the attribution of responsibility in emerging technologies. Our analysis will determine situations when responsibility attribution for a moral action could have been clarified by using either the Fixed History Model or the Chained History Mode. We will demonstrate how applying these models might help clarify ethical issues associated with distributed responsibility for software developers in some complex and interesting cases. Our analysis will show both the usefulness of and some of the shortcomings of the models proposed by Wolf, Miller, and Grodzinsky. From the analysis of the papers, we will make suggestions for a revision of the models that might be more robust when applied to cases on responsibility for downstream uses of software.

As an example, we have considered the paper "Digital health fiduciaries: protecting user privacy when sharing health data" where Chirag Arora (2019) argues that when it comes to privacy concerns surrounding health data, it is the responsibility of the digital health data controllers to take steps to protect the privacy of those whose data is being collected and stored. Arora argues for a fiduciary relationship between data subjects and the data controller. Arora's argument uses "security, anonymization, and data minimization as examples of contextualization and flexibility required to deal with privacy issues." Even though Arora brings up the WannaCry ransomware attack on the UK's National Health Service, the responsibility for the system failure is not mapped back to flaws in the

software that was infected by WannaCry, but rather to the failure to upgrade. Arora places the ethical breach at the feet of the data controller and makes no attempt to push any responsibility back to those who created the software with the flaw in it. In terms of our two models, Arora uses the Fixed History Model.

This analysis stands in contrast to the argument presented by Wolf et al. They argue that "the more sensitive the data accessed [are], the more responsibility [that] can be ascribed to the developer for [the software's] downstream use." Using this line of reasoning, Wolf et al. would likely use the Chained History Model in this case. Our analysis will compare and contrast these two opposing views.

As a second example, in the article" First steps towards an ethics of robots and artificial intelligence (RAI)," John Tasioulas (2019) investigates the problem of trying to build moral norms into RAIs. He distinguishes between RAIs that follow top down algorithms that are prescriptive and closed-rule and bottom up or stochastic algorithms that use machine learning. In the first case, the RAI is largely functional and failure to accomplish its task can be attributed back to the developer. "In ... RAIs operating on the basis of top-down algorithms that render their behavior highly predictable, the argument for attributing legal responsibility to manufacturers, owner, or users seems compelling (Tasioulas, 2019:70). Responsibility analysis can be served by the Fixed Model. In RAIs that use machine learning, the cases are more varied and complex. The author asks "...whether a good case exists for attributing legal personality to RAIs with corresponding legal rights and responsibilities …" (Tasioulas, 2019:70). The European Union and UNESCO have been grappling with this issue and it is beyond the scope of our paper to delineate the various arguments. However, Tasioulas does raise the question of traceability as particularly difficult with RAIs using bottom-up algorithms. He asks "... how do we ensure the 'traceability' of RAIs in order to be able to assign moral or legal responsibility in relation to them? Traceability involves being able to determine the causes that led an RAI to behave in the way that it did…" (Tasioulas, 2019:71). Applying the Chained Model might aid in the analysis of what Tasioulas calls one of the biggest challenges in the deployment of machine learning RAIs.

As these two examples show, the two models for downstream responsibility attribution can clarify thinking about different kinds of software. In the full paper, we will show where and why each of the two models has been used and argue whether doing do was appropriate. Additionally, we will identify cases where the choice between the models is not clear. We will also identify revisions to the models to make them more versatile.

KEYWORDS: responsibility, software developer responsibility, models of responsibility, ethical analysis.

REFERENCES

Arora, C. (2019). Digital health fiduciaries: protecting user privacy when sharing health data, Ethics Inf Technol 21: 181. DOI: 10.1007/s10676-019-09499-x.

Tasioulas, J. (2019). First steps towards an ethics of robots and artificial intelligence, Journal of Practical Ethics. 7(1), 49-83.

Wolf, M. J., Miller, K. W., & Grodzinsky, F. S. (2019). On the responsibility for uses of downstream software," Computer Ethics - Philosophical Enquiry (CEPE) Proceedings. *2019*, Article 3. DOI: 10.25884/7576-wd27, from https://digitalcommons.odu.edu/cepe_proceedings/vol2019/iss1/3.

ONCE AGAIN, WE NEED TO ASK, "WHAT HAVE WE LEARNED FROM HARD EXPERIENCE?"

William Fleischman, Jack Crawford

Villanova University (USA)

william.fleischman@villanova.edu; jcrawf15@villanova.edu

EXTENDED ABSTRACT

In this paper, we discuss the disconcerting structural similarities between the series of radiation therapy accidents caused by the Therac-25 in the 1980's and the accidents and near-accidents involving the Boeing 737 Max aircraft in 2018 and 2019. These similarities concern engineering and software design, testing, hazard analysis, documentation as well as responses to accident reports. Considering the lapse of time between the two cases and the publicity attending publication of the 2017 revision of the ACM Code of Ethics, we reflect on the role of codes of ethics in computing and make several suggestions of measures that might enhance their effectiveness.

The series of accidents caused by the Therac-25 computer-controlled radiation therapy machine is one of the most carefully studied and widely cited cases of accidents involving poorly conceived safety-critical software and deeply flawed engineering design. The Therac accidents are commonly and justifiably used as a fundamental case study in university courses in computer and engineering ethics and system safety. Although the accidents occurred more than thirty (and the design process more than forty) years ago, every aspiring computer or engineering professional should reflect on the deficiencies in software and engineering design, testing, safety analysis and documentation related to this case. In addition, they should think deeply about the ineffective and frequently dishonest responses of the device manufacturer to reports of harm to patients treated with the machine.

The currency of this case study is underscored by the striking similarities between the factors identified by Leveson and Turner (1993) in their investigation of the Therac-25 accidents and those, revealed in investigative articles in the recent press, relating to the contemporary series of accidents and near accidents involving the Boeing 737 Max aircraft.

In this paper, we begin by laying out in detail corresponding elements material to producing the harms that occurred in each of the two cases. We then reflect on the role of codes of ethics in the face of the persistence of identifiable patterns of unethical behavior and practice.

There is a common ground circumstance that links the cases of the Therac-25 and the Boeing 737 Max. Each involved the re-design of a system that, because of an engineering decision, required the introduction of new safety-critical software. Design changes were significantly driven by economic factors. Subsequent similarities extend to deficiencies in testing and documentation, failure to consider carefully the "ecology of use" of each system, and, most disturbingly, a pattern of evasion and dishonesty by company personnel in response to reports of problems.

The Therac-25, a radiation therapy machine or linac (linear accelerator) used in the treatment of cancer, was developed by Atomic Energy of Canada Limited (AECL). It was the successor to similar devices previously developed jointly by AECL with a French partner, CGR, under a collaboration agreement that had recently been discontinued. The Therac-25 had several novel features that made it an attractive investment for hospitals and cancer treatment centers. It was a single device that could deliver therapeutic doses of radiation in either electron or X-ray mode. Because of innovations in beam

technology, it was a more compact device, yet able to deliver therapy at higher energy than earlier models. A further considerable economy resulted from the decision to substitute software for costly mechanical interlocks as the means of ensuring safe operation. (Leveson & Turner 1993)

Responding to innovations by Airbus, its main competitor in the passenger plane market, Boeing sought to redesign its 737 aircraft. The re-design involved replacing the engines common to earlier 737 models by more efficient, but larger and heavier, new engines. However, the size of the new engines dictated that their mounts be moved upward and forward on the wing in order to provide safe ground clearance for take-off. (Campbell, 2019) This created a problem in aerodynamics that Boeing engineers decided to solve through changes to the software system that controls flight characteristics of the plane. The supplemental software implementing this functionality was called the Manœuvering Characteristics Augmentation System (MCAS).

In each of these cases, the complexity of the control problem and the difficulty of developing a solution in software were vastly underestimated.

For the Boeing 737 Max, as earlier for the Therac-25, there was a "frictionless" path to approval for commercialization under government regulations. All that was required was an affirmation by each of the companies that the systems were effectively identical to the earlier models that they replaced. In effect, this meant an assertion that performance characteristics of the systems were not affected by the inclusion of safety-critical software to replace or supplement existing control features.

The Therac-25 was approved for use under the FDA's pre-market notification process rather than the more rigorous and time-consuming process of pre-market approval in the early 1980's (Leveson & Turner, 1993), a time at which the argument that hardware safety features of the device could be replaced by software perfectly equivalent in function might plausibly, if mistakenly, have been advanced.

For the 737 Max, the "frictionless" path involved retaining the 737 designation (hence the name Boeing 737 Max) so that pilots already certified to fly earlier 737 models would not be required to undergo lengthy and expensive training to be recertified on a new aircraft. However, this decision implicitly involved the assumption that changes to the aircraft's flight control software were of such a minimal nature that pilots would need only a short, self-administered computer course instead of expensive classroom time and training on a flight simulator to be properly prepared to fly the new plane. (Campbell, 2019)

The perfunctory nature of testing and hazard analysis performed for the Therac-25 by AECL engineers was brought to light by the careful investigation of Leveson and Turner (1993). By means of review of proprietary documents and interviews with personnel involved in the development process, several teams of journalists (Campbell, 2019; Gates & Baker, 2019; Nicas, Kitroeff, Gelles & Glanz, 2019) have uncovered similar shortcomings in the testing and analysis of hazards associated with the 737 Max software modifications.

The Therac-25 study cited glaring deficiencies in documentation both for internal purposes and in presentation of information to operators of the device who were provided only cryptic error messages and uninformative user manuals. Leveson and Turner (1993) remark tartly, "Software specification and documentation should not be afterthoughts."

Campbell (2019) and Gates and Baker (2019) relate how the pressure to certify the new aircraft and preserve the 737 type certificate led to serious deficiencies both in regard to documents filed with the FAA (as the cognizant regulatory agency) and, critically, in information provided for the training of pilots. Most seriously, Boeing introduced, and subsequently modified, a critical software control feature, MCAS, about which there was no mention in training bulletins prepared for pilots.

The effect of these deficiencies is magnified because of the disregard they indicate for the "ecology of use" of safety- and life-critical systems. In the case of the Therac-25 the lack of concern for providing meaningful information to operators who were low-level hospital personnel led to a pattern of tolerance of numerous apparently innocent machine malfunctions that infected hospital discipline at all levels. The resulting poor culture of care was implicated in all of the radiation overdoses inflicted on patients.

By contrast, pilots who fly the 737 Max are highly trained professionals. Nonetheless, the absence of information about changes to the aircraft's flight control software left them in uncharted waters under severe time pressure when that software mistakenly started to force the nose of the plane downward. The testimony of experienced pilots to the U.S. Congress was that "the terror and tumult of such a moment would defeat many of the world's best pilots," and "I can tell you firsthand that the startle factor is real, and huge." (Laris, 2019) Although a recent article blames the accidents on inexperienced pilots and the economy airlines that hire them (Langewiesche, 2019), Boeing knew of these practices and should have taken even more care in providing for this altered ecology of use.

After the first accident, in response to questions from the attending radiation physicist as to whether the Therac-25 was capable of malfunctioning and burning his patient, AECL personnel insisted this was impossible. Rather than investigating carefully, they maintained this posture and asserted that no one had been injured by the Therac-25, as one accident followed another, until there was definite confirmation that a patient had been burned by the device.

In a similar manner, Boeing denied that its control software was implicated in the first accident even though there was circumstantial evidence from a flight the day before that the software could create the conditions for precipitating loss of the aircraft. They attempted to conceal knowledge of the flaw in their control system as they tried to repair it. However, a technical bulletin posted on the company's online portal for pilots and airlines, that made oblique reference to the conditions of the first accident without directly identifying MCAS, generated such a volume of angry demands for additional information that Boeing had to admit there was something fundamentally wrong with the aircraft and name and explain the nature of the faulty modification. (Campbell, 2019)

In one notorious episode involving the Therac-25, AECL engineers inspected an accident site, could not replicate the conditions that produced the malfunction, but speculated as to a possible cause. AECL engineered a "fix" for the hypothetical fault and then made the preposterous announcement that "analysis of the hazard rate of the new solution indicates an improvement over the old system of at least *5 orders of magnitude* [emphasis added]." (Leveson & Turner, 1993)

In its safety analysis of MCAS, Boeing identified several possible failure modes. One particular mode associated with the actual accidents was designated as "hazardous," something with the potential to cause serious or fatal injuries to a small number of people not, however, resulting in loss of the plane itself (a "catastrophic" failure.) Boeing calculated the probability of this failure as approximately one every 223 trillion hours of flight. Gates and Baker (2019) remark acidly, "In its first year in service, the MAX fleet logged 118,000 flight hours. On the basis of this analysis, Boeing downgraded the number of sensors required to confirm the condition from two to one, thus creating circumstances that increased dramatically the likelihood of this failure due to malfunction of or damage to the single sensor involved.

In cases of such breath-taking obtuseness, it is difficult to say which is the greater fault – failure to analyse correctly the true sources of danger, or the recourse to incomprehensibly large numbers to provide a misleading quantitative "fig leaf" of rationality for an unfounded assurance of safety.

Although Leveson and Turner do not comment directly on this, the fact that earlier Therac models had been collaboratively developed suggests that AECL may have been anxious to pre-empt possible competition from CGR by expediting the introduction of the Therac-25. This would explain in part the deficiencies in testing, hazard analysis and documentation they uncovered.

In the case of the Boeing 737 Max, these pressures were explicit. Boeing rushed to prevent Airbus from securing market dominance in sale of single-aisle passenger aircraft. Several authors detail the accelerated pace of releasing blueprints and pressures on all aspects of engineering, software development, and testing, all citing internal sources at Boeing. Even an experienced commercial pilot for a major airline was aware of the effects of procrustean efforts to preserve the 737 type certificate for the new aircraft. (Campbell, 2019; Gates & Baker, 2019; and Travis, 2019)

If we make the most favourable assumptions about the motivation and actions of computing professionals, comparison of the two cases suggests that at least some of the problems they encounter are rooted in the nature of their interactions with engineers and professionals from other disciplines, possibly acting in a supervisory role. The revised ACM Code of Ethics does speak of the duty to report risks, but in a document of many words, emphasis on this point is lacking and there is no explicit reference to the type of risks that arise in the interactions to which we have alluded. We believe computing professionals have certain explicit proactive responsibilities in regard to system development tasks they implement where specifications are set by other engineers and managers. **"Are you really asking me to write code to implement a potentially catastrophically dangerous manoeuvre of the aircraft on the basis of input from a single, fragile and easily compromised sensor?"** We will discuss some of the ideas suggested by our investigation.

We believe that including illustrative case studies highlighting these types of interactions, with explicit reference to real world experience, would be useful in alerting computing professionals to the pitfalls they are likely to encounter in such situations. We suggest an approach similar to that of the American Society of Civil Engineers (ASCE) whose code of ethics website includes a sidebar that features persuasive case studies based on the real experiences of practicing civil engineers. (American Society of Civil Engineers, 2017) In the present context, for example, a compact comparison of common problematic factors in the Therac and Boeing cases might be effectively presented among the case studies accompanying the ACM Code of Ethics.

KEYWORDS: System failure, safety-critical software, engineering design, codes of ethics.

REFERENCES

American Society of Civil Engineers (2017). ASCE Code of Ethics. Retrieved from https://www.asce.org/code-of-ethics/

Association for Computing Machinery (2018). ACM Code of Ethics and Professional Conduct. Retrieved from https://www.acm.org/code-of-ethics

Campbell, D. (2019, May 2). Redline: The many human errors that brought down the Boeing 737 Max. *The Verge*. Retrieved from https://www.theverge.com/2019/5/2/18518176/boeing-737-max-crash-problems-human-error-mcas-faa

Gates, D. & Baker, M. (2019, June 22). The inside story of MCAS: How Boeing's 737 MAX system gained power and lost safeguards, *Seattle Times*. Retrieved from https://www.seattletimes.com/seattle-

news/times-watchdog/the-inside-story-of-mcas-how-boeings-737-max-system-gained-power-and-lost-safeguards/

Langewiesche, W. (2019, September 18). What Really Brought Down the Boeing 737 Max? *The New York Times Magazine.* Retrieved from https://www.nytimes.com/2019/09/18/magazine/boeing-737-max-crashes.html

Laris, M. (2019, June 19). Changes to flawed Boing 737 Max were kept from pilots, DeFazio says. *The Washington Post*. Retrieved from https://www.washingtonpost.com/local/trafficandcommuting/changes-to-flawed-boeing-737-max-were-kept-from-pilots-defazio-says/2019/06/19/553522f0-92bc-11e9-aadb-74e6b2b46f6a_story.html

Leveson, N. & Turner, C. (1993). An investigation of the Therac-25 accidents. *IEEE Computer*, 26(7), 18-41.

Nicas, J., Kitroeff, N., Gelles, D., & Glanz, J. (2019, June 1). Boeing built deadly assumptions into 737 Max, blind to a late design change. *The New York Times*, Retrieved from https://www.nytimes.com/2019/06/01/business/boeing-737-max-crash.html

Travis, G. (2019, April 18). How the Boeing 737 Max disaster looks to a software developer. *IEEE Spectrum*, Retrieved from https://spectrum.ieee.org/aerospace/aviation/how-the-boeing-737-max-disaster-looks-to-a-software-developer

ORGANISATIONAL ETHICS OF BIG DATA: LESSONS LEARNED FROM PRACTICE

Maryam Nasser Al-Nuaimi

AL-Buraimi University College (Sultante of Oman)

maryam@buc.edu.om

EXTENDED ABSTRACT

The purpose of this paper is to give a synopsis of the ethical concerns of big data analytics. In information technology contexts, the implementation of data mining tools implies that an individual's activities are monitored and tracked online. This reality implies that the flow of personal information and the creation of new kinds of personal information have had enormous and extensive impacts on the controversial issues of individuals' privacy, autonomy, and security. By contrast, there is a considerable discrepancy between the collection and analysis of data on the one hand and the awareness of the ethical ramifications of big data analytics and the relevant regulations and institutional policies on the other hand. This paper sheds light on the ethical paradoxes surrounding big data analytics. More specifically, the paper raises ethical questions on big data analytics with a specific focus on higher education institutions.

The sphere of big data is a comprehensive multidisciplinary domain. For example, big data encompasses a wide range of data types and data mining applications used in areas such as artificial intelligence, electronic commerce, behaviour modelling and analysis, sustainability studies, and biomedical research (Mittelstadt & Floridi, 2016). Big data is an exponentially growing concept, but it is essentially defined as the enormous volume of data gathered via technological means. Such technological tools stockpile data at a high rate of rapidity that exceeds the pace of information processing required (Jurkiewicz, 2018). More specific designations of big data are geared towards classifying the dimensions of the information comprised in big data. These dimensions include volume, velocity, variety, and veracity (Herschel & Miori, 2017). The volume of big data is estimated to surpass 40 zettabytes by 2020, which is 300 times higher than the level in 2005 (Herschel & Miori, 2017). The information stored by big data is generated via a diverse multitude of resources, that is, posts to social media, purchase transaction records, and any information recorded by sensors and smart devices used to collect communicate metadata, to name but a few. Since data are accumulated by means of technological devices without the direction of a user, big data is considered passive data collection. Passive data collection, therefore, poses ethical concerns regarding privacy and informed consent (Maher *et al.*, 2019). In higher educational contexts, big data are closely associated with learning analytics. Learning analytics is the collection and analyses of both digital and analogue data about student behaviour to increase institutional efficiency (Rubel & Jones, 2016). In this regard, understanding the legal, ethical and social impacts of automated data generation, linkage, sharing, and exchange lags behind the technological capacities being employed (Mittelstadt & Floridi, 2016). Thus, research ventures prefer to define big data and big data analytics socially rather than technically. That is, big data could inform projects that could be conducted on a large scale to gain insights, create new forms of value, and change markets, organisations, and the relationship between citizens and government (Richards & King, 2014). This conceptual paper discusses the philosophical underpinnings of the ethical concerns in big data analytics. Accordingly, the paper raises ethical issues of big data in biomedical and health care research as well as big data ethics in higher education institutions to glean lessons learned from practice.

The ethical challenges exacerbated in big data analytics are triggered and aggravated by virtue of big data-induced hyper-networked ethical qualities. Such qualities shift moral responsibility from personal agency towards the moral accountability of those who operate big data. More specifically, big data is characterised by being (1) constantly and increasingly accumulating on an enormous scale, (2) organic (i.e., a proxy for the real world digitally much more authentically than statistically), (3) potentially global, and (4) contingent on correlations rather than causations to analyse and understand human behaviour (Zwitter, 2014). Ethical themes have emerged from meta-analyses of the big data-germane literature. For example, Mittelstadt and Floridi (2016) and Rubel and Jones (2016) have designated and emphasised some unique ethical aspects of big data analytics.

As big data brings to the fore the unforeseen correlations between strands of data and since this insight implies that the relations inferred between subjects and their utility from big data analytics are ambiguous at the time of data collection, informed consent in big data analytics presents a persistent ethical dilemma (Mittelstadt & Floridi, 2016). The problematic issue of informed consent lays stress on the controversy over the autonomy of the end user, individual consumer, or client, since data are linked from multiple sources without the permission of end users. In epidemiology and public health, clustering data together from linkable frugal secondary sources such as routine public health surveillance and bio-specimens demonstrates the difficulty of making big data analytics value-laden. Without a relevant research question and an evidently legitimate utility to public health, the generation of linkable data reservoirs poses potential risks to the personal and socioeconomic determinants of individuals' health and well-being (Salerno, Knoppers, Lee, Hlaing, & Goodman, 2017). Hence, in the utilitarianism ethical perspective, the maximum well-being and happiness outcomes of data mining are critical in ethically evaluating big data analytics. Although categorical imperatives in Kantian ethics are concerned with the rules underlying an action rather than its consequences, Kantianism conversely emphasises informed consent. On the other hand, the ethical perspective of social contract theory justifies that citizens submit rationally and willingly to the moral guidelines that govern citizens and governments. Nevertheless, the purposes of such guidelines should be expounded to the public (Salerno *et al*., 2017).

The conception and the intrinsic value of privacy are continually questioned and often compromised in big data analytics irrespective of the relative anonymisation and de-individualisation of data. While neither de-individualization nor anonymisation shields groups from being transparent, exposed, and even stigmatised, they do occlude the latent assumptions for inferring correlations between groups and behavioural patterns (Zwitter, 2014). Jurkiewicz (2018) argues that contesting group privacy would lead to substituting individual identity with collective identity. This has the potential of adversely affecting the less advantaged socioeconomic classes (Jurkiewicz, 2018). The concern of privacy in higher education contexts is defined by the necessity of the contextual integrity of the information flows as a benchmark of privacy. However, prima facia violations of contextual integrity occur when stakeholders, especially students, are not fully aware of and willing to endorse data collection, analysis and use (Jurkiewicz, 2018).

Table 1 Recorded history of big data

Year	Data size	Rate
2011	5 billion gigabytes	every two days
2013	5 billion gigabytes	every 10 min
2015	5 billion gigabytes	every 10 s

Source: Zwitter (2014, p.2)

The concept of ownership is sophisticated because it accounts for the nuances among the right to modify data, the right to benefit from the data and the rights of data redistribution. Restrictions on data access, modification, and redistribution to data analysts are necessary to preserve data integrity (Mittelstadt & Floridi, 2016). Choudhury, Fishman, McGowan, and Juengst (2014) contend that stakeholders, including subjects of learning analytics, should be able to use tracking and checking mechanisms to ensure that their data are not being modified for unacceptable purposes. In higher education institutions, students are expected to be made aware of the benefits and burdens of the learning analytics tools that are employed by instructors and management information systems (Rubel & Jones, 2016).

Like other institutions, higher education institutions adopt information systems security policies. Information systems security policies revolve around educative and preventive procedures that counteract the risks, threats, and vulnerabilities to the security of an organisation's information systems that arise because of the users' practices. In contrast, information systems security policies ignore the ethical concerns arising from big data analytics, especially in educational contexts. This state of affairs provokes ethical debates that can be encapsulated as follows:

- Do higher education institutions set ethical guidelines for justifying and approving the data analytics tools/applications used by instructors or information systems administrators to collect students' data?
- Does the official learning management system (LMS) in a higher education institution make it clear to students whether their data are collected and analysed and whether the LMS prompts students to give their informed consent?
- Are students' data and records disclosed by third parties? If so, under what condition and for which safeguards are they disclosed?
- Do educational institutions give students the right to know who has access to their records and for what purposes?
- Are higher education institutions lucid regarding preventative measures that protect students from phishing scams that can be attached with the big data enterprise?

To recap, this paper claims that fulfilling transparency requirements regarding the establishment of ethical codes and organisational policies that govern the use of big data analytics applications has become a critical mission for many higher education institutions today. The invasive nature of big data analytics makes the personal information of end users (i.e., students) vulnerable to privacy violations. Thus, it is imperative to set and clarify regulations that protect students' privacy, security, autonomy, and digital footprints in the digital age.

KEYWORDS: Big Data, analytics, ethics, privacy, consent, higher education.

REFERENCES

Choudhury, S., Fishman, J. R., McGowan, M. L., & Juengst, E. T. (2014). Big data, open science and the brain: lessons learned from genomics. *Frontiers in human neuroscience, 8*, 239. doi: 10.3389/fnhum.2014.00239.

Herschel, R., & Miori, V. M. (2017). Ethics & Big Data. *Technology in Society, 49*, 31-36. doi: https://doi.org/10.1016/j.techsoc.2017.03.003

Jurkiewicz, C. L. (2018). Big Data, Big Concerns: Ethics in the Digital Age. *Public Integrity, 20*(sup1), S46-S59. doi: 10.1080/10999922.2018.1448218

Maher, N. A., Senders, J. T., Hulsbergen, A. F. C., Lamba, N., Parker, M., Onnela, J.-P., … Broekman, M. L. D. (2019). Passive data collection and use in healthcare: A systematic review of ethical issues. *International Journal of Medical Informatics, 129*, 242-247. doi: https://doi.org/10.1016/j.ijmedinf.2019.06.015

Mittelstadt, B. D., & Floridi, L. (2016). The Ethics of Big Data: Current and Foreseeable Issues in Biomedical Contexts. *Science and Engineering Ethics, 22*(2), 303-341. doi: 10.1007/s11948-015-9652-2

Richards, N. M., & King, J. H. (2014). Big data ethics. *Wake Forest L. Rev., 49*, 393-432.

Rubel, A., & Jones, K. M. L. (2016). Student privacy in learning analytics: An information ethics perspective. *The Information Society, 32*(2), 143-159. doi: 10.1080/01972243.2016.1130502

Salerno, J., Knoppers, B. M., Lee, L. M., Hlaing, W. M., & Goodman, K. W. (2017). Ethics, big data and computing in epidemiology and public health. *Annals of Epidemiology, 27*(5), 297-301. doi: https://doi.org/10.1016/j.annepidem.2017.05.002

Zwitter, A. (2014). Big Data ethics. Big Data *& Society, 1*(2), 1-6. doi:10.1177/2053951714559253

PERCEIVED RISK AND DESIRED PROTECTION: TOWARDS COMPREHENSIVE UNDERSTANDING OF DATA SENSITIVITY

Yasunori Fukuta, Kiyoshi Murata, Yohko Orito

Meiji University (Japan), Meiji University (Japan), Ehime University (Japan)

yasufkt@meiji.ac.jp; kmurata@meiji.ac.jp; orito.yohko.mm@ehime-u.ac.jp

EXTENDED ABSTRACT

This study conducts exploratory consideration on the relationships between risks perceived by data subjects and their desired level of protection of their personal data when it is distributed among and/or used by organisations in the public and private sectors based on a quantitative survey of Japanese people. In many studies, personal data sensitivity has been measured in terms of potential damage data subjects suffer when their data is leaked or misused (Woodman et al., 1982; Ackerman et al., 1999). Therefore, certain types of personal data have been categorized as sensitive data based on the evaluation of such damage, and their handling is strictly regulated (e.g. special categories of personal data in GDPR).

The Japanese data protection act revised in 2015 defines a set of sensitive data called *yo-hairo* (extra care is required) personal data: "Special care-required personal information" in this Act means personal information comprising a principal's race, creed, social status, medical history, criminal record, fact of having suffered damage by a crime, or other descriptions etc. prescribed by cabinet order as those of which the handling requires special care so as not to cause unfair discrimination, prejudice or other disadvantages to the principal (3rd paragraph, Article 2; http://www.japaneselawtranslation.go.jp/ law/detail/?id=2781&vm=&re=) . However, some survey results have shown that ordinary Japanese have little interest in personal data in this category and there are differences between the legally defined sensitive data and the data ordinary people perceive as sensitive one (Cabinet Public Relations Office, 2015; Fukuta et al., 2017). To properly and effectively protect personal data including sensitive one, it is helpful to comprehensively understand the relationships between people's perceived risks and desired protection level related to the distribution and use of various types of personal data not only conceptually but also empirically.

As a first step towards the comprehensive understanding, a questionnaire survey of Japanese people was conducted in March 2019. The total sample size of the survey across five age groups (20s, 30s, 40s, 50s and over 60s) was 420. The questionnaire used in the survey consists of three sections. In the first section, respondents are asked to answer questions about their attributes and general attitude towards privacy and personal data. The next section measures respondents' perceived risks relating to distribution of the four types of personal data: political orientation, financial status, health status, and consumption behaviour. Respondents were required to evaluate the two aspects of the perceived risks. One is their subjective probability (0-100%) of occurrence (SPO) of the following five situations: public surveillance, unjust discrimination/prejudice, commercial use, embarrassing experience and crime damage. The other is their perceived magnitude of the damage (PMD) of each SPO, which is measured by 6-points rating scale (from 0=don't damage me at all to 5=damage me strongly). An average value of SPOs of a respondent and a total score of PMDs of him/her were regarded as the scores of his/her perceived risk concerning each data type. The desired protection level (DPL) of each type of personal data were measured in the last section of the questionnaire. Respondents were asked to answer their desired level of rigorousness of information management, legal protection and

agreement on treatment of the data using 6-points rating scale (from 0=don't desire at all to 5=desire strongly). The sum of these three scores was regarded as the score of DPL of a respondent.

The result of the statistical analyses of the survey data indicates the followings:

a. Table 1 shows the average value of SPOs, PMDs and DPLs for each data type. As it is difficult to assume that the SPO is normally distributed, median of SPOs is added in the list of sample statistics. As shown in the table 1, the personal data regarded as sensitive data was not necessarily perceived as data that could cause severe damage, and vice versa. For example, respondents tended to feel low level of SPOs, PMDs and DPLs toward personal data relating to their political orientation, though it is categorised as one of sensitive data in the Japanese revised personal data protection act. On the other hand, the average values and median of all variables of personal data concerning their financial state were higher than others despite it is not recognised as sensitive data.

Table1 Fundamental statistics of each variable

	Political Orientation	Health Status	Financial State	Consumption Behaviour
<SPO (N=420, Range:0-100)>				
Mean	25.84	29.10	31.80	28.56
Median	16	20	24	20
<PMD (N=420, Range:0-25, Total average=17.10)>				
Mean	16.72	17.06	17.58	17.03
S.D.	6.33	6.11	6.11	6.07
<DPL(N=420, Range:0-15, Total average=11.37>				
Mean	10.36	11.79	12.44	10.89n
S.D.	3.62	3.10	3.06	3.32

b. Table 2 shows the result of dummy-regression analysis conducted to explore causal relationship between SPOs and DPLs by each information type. As the distribution of SPO was distorted, SPO was categorized into four categories (below 5%, 5-20%, 20-50%, more than 50%) based on its quartile points and converted into three dummy variables. The result indicates two findings. As all regression coefficients of dummy variable 3 are significant at least 5% significance level, the respondents whose SPO are over 50% tend to desire a significantly higher level of protection than people who perceived SPO below it. Moreover, the result that all of significant regression coefficients have positive value indicates there is a general positive relationship between SPOs and RPLs.

c. To consider relationship between PMDs and DPLs, regression analysis controlling sex variable as a confounding factor was applied. Table 3 shows the results that, for all data types, PMD has a positive and significant effect on DPL at 0.1% significance level. According to the adjusted multi-correlation coefficients of each regression model, the effect of PMD on DPL observed in the case of health-relating and financial personal data is stronger than those observed in the cases of political and consumption-relating data. These indicates influence of PMD on DPL depends on the data type, and the higher the personal data's PMD is, the stronger the influence of PMD on DPL is. Furthermore, in the presentation, we will mention comparison between influence of SPO and PMD on DPL and discuss effect of combined score of SPO and PMD on DPL as well. These findings extracted from a quantitative survey will be complemented by the result of unstructured interviews.

Table 2 Results of dummy-regression analysis. (x=SPO)

Variables	Coefficient (Std)	P-value	VIF
<Political orientation: AdjR^2=0.069>			
Intercept	9.082	0.000	-
Dummy Vari.1 (1)	1.341(0.165)	0.003**	1.388
Dummy Vari.2	-0.074(-0.010)	0.875	1.845
Dummy Vari.3	1.756(0.190)	0.001**	1.452
Sex (2)	0.014(0.002)	0.968	1.025
<Health Status: AdjR^2=0.118>			
Intercept	10.359	0.000	-
Dummy Vari.1	0.396(0.056)	0.318	1.471
Dummy Vari.2	1.209(0.195)	0.003**	1.972
Dummy Vari.3	0.929(0.126)	0.024*	1.473
Sex	0.682(0.110)	0.019*	1.036
<Financial Status: AdjR^2=0.117>			
Intercept	10.745	0.000	-
Dummy Vari.1	1.031(0.136)	0.012*	1.384
Dummy Vari.2	0.516(0.085)	0.178	1.860
Dummy Vari.3	1.316(0.187)	0.001**	1.453
Sex	0.565(0.093)	0.049*	1.047
<Consumption Behaviour: AdjR^2=0.081>			
Intercept	9.397	0.000	-
Dummy Vari.1	0.954(0.117)	0.032*	1.338
Dummy Vari.2	0.341(0.051)	0.401	1.704
Dummy Vari.3	1.486(0.180)	0.001**	1.366
Sex	0.544(0.082)	0.086	1.034

(1) In this dummy conversion, based on the quartile point of the distribution of SPO, we assigned "1" to the response in which SPO was more than 5% in the dummy variable 1. Likewise, in the dummy variable 2, score 1 was allocated in the case that SPO was more than 20%. And in the dummy variable 3, score 1 was assigned to the answer SPO was more than 50%.

(2) Variable of sex was added as a confounding factor in this regression model. The dummy conversion allocated "1" to the female respondents.

* significance at 5% significant level, ** significance at 1% significant level.

Table 3 Results of regression analysis (x=PMD)

Variables	Coefficient (Std)	P-value	VIF
<Political orientation: AdjR^2=0.156>			
Intercept	6.627	0.000	
PMD	0.232(0.406)	0.000***	1.040
Sex (1)	-0.293(-0.041)	0.377	1.040
<Health Status: AdjR^2=0.193>			
Intercept	7.887	0.000	
PMD	0.212(0.418)	0.000***	1.033
Sex	0.569(0.092)	0.040*	1.033
<Financial Status: AdjR^2=0.202>			
Intercept	8.403	0.000	
PMD	0.218(0.435)	0.000***	1.038
Sex	0.433(0.071)	0.111	1.038
<Consumption Behaviour: AdjR^2=0.145>			
Intercept	7.238	0.000	
PMD	0.202(0.369)	0.000***	1.031
Sex	0.441(0.067)	0.147	1.031

(1) Variable of sex was added as a confounding factor in this regression model. The dummy conversion allocated "1" to the female respondents.

* significance at 5% significant level, *** significance at 0.1% significant level.

KEYWORDS: Personal data, Sensitive data, Perceived risks, Desired protection level.

REFERENCES

Ackerman, M.S., L.F. Cranor and J. Reagle (1999). Privacy in e-commerce, *Proceedings of the 1st ACM Conference on Electronic Commerce*, pp.1-8.

Fukuta, Y., K. Murata, A.A. Adams, Y. Orito, A.M. Lara Palma (2017). Personal data sensitivity in Japan: An exploratory study, *Orbit Journal*, Vol. 2. DOI: 10.29297/orbit.v1i2.40.

Cabinet Public Relations Office. Retrieved from https://survey.gov-online.go.jp/tokubetu/h27/h27-kojin.pdf (confirmed in October, 2019, a material in Japanese)

Personal data Protection Commission Japan. Retrieved from https://www.ppc.go.jp/files/pdf/Act_ on_the_Protection_of_Personal_Information.pdf (confirmed in October, 2019)

Woodman, R.W., D.C. Ganster, J. Adams, M.K. McCuddy, P.D.Tolchinsky and H. Fromkin (1982). A survey of employee perceptions of information privacy in organizations, *The Academy of Management Journal*, 25(3), pp.647-663.

PRIVACY DISRUPTIONS ARISING FROM THE USE OF BRAIN-COMPUTER INTERFACES

Kirsten Wahlstrom, Helen Ashman, Sara Wilford

University of South Australia (Australia), De Montfort University

Kirsten.Wahlstrom@unisa.edu.au; Helen.Ashman@unisa.edu.au; Sara.Wilford@dmu.ac.uk

EXTENDED ABSTRACT

The research described in this paper investigated privacy with respect to a group of emerging technologies called Brain-Computer Interfaces (BCIs). BCIs facilitate direct communication between a user's brain and a computing device in order to control external devices, such as wheelchairs (Kobayashi & Nakagawa, 2018), prostheses (Murphy et al., 2017), and avatars in games (Kerous, Skola, & Liarokapis, 2018). Research into the effects of BCIs on privacy has been suggested by various authors (Clausen, 2014; Grübler et al., 2014; Jebari & Hansson, 2013; Klein & Nam, 2016; Li, Ding, & Conti, 2015; Moreno, 2003) and the research described in this paper responds to these suggestions.

The research has two motivations. First, BCIs bring new information sharing practices into being, placing neural signals simultaneously within an intimate domain (a person's brain) and a systems domain. Thoughts (ie neural signals) may be influenced by the presence of others, by monitoring, and by surveillance. On the other hand, people think freely when in privacy and in doing so, develop identity, preferences, integrity, independence, and so on. Therefore, it may be that bringing neural signals to a systems domain is problematic. Second, the elicitation of neural data for interpretive purposes in a consumer context is unprecedented and therefore there is a high risk it may occur in a policy and design vacuum. These motivations give the research its purposes, which are firstly to establish whether BCIs disrupt privacy and secondly, if privacy is disrupted, to explain how so. Expressed as research questions, these purposes are (a) do BCIs disrupt privacy? and (b) If so, how?

The literature on privacy encompasses diverse concepts and themes (Tavani, 2008), which will be outlined in the full paper. Given this diversity, privacy research projects that are premised on clearly stated foundation concepts of privacy are more readily understood, interpreted, and reviewed. Literature on the nature of privacy will be explored in the full paper in order to establish the foundation privacy concept. Thus, the research is premised upon prior research on privacy and its findings expand the field.

Privacy is important because it enables such things as autonomy and independence, identity and integrity, interpersonal relationships, safety and security, trust, and so on (Nissenbaum (2004, p146) provides an overview). According to the Oxford English Dictionary Online, privacy is "The state or condition of being alone, undisturbed, or free from public attention, as a matter of choice or right; seclusion; freedom from interference or intrusion."

This definition provides a starting point for a more detailed and nuanced concept of privacy. This concept is developed through the analysis of prior work on the meaning of privacy and it is that privacy and its social contexts are understood to be mutually pliable, taking diverse forms in response to emergent and influential factors; in other words, privacy is both context dependent and context forming.

One such emergent and influential factor may be the use of BCIs, which may shape and re-shape the norms that give rise to privacy expectations, experiences, and practices. We follow Solove's example

(Solove, 2006) in referring to the shaping and re-shaping of privacy norms as *privacy disruption*. Whether privacy disruptions carry intrinsic positive and negative connotations is not considered in this research, however, the well-documented tension between privacy and security is acknowledged.

This foundation privacy concept is the foundation for this project's research methods and therefore, it also circumscribes the findings. In extending our research from the foundation privacy concept, we apply a novel method informed by Habermas's discourse ethics (Habermas, 1985a, 1985b). This method uses a pre-selection survey to recruit participants with diverse privacy attitudes into focus groups. Each focus group participated in a discussion on privacy norms; next, individual participants used a BCI in order to learn the capabilities of the technology; participants then established their views on BCIs and privacy; finally, the focus groups were re-formed and the discussion was held again. This method engaged participants in communicative actions on privacy, inviting them to articulate conceptions of privacy that arose in lifeworld experiences. The qualitative data collected in the focus group discussions were analysed using NVivo 12.

This method was triangulated with a factorial vignette survey implementing a contextual integrity approach to identifying privacy disruption (Nissenbaum, 2009). This survey was conducted after participants had used the BCI and prior to the second focus group discussion. Thus, it served the purpose of supporting participants in establishing their views on privacy with respect to BCIs. Quantitative data collected with the factorial vignette survey were analysed with an ordinal logistic regression. Both methods found that BCIs disrupt privacy in contexts in which agency, fairness, self-determination, autonomy, justice, and power are also prominent.

KEYWORDS: Privacy, Brain-Computer Interfaces, Contextual Integrity.

REFERENCES

Clausen, J. (2014). [Ethical Implications of Brain–Computer Interfacing]. In J. Clausen & N. Levy (Eds.), *Handbook of Neuroethics* (pp. 699-704). Dordrecht, Netherlands: Springer Science+Business Media.

Grübler, G., Al-Khodairy, A., Leeb, R., Pisotta, I., Riccio, A., Rohm, M., & Hildt, E. (2014). Psychosocial and Ethical Aspects in Non-Invasive EEG-Based BCI Research—A Survey Among BCI Users and BCI Professionals. Neuroethics, 7(1), 29-41.

Habermas, J. (1985a). *The Theory of Communicative Action, Volume 1: Reason and the Rationalization of Society*. Cambridge, UK: Polity Press.

Habermas, J. (1985b). *The Theory of Communicative Action, Volume 2: Lifeworld and System: A Critique of Functionalist Reason*. Cambridge, UK: Polity Press.

Jebari, K., & Hansson, S.-O. (2013). European Public Deliberation on Brain Machine Interface Technology: Five Convergence Seminars. Science and Engineering Ethics, 19(3), 1071-1086.

Kerous, B., Skola, F., & Liarokapis, F. (2018). EEG-based BCI and video games: a progress report. Virtual Reality, 22(2), 119-135.

Klein, E., & Nam, C. S. (2016). Neuroethics and brain-computer interfaces (BCIs). Brain-computer interfaces, 3(3), 123-125.

Kobayashi, N., & Nakagawa, M. (2018). BCI-based control of electric wheelchair using fractal characteristics of EEG. IEEJ Transactions on Electrical and Electronic Engineering, 13(12), 1795-1803.

Li, Q., Ding, D., & Conti, M. (2015). Brain-Computer Interface applications: Security and privacy challenges. Paper presented at the 2015 IEEE Conference on Communications and Network Security (CNS).

Moreno, J. (2003). Neuroethics: an agenda for neuroscience and society. Nature Reviews Neuroscience, 4(2), 149-153.

Murphy, D., Bai, O., Gorgey, A., Fox, J., Lovegreen, W., Burkhardt, B., . . . Fei, D.-Y. (2017). Electroencephalogram-Based Brain-Computer Interface and Lower-Limb Prosthesis Control: A Case Study. Frontiers in neurology, 8, 696-696.

Nissenbaum, H. (2004). Privacy as Contextual Integrity. Washington Law Review, 79(1), 119-157.

Nissenbaum, H. (2009). *Privacy in Context: Technology, Policy, and the Integrity of Social Life*. Stanford, US: Stanford Law Books.

Solove, D. (2006). A Taxonomy of Privacy. University of Pennsylvania Law Review, 154(3), 477-560.

Tavani, H. (2008). [Informational Privacy: Concepts, Theories, and Controversies] *The Handbook of Information and Computer Ethics* (pp. 131-164): John Wiley & Sons, Inc.

RESPONSIBILITY IN THE AGE OF IRRESPONSIBLE SPEECH

Benjamin Mitchell, William Fleischman

Villanova University (USA), Villanova University (USA)

benjamin.r.mitchell@villanova.edu; william.fleischman@villanova.edu

EXTENDED ABSTRACT

We discuss the impact of language on some ethical problems surrounding the interactions of technology and society. We focus on problems of careless and irresponsible speech in the contexts of artificial intelligence and social media. As these areas are central to modern public discourse, the inappropriate use of language by computer professionals in these contexts has the potential for serious harm.

The language used to express a concept is important, particularly when the concept is a new one. This is by no means a novel insight: Joseph Weizenbaum commented on this in the context of computing many years ago (Weizenbaum, 1972). Current public discourse suggests that a reminder and an update may be needed. Particularly when we attempt to re-purpose an already existing word into a new context, there is always some conceptual "bleed-through." Connotations of the original usage are ascribed to the new even when they are not truly warranted.

Creating new words from scratch is difficult, and many attempts to get such terms into circulation fail. It is therefore perhaps unsurprising that the field of computing has a long history of simply borrowing conceptually linked terms and re-purposing them, rather than attempting to define new words. Even the term "computer" originally referred to a human who performed mathematical calculations as a career. But as convenient as repurposing existing terms is, there are clear hazards to doing so, and it must be the responsibility of computing professionals to ensure that the terms we use are not mis-interpreted by those who lack the background to directly understand their intended use.

Sometimes this careless use of language is unintentional, but frequently it appears to be done with malice aforethought. The dominant mode of political rhetoric in our society, for example, seems to revolve around the idea that perception is more important than truth, and that carefully selected terminology can be used to appeal to the baser instincts of the political base while still maintaining some façade of impartiality. To take one example from a story currently dominating the news in the United States, informed reporters, government officials, and casual observers all commonly repeat the mantra "We've seen the transcript [of the phone call between the presidents of the U.S. and Ukraine that may result in articles of impeachment entered against the American president] …" In fact, there are very few individuals who have actually seen the full, accurate transcript of that call, because its potentially incendiary nature led to the "reconstructed transcript" being quickly locked down on a server in the White House's most classified computer system. But by now, everyone in the public "understands" that the transcript of that call is a matter of common knowledge. This imprecision is convenient for those who wish to dismiss the importance of the conversation since at least one national security individual, who listened in on the call as a matter of his official duties, has openly criticized the omission of crucial words and phrases in the publicly disclosed "transcript." (Barnes, J., Fandos, N., & Hakim, D., 2019)

Whether inadvertent, reflexive, or calculated, imprecise or careless speech can have serious consequences and influence the thoughts and actions of individuals and collectives. In this paper, we

consider the dangers of such speech in two distinct contexts: First, in public understanding of the capabilities and limitations of machine learning and, more generally, artificial intelligence; and in the ways in which careless and irresponsible speech by prominent executives of social media companies can undercut responsible behavior by computing and information professionals and frustrate efforts to find sensible measures to regulate the practices of social media platforms.

There is a complex interplay between science, science fiction, and public perception. Artificial Intelligence (AI) has always been deeply entangled in science fictional narratives. This manifests in many ways, and affects both researchers and members of the public at large. The result is that many people's reasoning about the world is based on a mythologized version of AI that can lead to dangerous conclusions.

AI has a long history of underestimating the difficulty of its core problems. In one memorable anecdote from the founding era of the field, several prominent computer scientists estimated that ***programming a computer system to replicate all the important functionality of a human mind might take several graduate students as much as a few months to accomplish.*** Some seventy years later, we have yet to even come close. This has not stopped popular portrayals of AI from ascribing human-like behaviors and capabilities to these systems. HAL 9000 from *2001: A Space Odyssey*, Skynet from *Terminator* and the eponymous cute robot from WALL-E are iconic examples. Whether implacable foe or compassionate helper, these systems are presented as being "not so different from you and me." Perhaps they have certain affective deficiencies, but nothing outside the range of behaviors displayed by actual humans.

These narratives paint a highly misleading picture of the capabilities of real-world AI systems. In actuality, all "Artificial Intelligence" systems to date are just purpose-built software tools designed to automate specific and narrowly defined processes. An "AI" has less in common with a human, and more in common with a kitchen knife; both are useful tools for assisting a human to get something done faster and better, but neither has any "agency" of its own. We give human names to these systems (Eliza, Siri, Alexa, Watson, etc.), although they are no more 'human' than a toaster. We use words that imply human-like thought processes (attention, understanding, belief, etc.) as labels for simple mathematical equations and algorithms that have only loose conceptual ties to the conventional meaning of the terms. Once these anthropomorphic characterizations of "AI" are internalized by the public, sweeping extrapolations of these tools' potential are almost inevitable, resulting in misplaced trust in the capabilities of such systems.

The flip side of this exaggerated conception of the power of "AI" is something that perhaps deserves the name "silicomorphization," the reductive view of human intelligence based on the conflation of human intelligence with what can be computed. Naturally, in any comparison of capabilities based on this view, humanity comes off rather badly. The result is a systematic denigration of the reach and richness of human intelligence and the robustness of human judgment. This devaluation of human capacity seems particularly harmful when taken as received wisdom concerning the relations between humans and machines, and the future of humanity itself.

Helen Nissenbaum (1994) urged the adoption of a robust standard of accountability for computing professionals. But how can accountability survive in an atmosphere in which Mark Zuckerberg can deny and distance himself, through evasive and disingenuous speech, from one scandal after another – the misappropriation of user data by Cambridge Analytica, the dissemination of false information in its newsfeed, strange policies regarding whether political advertisements, a lucrative source of revenue to Facebook, may contain verifiably false information. If ignorance, or the pretense of ignorance, is an effective defense for the well-placed, why should it not be equally available to the

subordinate? (Lee, 2018) We'll say more about this and explore in depth connections with the section that follows in the full paper.

In his profound and prophetic paper, "On the Impact of the Computer on Society," Joseph Weizenbaum exhorts us, as computer professionals, to recognize that "[t]he nonprofessional has little choice but to make his attributions to computers on the basis of the propaganda emanating from the computer community and amplified by the press. The computer professional therefore has an enormously important responsibility to be modest in his claims." (Weizenbaum, 1972)

The mid-20th century judgment of Friedrich Dürrenmatt seems also uncannily pertinent – with particular application to the evasions, rationalizations and outright dishonesty that attend technological failures – to our moment in history: "In the Punch-and-Judy show of our century … there are no more guilty and also, no responsible men. It is always, 'We couldn't help it' and 'We didn't really want that to happen.' And, indeed, things happen without anyone in particular being responsible for them. … That is our misfortune, but not our guilt… Comedy alone is suitable for us." (Dürrenmatt, 1964) The comedy, alas, is often of a rather mordant nature (in which we are the bitten.)

KEYWORDS: responsibility, language, artificial intelligence, machine learning, social media.

REFERENCES

Barnes, J., Fandos, N. & Hakim, D. (2019). White House Ukraine expert sought to correct transcript of Trump call. The New York Times, October, 2019. Retrieved from https://www.nytimes.com/2019/10/29/us/politics/alexander-vindman-trump-ukraine.html

Dürrenmatt, F. (1964, at 31) *Problems of the Theatre*, translated by Gerhard Nellhaus. Grove Press, New York.

Lee, D. (2018), Mark Zuckerberg, missing in inaction, BBC News, Retrieved from https://www.bbc.com/news/technology-46231284

Nissenbaum, H. (1994), Computing and accountability. *Communications of the ACM*, vol. 37, no. 1, pp. 72-80.

Weizenbaum, J. (1972), On the impact of the computer on society: How does one insult a machine? *Science*, vol. 176, no. 4035, pp. 609-614.

THE 'SELFISH VISION'

Peter Vistisen, Thessa Jensen

Aalborg University (Denmark)

vistisen@hum.aau.dk; thessa@hum.aau.dk

EXTENDED ABSTRACT

This paper seeks to investigate the ethical implications of applying so-called 'vision videos' as tools for technology speculating. Vision videos a genre of filmic representations, which utilises diegetic prototypes – prototypes which are simulated inside a narrative context, and which might not (yet) exist or be feasible in praxis (Kirby 2010). Through the film medium, researchers as well as corporations can demonstrate to a larger audience a technology vision's need, benevolence, and viability, with a major rhetorical advantage over true prototypes: in the diegesis of storytelling these technologies are portrayed as already implemented, and already in use by people. In this regard, vision videos, and their use of diegetic prototypes, are related to the emerging field within human-computer interaction and design research of 'design fiction', which uses narrative storytelling to establish discursive spaces to critically reflect upon technological practices (Markussen & Knutz 2013). The main difference, between vision videos and design fictions is how the discursive space is framed; where the former tends to be strategically oriented, the latter has a critical focus.

Vision videos have a history of being either internally oriented (an example is the 'Apple Knowledge Navigator' vision video from 1989), and often only reached the public as re-appropriated marketing material for a company's ethos. In the latest decade, with rise of social media and shareable platforms, externally oriented uses of vision videos have emerged from large international companies as diverse as Google, Microsoft, Jaguar Land Rover, and IKEA.

Previously, we have seen vision videos have a potential in strategic concept development, and how such strategic uses also increase the need of engaging in an explicit ethical contract with the users receiving the vision video (Vistisen & Bolvig 2017). Here, an ethical contract is understood as a clear communication from the creator of the vision video of what kind of discourse they wish to engage in regarding the proposed technology vision, and how they will react and respond to stake-holders' feedback (be it internal or external).

In these contributions, the potential pitfalls of releasing vision videos and not actively listening or engaging in a dialogue with the participatory culture forming around the vision videos online has been discussed (Vistisen & Bolvig 2017, Vistisen & Jensen 2018a, Jensen & Vistisen 2017). But what happens when something being intended as an internal (and perhaps speculative) vision reaches an external crowd unintentionally, and without any explicit guidance in the vision video for what kind of discourse it should be understood within?

To investigate this question, we choose a leaked vision video from Google, titled 'The Selfish Ledger', from 2018, as the case for this study. Google's *The Selfish Ledger* is an interesting example of this potential pitfall of not addressing or explicitly preparing speculative design fictions for the public discourse it potentially can create, and thus also distort the potential prolific debate regarding the desirability of emerging technological possibilities.

The main theme in *The Selfish Ledger* shows how the collection of data can enhance the user's life. The name, *Selfish Ledger*, is inspired by Richard Dawkins' book *The Selfish Gene* (Dawkins 1976). Like

the gene, the ledger autonomously collects needed data in a ledger, which follows one particular user from birth to death and beyond. The data continues to exist and inform the Ledger's future data collection. Streamlining the data collection, the Ledger proposes to use Google's machine learning algorithms to collect missing data autonomously. The example given in the video is about data on a particular user's weight. In the vision video, to get these data, the Ledger orders a weight, considering the particular user's taste. All is done independently of the user.

As such, the vision for the ledger system is portrayed as more than a data collecting tool. It is a marketing device, which becomes clear with the second The Verge article. Here, the author cites various patent applications for parts of the ledger system and reveals that though the vision video is itself a diegetic prototype in a narrative scenario Google holds all the patents to actually realise the vision (Savov, 2018 May 19).

The vision video does not explicitly reveal itself as speculation – neither in its rhetoric nor its aesthetics. Rather, it portrays itself in the genre of 'traditional'/'real' product videos, showing the technology from its finest, not discussing or even addressing the potential concerns or edge cases which may manifest themselves in another 'what if' scenario.

Analyzing the reaction to the video, several points need to be addressed. First of all, it is a leaked video, reaching the press after being shown and labeled 'for internal use' by Google (Svavov, 2018 May 17). However, a later article explains about Google having applied for a patent on the technology needed for the ledger (Savov, 2018 May 19), seemingly undermining the 'speculativenes' of the video, leaving the public to trust the benevolent intentions of a multinational technology company.

The framing provided by the original and subsequent article in the Verge would seem to open up for a discourse on the idea of the Ledger. Our preliminary findings show a very divided and short-lived discussion among users reacting to both the video and the patents. Either, the commentators are pro-Google or con-Google. There is no middle ground. The discursive space is narrowly focusing on Google as a corporation, only dealing with the ethical challenges of the Ledger as side notes.

This is contrasted by the comment section on the Youtube video of the leaked vision video ("Google's The Selfish Ledger (leaked internal video)—YouTube," 2018 May 17). Here, the discussion is more nuanced. Some commentators engage in longer arguments, keeping it mostly civil. Also, the discussion is still ongoing, while the comment sections on the Verge articles closed down a few days after publication.

The question remains if the 'leak' actually ends with fulfilling the ideal of speculative design – to create a space for discourse. If it does, this will probably not happen on the premises Google would have preferred. In fact, Google released a press statement regarding the leak, addressing the video as 'speculation' and not to be taken literal (Svavov, 2018 May 17). This rebuttal was largely ignored by both the concurrent follow up articles from the Verge and other tech news outlets as well as in the online comments made from users. What does this teach us about how to create not just critical design fictions, but also how corporate visions must be prepared for reaching a public participatory culture, and thus engage in a mutually beneficial discourse on the proposed 'what if' question of the design fiction?

A further analysis of the online discussions and comments found on the different platforms will be conducted as part of this paper. This analysis is conducted through sampling the comments from the Verge's journalistic articles regarding the vision video, as well as the comments made on the Youtube video. We analyse these comments through how they express ethical concerns from the users, and how this discourse can be compared to other corporate vision videos with a more explicit ethical contract on how the discursive space should be framed. With the preliminary findings in mind, the

paper will examine how different platforms and their functionality provide for opening up for or closing down on possible discussions and discourses about the Ledger.

We point to a series of insights for, how the ethical contract, between the producer of vision videos and the users (internal as well as external) can be formed, to ensure the discursive spaces actually enable reflections on the potentials, and pitfalls of the technology vision – both from an internal and external audience, as well as from an intended and unintended one.

If you avoid making the ethical contract of design fiction explicit, you are bound to get misinterpretations. Sterling's original words: how do we let the user 'go' again? (Sterling 2009) should be seen as a wake-up call for producers of speculative fiction. Not taking Sterling's words into account, creates various secondary and tertiary discourses differing completely from the original 'what if' question of the vision video. This is due to the viral effects of participatory culture enforced by the lack of pre-emptive preparation of the vision video's role as a piece of speculative design. In our case, there even is a significant ambiguity regarding the sender and the framing of the design fiction.

Further differences can be found in design fictions made by design researchers, independent design consultancies, and individual persons on one side, and institutions, such as governmental organs and private corporations, on the other side (Vistisen & Jensen, 2018b). The first group has an easier task of addressing the speculative nature of their creations, thus making the need for explicitly stating the ethical contract of the vision video less apparent. However, when the speculative 'what if' is raised from an institution, we are immediately forced to address the following questions: are you actually making this? When will you be doing this? As well as: why do you make this scenario, if you are not actively pursuing the idea?

The ethical contract of institutional design fiction is just as much about giving the receiver the framing to 'read' the speculative 'what if' as about inviting for debate and critical discourse – rather than being a piece of marketing for an 'upcoming product or service'.

Depending on the nature of the speculation, this needs to be explicitly done to avoid the trap of distorting discourse rather than promoting it – from discreetly stating the video as a 'virtual prototype (who outside of design research would know what that is?), to more actively asking the 'what if' question in the vision video itself, and perhaps even discuss the ambivalences of the product or service proposed.

KEYWORDS: Google Selfish Ledger; design fiction; vision video; data collection; ethical contract.

REFERENCES

Dawkins, R. (1976). *The Selfish Gene*. Oxford University Press.

Google's The Selfish Ledger (leaked internal video)—YouTube. (2018, May 17). Retrieved October 25, 2019, from https://www.youtube.com/watch?v=QDVVo14A_fo

Jensen, T. & Vistisen, P. (2017). Ethical Design Fiction: Between storytelling and world building. In *ETHICOMP 2017 Conference Proceedings: Values in Emerging Science and Technology* (2 udg., Bind 1). Ethicomp https://doi.org/10.29297/orbit.v1i2.56

Markussen, T., & Knutz, E. (2013, September). The poetics of design fiction. In *Proceedings of the 6th International Conference on Designing Pleasurable Products and Interfaces* (pp. 231-240). ACM.

Savov, V. (2018, May 17). Google's Selfish Ledger is an unsettling vision of Silicon Valley social engineering. Retrieved October 25, 2019, from *The Verge* website: https://www.theverge.com/2018/5/17/17344250/google-x-selfish-ledger-video-data-privacy

Savov, V. (2018, May 19). Google's Selfish Ledger ideas can also be found in its patent applications. Retrieved October 25, 2019, from *The Verge* website: https://www.theverge.com/2018/5/19/17246152/google-selfish-ledger-patent-applications

Sterling, B. (2009). Design fiction. *interactions*, 16(3), 20-24.

Vistisen, P. & Jensen, T. (2018a). The Ethical Contract of Using Online Participation from Vision Videos in Design. In *Proceedings of the 5th Participatory Innovation Conference, PINC-2018* (s. 310-318)

Vistisen, P. & Jensen, T. (2018b). Designers as fans: bottom-up online explorations of new technology concepts as a genre of design fan fictions. *First Monday*, 23(12). https://doi.org/10.5210/fm.v23i12.9298

Vistisen, P. & Poulsen, S. B. (2017). Return of the Vision Video: Can corporate vision videos serve as setting for participation? *Nordic Design Research (NORDES)*, (7).

THE ANTICIPATORY STANCE IN SMART SYSTEMS AND IN THE SMART SOCIETY

Sabine Thuermel

Technische Universität München (Germany)

sabine@thuermel.de

EXTENDED ABSTRACT

"Smart" means "operating by automation" (Merriam-Webster, definition 7b) when used to describe the smart systems currently under development. They aim at regulating and controlling resources by means of autonomous IT systems based on predefined objectives. The goal is identical for the smart home, for smart energy systems, for the smart mobility infrastructure, and for the smart medicine environment supervising the chronically ill. All these smart systems are intended to anticipate the needs of their users and act accordingly.

Smart can also mean "mentally alert" (Merriam-Webster, definition 4a) and that is what a prudent society should be when pondering which smart systems to develop and to use. Guiding the process of responsible invention and deployment of AI the following core ethical principles are proposed by Floridi and Clowls (2019) based on the latest most prominent proposals in literature, i.e.:

- Beneficence: promoting well-being, preserving dignity, and sustaining the planet
- Non-maleficence: privacy, security and 'capability caution'
- Autonomy: the power to decide (to decide)
- Justice: promoting prosperity, preserving solidarity, avoiding unfairness
- Explicability: enabling the other principles through intelligibility and accountability.

These principles may well serve the smart society when anticipating a future where AI plays a predominant role. The aim of such an anticipatory process is "preparing for the unexpected in the world as we know it" (Nordmann 2014). Poli and collegues distinguish between the past-oriented forecast methods identifying macro trends as the Kondratieff waves, the future-oriented foresight exploring possible scenarios, and the present-oriented anticipation, where "innovation is about creating (anticipating) changes in the context of shared ideas of the future (including preferred futures)" (Poli&Valerio 2019, p.5). In Poli's and his collegues' view on the project anticipation webside (Poli et al. 2019) forecast and foresight are both predictive (the latter within scenarios) and predicative, which equates with deterministic for them. Both perspectives take structures into view: Forecast focuses on a closed system and foresight on a semi-closed system. Anticipation is characterized as non-predictive and impredicative taking open systems and their functionality into view. They contrast a reactionary stance taking past events into account to anticipation described as a forward looking activity.

In current smart systems we find both predictive and prescriptive algorithms for the engineering of the not yet. The purpose of these algorithms is on providing knowledge under conditions of uncertainty in order "to know ahead and act before" intending to streamline processes towards enhanced efficiency. These systems combine data gained in the past for influencing the immediate behaviour in sociotechnical systems. The focus is on anticipating and forming the near future at the

same time. Thus, we need a slightly different definition of anticipation to characterize the anticipatory stance of smart systems. In order to provide such a concept the chapters “prediction and anticipatory action” and “prescription und anticipatory measures” lay the base. Anticipation in smart systems will then be contrasted with anticipation in the smart society by giving particular attention to “imaginaries and the art of anticipation”.

Let us now take a closer look at smart predictive systems and the anticipatory action(s) they provide: In our fast-paced world interest is growing in near-term decision support and especially in the automation of tactical decisions. Big data approaches promise assistance when decisions under uncertainty have to be taken. Big data analytics aims at detecting hidden patterns. Recommendations for action are generated nearly in real-time. When such proposals are put into action new data are produced and a new round of analysis and action can start. Applications are found in the computational sciences, i.e. in the scientific approach to use advanced computing capabilities to understand and solve complex problems in the natural sciences, in medicine, pharmacy and engineering. A second focus is in the behavioural sciences, i.e. in behavioural psychology and in behavioural economics. The scientific application of algorithmic procedures leads to a second feed-back loop. It runs from the sciences to the analytic methods and back to the sciences. In the centre of these two feed-back loops lie big data analytics methods, i.e. the predictive algorithms. Mostly deep learning algorithms are deployed. The goal is to shape the (near) future by these procedures either indirectly by influencing human decisions and actions or directly by automatizing decisions and actions.

“Actionable data”, the output of the predictive algorithms, intend to provide a link between “knowing that” and “knowing how”. In drug discovery systems they lead the way to further optimizing the discovery path given (“knowing that”) the explorative experimentation until that moment and the options presented at that moment (“knowing how”). In preventive health care systems the “actionable data” are aimed at nudging or even pushing the patients towards a healthier life style.

In the following smart prescriptive systems and the anticipatory measures they can offer are presented: when the transition from prediction to prescription takes place future behaviour is not only anticipated but formed. Context-specific, adaptive micro-directives (Casey/Niblett 2015) 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 behaviour. Thus, in smart systems two variants of anticipatory behaviour exist: when deploying smart predictive systems anticipatory actions are executed both by machines and humans. In smart prescriptive systems anticipatory measures are delegated to the systems and human supervisors are at best in the loop. Such approaches may be used for social engineering based on micro-directives.

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, p.98). 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 chance of subverting it.

Such systems are autonomous only in so far as they are capable of operative and strategic control. The normative control remains in the hands of the human operators (Gransche et al. 2014). Thus, core ethical principles e.g. as proposed by Floridi and Clowls (2019) and enumerated in the introduction of this paper may be embedded in the smart systems of the future. If the users remain discontent, civil disobedience undermining the systems from within is another potential option. However, since the anticipatory guidance provided by these systems is opaque, it might be quite difficult to fight against it.

Both predictive systems and prescriptive systems rely on actionable data gained in the past in order to "know ahead and act before". Their focus is on process-oriented efficiency based on pregiven goals. In contrast, a smart society, a mentally alert one, intends to imagine ahead and act accordingly. Support may be found in the perspectives and techniques provided by the humanities and the arts (see e.g. Miller 2018). "Futures literary" may be also supported by games and simulations of possible worlds letting the gamers explore potential utopias as well as dystopias to be avoided. In a first step its exploration of the unknown relies more on science fiction than science fact. In a second step feasible innovations may be explored.

Thus, anticipation in the smart society reflects the cultural accomplishments of a society. It takes the core values of its citizens into account. Yet, it is open to novel developments. It is an artform and not a technology driven perspective. The development of future smart systems will profit from such an approach.

To sum it up: Both predictive smart systems and prescriptive smart systems rely on actionable data in order to "know ahead and act before" thus anticipating and shaping the future. Their focus is on process-oriented efficiency. In contrast, a smart society, needs to imagine ahead and act accordingly. Its perspective should be open and broad profiting from the insights provided by the humanities and the arts. The anticipatory stance in the smart society should not be technology driven. The art of anticipation – not the technology of anticipation - may open up new perspectives, and provide stimuli for smart and ethically sound innovations.

KEYWORDS: smart systems, predictive algorithms, prescriptive systems, smart society, anticipatory stance.

REFERENCES

Casey, A. & Niblett, A. (2015). The Death of Rules and Standards, University of Chicago Coase-Sandor Institute for Law & Economics Research Paper No. 738.

Foucault, M. (1980). Power/Knowledge: Selected Interviews and Other Writings 1972–1977, Harvester Press: London.

Floridi, L. & Cowls J. (2019). A Unified Framework of Five Principles for AI in Society, Harvard Data Science Review, Issue 1, DOI 10.1162/99608f92.8cd550d1

Gransche, B., Shala, E., Hubig, Chr., Alpsancar, S., Harrach S. (2014). Wandel von Autonomie und Kontrolle durch neue Mensch-Technik Interaktionen. Grundsatzfragen autonomieorientierter Mensch-Technik-Verhältnisse. Frauenhofer Verlag: Stuttgart.

Merriam-Webster Dictionary. Smart. Retrieved September 18, 2019 from https://www.merriam-webster.com/dictionary/smart

Miller, Riel (ed) (2018). Transforming the Future: Anticipation in the 21st Century, Routledge: London.

Nordmann, Alfred (2014). Responsible innovation, the art and craft of anticipation, Journal of Responsible Innovation, 1:1, 87-98, DOI: 10.1080/23299460.2014.882064

Poli, R. & Vaerio, M. (eds) (2019). Anticipation, Agency and Complexity, Springer Nature Switzerland

Poli, Roberto et al. (2019). The Project Anticipation. Retrieved September 18, 2019 from http://www.projectanticipation.org

THE EMPLOYMENT RELATIONSHIP, AUTOMATIC DECISIONS, AND THEIR LIMITS - THE REGULATION OF NON-UNDERSTANDABLE PHENOMENA

Enrico Gragnoli

Università degli studi di Parma (Italy)

enrico.gragnoli@studiogragnoli.eu

EXTENDED ABSTRACT

The first EEC laws, addressing the problem of possible limitations to the decision making of decisions regarding workers/employees based on automatic systems that in the end caused choices to not be made by humans date back to the mid-90s. This issue has now become extremely relevant, especially in contemporary labour law, for quite evident reasons; however, law scholars are the victims, not the protagonist of this debate since they must discuss the regulation of profiles and elements which they do not understand, not even superficially. This is the basic question and, that is, how can law interact according to reason towards decision-making methods based on mathematical resources so complex even their fundamental features are impossible to understand. This can very well be called a challenge to regulate the unknown. Can there be a rational regulation of a phenomenon incomprehensible to law and its scholars?

Contrary to what one might think, the issue does not only concern the work organized by the so-called digital platforms, but also (and above all) the industrial and commercial company structures with a more traditional organisation. The continuity of the relationship and the inclusion of the performance in a business context makes automatic decision paths a common feature of many employment relationships, and obviously companies exploit this feature with increasing frequency. The question we should ask ourselves is how the legal systems should react. The analysis does not have to be so much about positive law, that is focussed on outlining the existing approaches and directions, in particular EU ones, but rather focus on the possibilities of a regulatory system.

Which strategies can be adopted with automatic decision-making systems? Barring the anachronistic prohibition of similar mechanisms, since Luddites have never had much space in Western thought, the traditional idea was to force the company to disclose the basic assumptions and premises employed by these automatic evaluation systems so as to allow the workers to understand how they are judged. The public availability of said data would help workers/employees to react accordingly. Although this strategy dates back to the mid-1990s and is still valid, it has two obvious limitations. On the one hand, given the current structure of the civil trial, even if the company reveals how it carries out its automatic judgments, verifying whether their statements are true has unrealistic costs and times and, therefore, there is no realistic penalty for those who lie. Since significant case law precedents are lacking, it is not possible to establish whether and to what extent companies are tempted to conceal their real valuation systems behind convenient descriptions. However, due to the complexity of the mathematical models and the nature of a civil law trial, even if the companies presented unreliable statements, said unreliability would never be discovered, in the ordinary dynamic of a judgment, in which a reconstruction of the setting of an algorithm is actually impossible. This statement may surprise and the demonstration is not easy, if one turns to those who have no experience of a civil law trial. However, the conclusion is obvious, because a trial is not the place where such complex scientific questions can be explored.

On the other hand, it is indeed possible to discuss the protection afforded to the worker from prior information on the significant elements for the automatic assessment; the idea is trying to recreate organizational situations similar to those of the decisions entrusted only to men, while the technological innovation causes a completely different context. When examining an algorithm, the position of those who feel the effects of an automatic decision is not comparable to that of those who are evaluated by a human , which is ruled by the emotions and by that complex articulation of opinions and suggestions typical of our mind . The algorithm cannot be considered similar to man, for the profound difference in the way of assuming a determination.

So there is an alternative solution; the company must be free to use mathematical tools of its choosing for decision-making, but the laws must set down and codify rights with binding prerequisites for fundamental measures. In these areas, regardless of the results of the automatic analyses, the decisions must correspond to selected reasons, and the employer must provide a demonstration. To comply with the law the employer should set their computer aids (however sophisticated) using as a reference a conditioning evaluation grid, compliance to which they should be able to prove in court. In other areas, the employer would be free to use their preferred decision-making strategies. The protection would be selective and would not concern any decision, but only those on crucial assets , of extreme importance for the worker This stronger but limited protection would force companies to compare their merits with mandatory regulatory indications. In other words we should move from regulating all automatic decision-making procedures to a much more intense protection only for some, for example those regarding pay, of assignments, transfers, of training opportunities and dismissal.

The protection based on the explanation of decision-making methods is illusory; the worker cannot possibly dominate the algorithms that surround him and that are used for decisions against him. We should rather, with a more modest but also more effective approach, be content with substantial constraints on the decisions made in some matters, in which, free to process information with algorithms, the company should motivate according to predetermined selection lines, while it would remain free from constraints in other areas. The inversion in the regulatory approach would be significant and, precisely for this reason, a jurist with no computer science knowledge would appreciate discussing with competent people, to verify the credibility of an approach with which the legal scholar ... seeks to regulate matters that are beyond their knowledge and understanding.

Let us give some examples; if an automatic system has to select the employee worthy of a prize, it is foolish to think that the decision-making processes may be made clearer by disclosing in advance the parameters used for setting the algorithm, it goes without saying that if it is complex, if it uses many information and applies to a large number of employees. In the event of a dispute and in the context of a civil proceeding, it will never be possible to clarify the actual functioning of the algorithm and, therefore, to establish whether the statements made by the company are true or whether, on the contrary, the information given on the setting of the selection are incorrect. Nor can one think of banning such an initiative, since the strategy would be anti-historical. A different case would be whenever, in a litigation started by a dissatisfied worker, the company is forced to demonstrate positively (with the ordinary procedural instruments) the fact that the choice corresponds to the criteria established by law. This solution could be excessive in the cases pertaining to rewards, since these would be about a benefit exceeding the minimum remuneration and connected to individual merit.

But what about the cases of transfers to workplaces located very far from the place of residence or promotions? In such cases, little it matters to establish which parameters are used by the algorithm, since, in fact, it is impossible to verify in court a similar statement. Having established the legal limits to the decision, the company must be free to resort to the preferred technological solutions but, in

case of a trial, it must prove to have based its decision on a motivation consistent with the laws in force. For example, with regards to the transfer, the law can prescribe that only management profiles shall be considered for transfers and that, when evaluating profiles meeting the professional requirements, family requirements shall be taken into account.

Consider the case (even too often referred to) of the so-called digital platforms and of the mechanisms for selecting workers in charge of a service. Even if the company owning the so-called platform declared which facts are relevant for the purposes of the decisions taken with complex mathematical tools, one would never be able to establish in judgment the veracity of such statements and any attempt would have appalling costs, in no way proportionate to the economic interest of the individual worker. The law must take an opposite stance; the company decides as it wishes, with its instruments, but, in the event of a reliable objection, based on data that reveals an objective problem, it must demonstrate that no discrimination has been made and that the functioning of the algorithm does not affect subjects basing on of sex, religion, race, and so on.

There are countless examples but our intent is not to discuss specific norms, but the general framework of a rational regulation and to understand how the law should deal with a topic on which the complexity of mathematical tools prevents a direct comparison between technology and law. Article. 22 of the European Union regulation n. 679 of 2016 only apparently prohibits decisions based on automatic processing, since it admits them in the execution of contracts and, in particular, of work and, in this case, imposes a sort of confrontation between the company and the employee, often with the simple dissemination of the elements considered by the mathematical model. Nor do we find convincing of feasible the idea of refusing to use sophisticated algorithms, proposed by the same regulations, in contradiction with the current technological evolution.

KEYWORDS: Labour law, employment relationship, automatic decision, making systems, limits, regulation.

REFERENCES

Appelbuam, E., Batt, R. (2011), *Private equity at work. When Wall Street manages main street.* New York.

Bayern, S. (2014), Of bitcoins, independently wealthy software, and the Zero–member LLC*. Northwestern University law review,* n. 108, 1485 ss..

Bostrom, N. (2014), *Super intelligence. Paths, dangers, strategies*. Oxford University press, Oxford;

Cherry, M. A. (2015), Beyond misclassification: the digital transformation of work. *Comparative labor law and policy journal,* n. 37, 577 ss..

Dan, M., Cohen (2016), *Rights, persons and organizations: a legal theory for bureaucratic society,* New Orleans.

Domings, P. (2015), *The master algorithm: how the quest for the ultimate learning machine will remake our world,* New York.

Faioli, M. (2018), *Tasks and Intelligence machine,* Turin.

Lopucki, L. M. (2018), Algorithmic entities. *Washington University law review,* n. 95, 887 ss..

Prassl, J. (2017), *Humans a service. The promise and the perils of work in the gig economy,* Oxford University press, Oxford.

Thomaz, A. L., Hoffman, G., Cakmak, M. (2016), Computational human–robot interaction. Foundations and trends*. Robotics,* n. 4, 105 ss.

Weaver, J. F. (2013), *Robots are people too: how Siri, Google car and artificial intelligence will force us to change our laws,* Preager, Santa Barbara.

THE POWER TO DESIGN: EXPLORING UTILITARIANISM, DEONTOLOGY AND VIRTUE ETHICS IN THREE TECHNOLOGY CASE STUDIES

Kathrin Bednar, Sarah Spiekermann-Hoff

Institute for Information Systems and Society,
Vienna University of Economics and Business (Austria)

kbednar@wu.ac.at; spiek@wu.ac.at

EXTENDED ABSTRACT

Today, a variety of technological artefacts mediates our lives and actions. Especially with new technologies reaching into sensitive areas such as our digital privacy, the call for an ethically aligned technology design has become more and more prominent among IT design and innovation scholars. Companies following a conservative approach to IT product design and innovation develop product roadmaps (Albright and Kappel 2003) or technology strategies (Cooper and Edgett 2010) that focus on linking strategy and operations with technological capability. The resulting value propositions emphasize function instead of actual values for the future users and customers. They also imply a narrow view of risks and potentials of the envisioned product, which is risky, considering that new technological products often show their negative potential only once they are in widespread use. This motivates the consideration of a technological product's impact early in the product design process, when its characteristics can still be adjusted. But how can we assess what makes a good IT product that should be designed and developed?

For thousands of years, people have thought about what it means to do the right thing and how to come to a justified judgment in this matter. Over the years, different theories of ethics have been established and the philosophical discourse on each theory's benefits and downfalls is ongoing. More recently, both theoretical works and practical approaches have been developed for an adequate approach of ethics with regard to technology (e.g. Brey 2015; Friedman et al. 2006; Vallor 2016). Interestingly, approaches focusing on the incorporation of values in technology design differ in their underlying ethical foundation. While value sensitive design (Friedman et al. 2006) is based on a utilitarian reasoning, value-based design (Spiekermann 2016) seeks to combine the perspective of different theories of ethics. Still, there is little empirical research that compares different ethical approaches and their usefulness for technology design.

In the study presented here, we address this gap in research by applying three traditional theories of ethics in an early product design phase. While there is an ongoing discourse in the field of ethics on the unique contributions of the theories that have developed over time, consequentialism, deontology, and virtue ethics can be considered as three grand theories for the Western world. Because of their unique ethical reasoning, they are incompatible in their view of what is good or right and exhibit different advantages and problems from a philosophical view. Their different perspectives on human character and actions may thus lead to different results when deployed in a practical setting such as technology design. Utilitarianism most likely inspires ideas that focus on a technology's consequences, deontology emphasizes the designer's moral duties and virtue ethics can help to identify a technology's impact on the character of its users. We wanted to explore how ideas resulting from a traditional product-technology roadmap differ from product analyses inspired by these three theories of ethics, and furthermore, whether their underlying ethical reasoning lead to different results.

We present a two-study mixed-method research project where participants analysed three IT products with different use contexts and form factors. In study 1, participants (*N* = 40; age: *M* = 23.9, *SD* = 2.6; 47.5% female; 21 different nationalities) were split up into two groups working on a digital teddy bear dedicated to the entertainment of children and the *Foodora* mobile application for bike couriers delivering food. In study 2, participants (*N* = 35; age: *M* = 24.56, *SD* = 2.61; 38.2% female; 14 different nationalities) analysed a telemedicine platform that connects patients to specialized doctors through an online interface. Participants were instructed to first develop a product-technology roadmap and then run three ethical product analyses. Before applying the perspectives of utilitarianism, virtue ethics, and deontology, they learned about proponents and the core ethical reasoning of each of the theories. In the utilitarian analysis, participants were encouraged to think of potential consequences of the envisioned IT product and identify harms or benefits. In the virtue ethical analysis, participants focused on how the character of stakeholders could be influenced by the IT product. The deontological analysis asked participants to think of moral rules with universal significance that are supported or undermined by the IT product. To have a common level for the comparison of ideas and the resulting value propositions for each product, participants related the ethical issues and potentials that they identified to values (and virtues) and affected stakeholders.

In total, participants came up with 592 ideas for the Foodora case, 918 ideas for the digital teddy bear, and 809 ideas for the telemedicine platform. Despite the different settings of the three case studies, results show similar patterns. As expected, the product roadmap focused on specific technical or physical product features. The ethical analyses on the other hand resulted in a broad range of value-related ideas. Next to an overlap of some ideas, each analysis also produced unique ideas evolving around job positions and a company's visibility (utilitarianism), temperance, humour and flexibility (virtue ethics) or trustworthiness, self-care and personal growth (deontology).

Overall, the utilitarian analysis clearly yielded the highest number of ideas, followed by the virtue ethical and the deontological analysis. Thinking about consequences of an IT product and how stakeholders are affected did not only inspire many ideas, but also very diverse ideas. When we grouped values according to the entity that they affected (i.e. individual, society, business, technology; see Figure 1), utilitarianism showed the broadest range of value-based ideas; it also covered most stakeholder groups (see Figure 2). Interestingly, the deontological analysis, per se focusing on moral duties, elicited fewer ideas, but ideas that are almost as diverse as the ones inspired by utilitarianism. Moreover, deontology inspired the highest proportion of ideas focusing on values with social import and produced unique ideas focusing on generic values such as living a better life. The product analysis inspired by virtue ethical thinking elicited mostly ideas related to individual values and virtues. This is comprehensible because of its focus on a person's good character and behaviour. The virtues that the ideas were related to included responsibility, kindness, flexibility, patience, and courage, among many others.

When we looked at the length of the description of ideas, we discovered that the deontological analysis yielded the most elaborate ideas. A high degree of personal involvement, triggered by the instruction to think about one's own moral duties, could explain this tendency. Interestingly, deontology also differed in the overall evaluation of the IT products. Not only did it identify more issues than potentials, participants also were most critical after the deontological analysis when asked whether they would invest in the analysed product: more than half of the participants decided against investment (53.1%) while the other two analyses yielded a much lower percentage of negative decisions (utilitarianism: 21.3%, virtue ethics: 32.3%).

Figure 1. Value groups represented by the aggregated pool of ideas

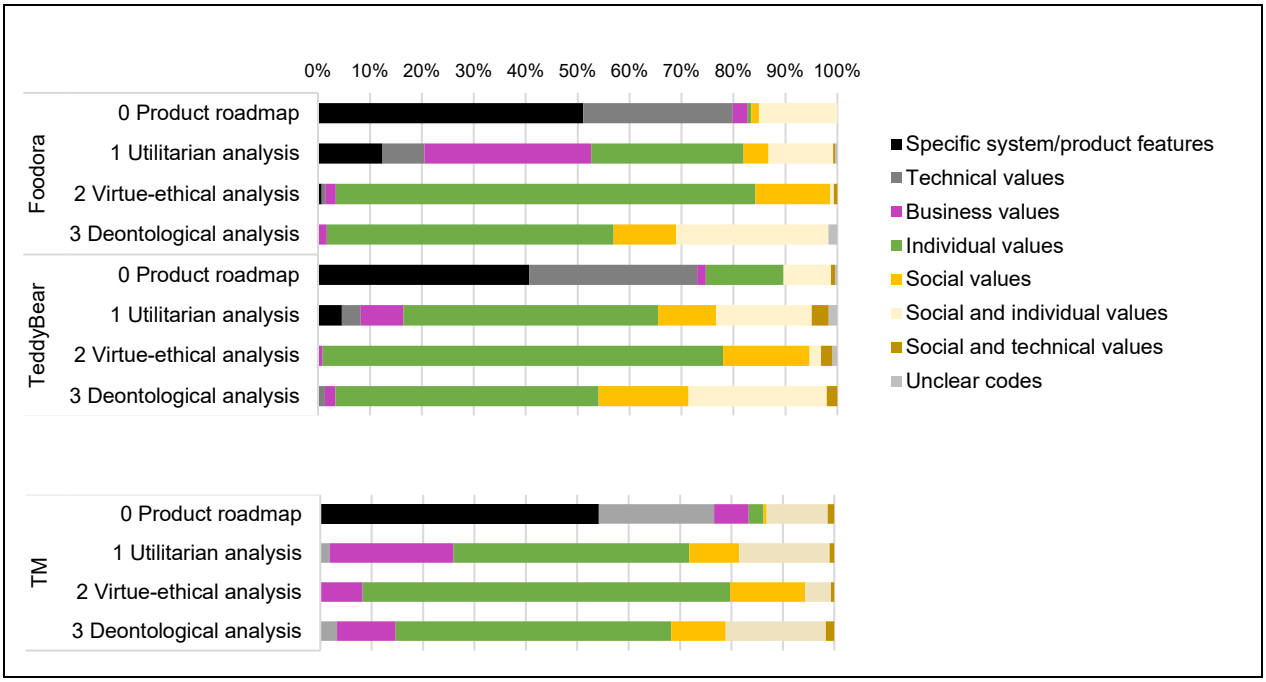
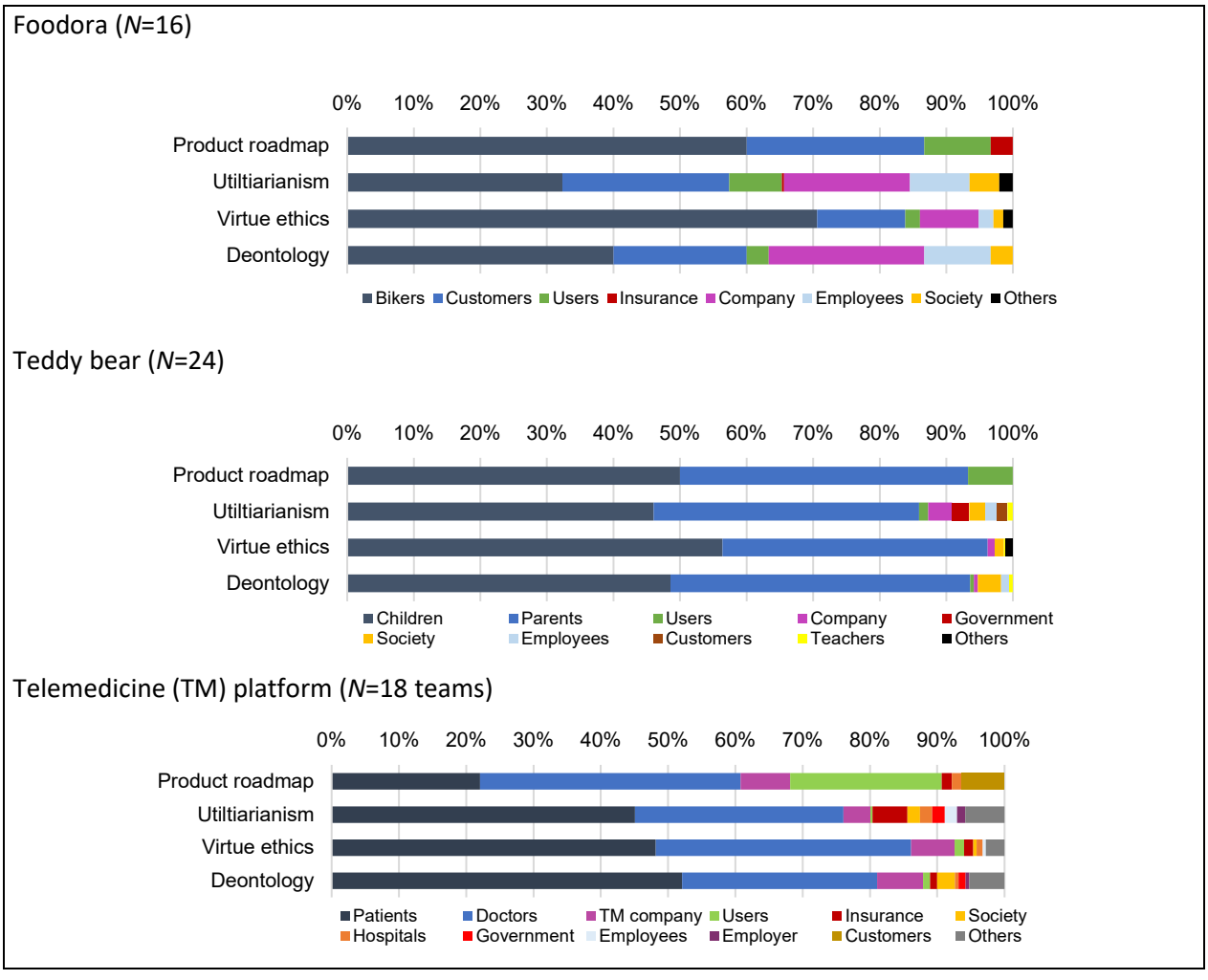


Figure 2. Stakeholder groups reflected in the aggregated pool of ideas across all analyses and product case studies



These preliminary results illustrate that each of the ethical analysis has its unique focus, with utilitarianism considering the broadest range of stakeholders and values in technology design, a clear focus by virtue ethics on individual growth and development, and an emphasis on more generic concepts and values by deontology. With these different contributions, every ethical theory serves a unique role in the identification of issues and potentials of IT products. Thus, choosing one perspective over the other for technology design needs to be well argued. Because of their unique foci, we also see a potential to combine the three ethical perspectives in the elicitation of value-related ideas, as has been suggested by Spiekermann (2016). The use of different ethical perspectives can provide an ethical grounding for design ideas and thus improve the ethical constitution of technological products for the people that are directly affected as well as for society as a whole.

KEYWORDS: technology design, ethics, utilitarianism, deontology, virtue ethics, empirical study.

REFERENCES

Albright, R. E., & Kappel, T. A. (2003). Roadmapping in the Corporation. *Research-Technology Management*, *46*(2), 31–40.

Brey, P. (2015). Design for the Value of Human Well-Being, in J. van den Hoven, P. Vermaas, and I. van de Poel (eds.), *Handbook of Ethics, Values, and Technological Design: Sources, Theory, Values and Application Domains* (pp. 365–382). Dordrecht: Springer.

Cooper, R. G., & Edgett, S. J. (2010). Developing a Product Innovation and Technology Strategy for Your Business, *Research Technology Management*, *53*(3), 33–40.

Friedman, B., Kahn Jr., P. H., & Borning, A. (2006). Value Sensitive Design and Information Systems, in P. Zhang and D. Galletta (eds.), *Human-Computer Interaction and Management Information Systems: Foundations* (pp. 348–372). Armonk, NY: M.E.Sharpe.

Spiekermann, S. (2016). *Ethical IT Innovation: A Value-Based System Design Approach*, Boca Raton: CRC Press.

Vallor, S. (2016). *Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting*, New York: Oxford University Press.

THE ROLE OF DATA GOVERNANCE IN THE DEVELOPMENT OF INCLUSIVE SMART CITIES

Damian Okaibedi Eke, John Obas Ebohon

De Montfort University (United Kingdom), London South Bank University (United Kingdom)

damian.eke@dmu.ac.uk; ebohono@lsbu.ac.uk

EXTENDED ABSTRACT

In the face of the current big data economy, the idea of 'smart cities' have gained more traction in Europe and beyond. As more cities turn to information and communications technology (ICT) for the efficient delivery of public services, there are huge potential to improve access to better infrastructure and services, including water supply and waste disposal facilities, urban transport networks, safer public spaces and improved public engagement or interaction. However, smart city initiatives have been criticised for overemphasizing technological solutions and business interests over social inclusion (Paskaleva et al., 2017) - an integral part of sustainable urban development. In a 2018 discussion panel on 'The Invisible Smart city', urban designer Gil Peñalosa stated that ''we currently design our cities as though everyone is 30 and active'' which excludes the majority of the population that doesn't fall into this athletic age group. According to the findings of the Microsoft-backed initiative- Smart Cities for All, ''most of today's smart cities, in both the global north and the global south, are not fully accessible''. These indicate that smart city designs reflect traditional urban design biases that exclude parts of the communities such as children, women, older population, the disabled, low income households and the mentally ill (O'Dell et al., 2019). With about 15% of the world population living with some sort of disability (WHO, 2011), about 12.3% of the global population over the age of 60 (ONS, 2015) and with nearly half of the world's population living under the poverty line (World Bank, 2019), there is great need for smart city initiatives to prioritise inclusion in urban development. This is particularly important because making "cities and human settlements inclusive, safe, resilient and sustainable" is a key goal of the UN's 2030 Agenda for sustainable development. Sustainable Smart cities therefore, should advance or reinforce inclusion; appreciating the diversity of different communities while removing identifiable digital barriers to social inclusion.

At the heart of a successful smart city is a big and robust data ecosystem that generates insights, stimulate innovation and efficiency, improves productivity and delivers wider social benefits (Bibri, 2018; Hashem et al., 2016). ''Big data'' refers to the datasets that represent relevant activities that are characteristically big in volume, velocity, variety, veracity and value (Chen et al., 2012; Fothergill et al., 2019). Data plays a central role in the services provided in smart cities. Digital data platforms and cloud-based systems enable smart cities to collect multimodal, cross-functional, big, complex but mostly unstructured data (Chen et al., 2014) of residents activities with associated individual and collective risks like; data protection, privacy, data sharing, environmental neglect, economic discrimination, social bias and data subject rights. Data are extracted from sources like healthcare systems, transportation, power grids, crime records, irrigation systems and other public service networks which are then used to recognize patterns and needs of the residents. While these different types of data can fuel innovation in smart cities, they can also facilitate exclusion. For instance, many of the smart city data are collected using facial recognition software but a recent study has revealed that commercial facial-recognition software show error rate of 0.8 percent for white male and 34.7 percent for black females (Buolamwini and Gebru, 2018; Raji and Buolamwini, 2019). The findings of

the study demonstrate inherent racial and gender bias and further evaluation into the cause of this evident bias in the technology shows that the algorithms are informed by datasets that were lacking in diversity (Buolamwini and Gebru, 2018). This is a further evidence that datasets have large influences on how technology excludes or unites us in today's society.

The nature of the smart city data, its method of collection and usage have great impact on issues such as; respect for human rights, benefit sharing, respect for diversity and inclusion. While available regulations particularly the EU's General Data Protection Regulation (GDPR) focuses on data subject rights, little attention is paid to the impacts of the inferential decisions made with the data which constitute exclusion of some sections of the population. The nature of the available data for such decisions can advance or impede inclusion in cities. Therefore, harnessing the benefits of smart cities for all communities is dependent on a functional data economy with good quality data and responsible governance approaches. The paper identifies the roles data governance can play in smart cities with regard to fostering inclusion. With the understanding of a smart city as ''a blend of institutions, processes, people, and technology'' (Paskaleva et al., 2017), this paper argues that an inclusive and sustainable smart city requires a sustainable data governance characterized by diverse datasets and approaches that address community, environmental and economic risks and concerns. The argument here is that the UN's SDG goal 11 cannot be achieved without a collaborative, dialogical approach to data governance where datasets, that can reflect the diverse nature of the population should be used to make inferential decisions on the people, the environment and the economy.

In this paper, two major questions are addressed. First, what is the relationship between data processing in smart cities and inclusion/exclusion? And second, what is the role of data governance in fostering inclusion in smart cities? Answers to these questions are provided through a multi-method empirical approach of critical literature review and case studies. The literature review provides conceptual perspectives on the relationships between smart cities, data governance and inclusion. The case studies on the other hand provide empirical insights on the practical impact of data governance on inclusion. This paper offers a unique contribution to the general discourse on the creation of sustainable and inclusive smart cities. The focus on data governance illustrates the interrelatedness of data and wider social issues. One common understanding is that smart city technologies are built with the principles of artificial intelligence - to unpick statistical relationships. The conclusions in this paper contribute to the ever-growing discussion on the responsible data governance for AI. These will not only be of interest to developers of smart cities but also other experts working on AI systems and inclusion.

This paper argues that sustainable data governance approaches advance inclusion in smart cities. There is a need to create data governance principles that are sustainable, responsible and consistent with the nature of data processing and fit for an AI-driven world. Sustainable data governance focuses on the **community** (creating awareness and knowledge regarding the value, utility and relevance of data at all levels for diverse population of dataset owners with common interests) **environment** (translates to the skills, regulations, policies, ethics, knowledge and tools required to manage data as an asset. And more specifically, it translates to the mechanisms required to both develop and invest in these skills and capabilities) and the **economy** (value creation, impacts and outcomes of the data governance process). There are two major areas of sustainability in a sustainable data governance - data and the approach applied. It emphasizes **diversity of datasets** about the people. Datasets about the citizens should reflect their varied identities, their environment and their economic value. Lack of diversity of citizen's datasets (including their cultural, health, economic and social differences) results in unintended bias in smart city decisions. The second area of sustainability is in the **process of governance** that should apply skills and tools to create value that will further the common interests of all communities in the city. Full representation of all communities (gender, race, religion etc) and the

environments (health, transportation, water systems) and the economy in the datasets and the process of decisions making are the focal points. To enhance inclusion therefore, smart cities should look to sustainable data governance that will address risks associated with the people, the environment and the economy and that is open, transparent, dialogical, responsible and a continuum of proactive practices.

KEYWORDS: Data governance, smart cities, inclusion, sustainability, diversity.

REFERENCES

Bibri, S.E. (2018). The IoT for smart sustainable cities of the future: An analytical framework for sensor-based big data applications for environmental sustainability. *Sustainable Cities and Society* 38, 230–253. Retrieved from https://doi.org/10.1016/j.scs.2017.12.034

Buolamwini, J., Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification, in: *Proceedings of Machine Learning Research. Presented at the Conference on Fairness, Accountability, and Transparency*, pp. 1–15.

Chen, H., Chiang, R.H.L., Storey, V.C. (2012). Business Intelligence and Analytics: From Big Data to Big Impact. *MIS Quarterly* 36, 1165–1188. Retrieved from https://doi.org/10.2307/41703503

Chen, M., Mao, S., Liu, Y. (2014). Big Data: A Survey. *Mobile Netw Appl* 19, 171–209. Retrieved from https://doi.org/10.1007/s11036-013-0489-0

Fothergill, B.T., Knight, W., Stahl, B.C., Ulnicane, I. (2019). Responsible Data Governance of Neuroscience Big Data. *Front. Neuroinform*. 13, 28. Retrieved from https://doi.org/10.3389/fninf.2019.00028

Hashem, I.A.T., Chang, V., Anuar, N.B., Adewole, K., Yaqoob, I., Gani, A., Ahmed, E., Chiroma, H. (2016). The role of big data in smart city. *International Journal of Information Management* 36, 748–758. Retrieved from https://doi.org/10.1016/j.ijinfomgt.2016.05.002

Paskaleva, K., Evans, J., Martin, C., Linjordet, T., Yang, D., Karvonen, A. (2017). Data Governance in the Sustainable Smart City. *Informatics* 4, 41. Retrieved from https://doi.org/10.3390/informatics4040041

Raji, I.D., Buolamwini, J. (2019). Actionable Auditing: Investigating the Impact of Publicly Naming Biased Performance Results of Commercial AI Products, in: Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society - AIES '19. Presented at the the 2019 AAAI/ACM Conference, ACM Press, Honolulu, HI, USA, pp. 429–435. Retrieved from https://doi.org/10.1145/3306618.3314244

UNDERSTANDING PUBLIC VIEWS ON THE ETHICS AND HUMAN RIGHTS IMPACTS OF AI AND BIG DATA

Laurence Brooks, Bernd Stahl, Nitika Bhalla

De Montfort University (UK)

laurence.brooks@dmu.ac.uk; bstahl@dmu.ac.uk; nitika.bhalla@dmu.ac.uk

EXTENDED ABSTRACT

There is significant discussion in academia, media and policy studies about ethical issues of artificial intelligence. Interventions in this discourse include those from national and European policymakers (European Commission, 2019; Artificial Intelligence Select Committee, 2018), various academics and labs working on ethics and AI (Dignum, 2018), but also accounts written for lay audiences (eg. https://www.wired.co.uk/article/artificial-intelligence-ethical-framework) and mainstream media (eg. 'The Ethical Dilemmas AI Poses for Health Care', https://www.wsj.com/articles/the-ethical-dilemmas-ai-poses-for-health-care-11571018400). What most of these contributions have in common is that they are based on anecdotal evidence or are purely conceptual in nature.

In this paper we go beyond this current state and offer a broader and more empirically-based view of ethics and human rights of AI and big data. The paper will present the initial findings of an online survey that elicited responses from a well-informed audience. The survey questions were drawn from prior empirical research that looked at current and likely future applications of AI and big data. The research aimed to answer the following research question:

> *From the perspective of a well-informed lay public, which ethical and human rights issues relating to AI and Big Data are perceived as particularly problematic and how should they be addressed?*

An answer to this question is highly important from both a practical and policy perspective. Understanding which issues, application areas and possible interventions are perceived as important or promising is important to ensure that conceptual and empirical understanding of these issues can be used as the basis for useful and evidence-based policy.

The survey is currently live and still in data collection mode. We can therefore not give an indication of findings and will confine the abstract to a description of the methodology. However, we currently have 115 complete and 214 partially completed responses (329 total), drawn from people based in 21 different countries. The survey is expected to be completed by the end of 2019 with analysis in early 2020. Therefore, the full paper will also be able to report on the empirical findings, analysis and discussion.

The research that informs the online survey is based on the recognition that the way in which ethical and human rights issues are perceived and subsequently addressed in social reality depend on local contexts. Enough understanding of these issues therefore requires more than a conceptual analysis and benefits from empirical insights. Prior to this survey we had undertaken 10 case studies of AI and big data applications (referred to as smart information systems or SIS), across 10 different application scenarios: IoT, government, agriculture, smart cities, science, insurance, energy and utilities, communication and media, retail and supply chain management. In addition to these case studies, we

developed 5 policy-oriented scenarios that described the use of AI and big data in areas that are already well under development and likely to become socially relevant soon: predictive policing, mimicking technologies (deep fake), self-driving cars, AI in warfare and education.

Based on the case studies and scenarios, we developed a list of 35 ethical issues that were recognisable in our research (see appendix A). For each of these issues we provide a brief definition that derives from our insights from the underlying research. The survey then aims to identify which of these ethical issues are perceived as being significant and important and which are of less relevance. Each question was answered using a Likert scale ranging from 1 to 5, with 1 being "not at all important" to 5 "very important''. This was followed by asking about the expectations of the respondents with regards to the future use of SIS, also drawn from the case studies and scenarios and human rights analysis. It asks about the application areas where AI and big data may become relevant. Respondents are asked to indicate whether they think the ethical and human rights issues arising in these areas are likely to become more, or less important in the future. The application areas are again drawn from the case studies and scenarios (see appendix B).

We then ask about some prominent specific issues, i.e. data privacy, transparency and fairness, bias, trust and accuracy, and inequalities and whether these are more or less likely to be important in the future and whether regulation or education would help address them. This is followed by a request to assess likely future developments, again using a matrix approach with a five-point Likert scale.

The final question refers to possible options to address these issues. Respondents are asked to state whether they find the following options likely to be successful or unsuccessful:

- Current legislation/regulation to support human rights
- Future legislation/regulation to support human rights
- Creation of new a regulator for AI/big data
- Ethical guidelines/codes of conduct for SIS developers
- Ethical guidelines/codes of conduct for SIS users
- Standardisation
- Certification
- Technical options
- Education

Respondents are then given space to capture any other input, as well as some demographic data, to allow researchers to assess the sample.

Following the definition of the survey questions and ethics approval from the relevant University research ethics committee, the survey was deployed during the middle of October 2019. It was sent out to more than 1000 individual respondents who had been identified as stakeholders with an interest in AI and big data during an earlier phase of the project. In addition, it was sent out to several email lists, groups and individual contacts using a snowball approach. Prior to the dissemination of the survey to be intended recipients, we undertook two pilot studies. The first pilot study was undertaken by a discussion with the consortium of the research plan during a fiscal consortium meeting. Following this, the finalised version of the survey using the intended tools for inviting participants and collecting the data was sent to all project participants. These pilot exercises provided valuable insights and helped clarify questions and improve the wording of the survey. The final paper will provide selected results and insights from the survey, on topics of interest to the ETHICOMP community and the conference participants.

Appendix A: Ethical Issues

Ethical Issues	Brief Explanation
Access	Related to the potential to favour people with more money to access SIS (ie. poorer people may not be able to afford access or the knowledge to access these technologies), at the local national or even global level
Accountability and liability	Related to the need to explain and justify one's decisions and actions to its partners, users and others with whom the SIS interacts; Regarding liability, it is related to the sense that a person who has suffered loss because of a decision made by SIS may be owed a duty of care
Accuracy of Data	Related to using misrepresentative data or misrepresenting information (ie. predictions are only as good as the underlying data) and how that affects end user views on what decisions are made (ie. whether they trust the SIS and outcomes arising from it)
Accuracy of Recommendations	Related to the possibility of misinterpreting data, implementing biases, and diminishing the accuracy of SIS recommendations
Bias	Related to the samples people that might be chosen/involved in generating data
Control	The degree to which people perceive they or the SIS are in control
Democracy	The degree to which all involved feel they have an equal say in the outcomes, compared with the SIS
Discrimination	Related to discrimination in terms of who has access to data. For example, discrimination in algorithms may be conscious or unconscious acts by those employing the SIS, or a result of algorithms mirroring society by reflecting pre-existing biases
Economic	Related to the potential for SIS to boost economic growth and productivity, but at the same time creating equally serious risks of job market polarisation, rising inequality, structural unemployment and emergence of new undesirable industrial structures
Fairness	Related to how data is collected and manipulated (ie. how it is used), also who has access to the data and what they might do with it as well as how resources (eg. Energy) might be distributed according to the guidance arising out of the data
Freedom	Related to the manipulative power of algorithms results in nudges towards some preferred behaviours, free will and the self-determination of people, which are the preconditions for democratic constitutions, run the risk of being compromised
Health	The use of SIS to monitor an individual's health and how much control one can have over that
Human Contact	The potential for SIS to reduce the contact between people, as they take on more of the functions within a society
Digital divide	Related to the potential for SIS to favour people with more money (ie. poorer people may not be able to afford access or the knowledge to access these technologies)

Ethical Issues	Brief Explanation
Dignity and care for the elderly	The level at which SIS is seen as impacting on the dignity and care for older people, for example how much a care robot might exert over an older person's life and 'tell them what to do'
Dual use	Concerns over the potential use of SIS for both military and non-military use
Discrimination	Related to discrimination in terms of who has access to data. For example, discrimination in algorithms may be conscious or unconscious acts by those employing the SIS, or a result of algorithms mirroring society by reflecting pre-existing biases
Economic	Related to the potential for SIS to boost economic growth and productivity, but at the same time creating equally serious risks of job market polarisation, rising inequality, structural unemployment and emergence of new undesirable industrial structures
Fairness	Related to how data is collected and manipulated (ie. how it is used), also who has access to the data and what they might do with it as well as how resources (eg. Energy) might be distributed according to the guidance arising out of the data
Freedom	Related to the manipulative power of algorithms results in nudges towards some preferred behaviours, free will and the self-determination of people, which are the preconditions for democratic constitutions, run the risk of being compromised
Health	The use of SIS to monitor an individual's health and how much control one can have over that
Human Contact	The potential for SIS to reduce the contact between people, as they take on more of the functions within a society
Digital divide	Related to the potential for SIS to favour people with more money (ie. poorer people may not be able to afford access or the knowledge to access these technologies)
Dignity and care for the elderly	The level at which SIS is seen as impacting on the dignity and care for older people, for example how much a care robot might exert over an older person's life and 'tell them what to do'
Dual use	Concerns over the potential use of SIS for both military and non-military use
Environment	Related to the use of SIS resources contributing to the production of greenhouse emissions as well as impacting the environments they are built on
Individual Autonomy	Related to how algorithms used in SIS affect how people analyse the world and modify their perception of the social and political environment
Inequality	Related to the digital divide and the potential for SIS to favour people with more money (ie. poorer people may not be able to afford access or the knowledge to access these technologies), at the local national or even global level; also related to discrimination in terms of who has access to data

Ethical Issues	Brief Explanation
Informed Consent	Related to informed consent being difficult to uphold in SIS when the value and consequences of the information that is collected is not immediately known by users and other stakeholders, thus lowering the possibility of upfront notice
Integrity	The internal integrity of the date used as well as the integrity of how the data is used by a SIS
Justice	The use of SIS within judicial systems, for example AI used to 'inform' judicial reviews in areas such as probation
Ownership of Data	Where ownership of data sits, and how transparent that is, for example when you give details to an organisation, who then 'owns' the data, you or that organisation
Manipulation	What is done with and to the data, for example when used with other data points to make a dataset, how is this done, what basis and who is making sure that it is not in some way abused
Military, Criminal, Malicious Use	Related to the use of SIS to make predictions about future possible military, criminal and malicious scenarios that can elaborate and improve strategies for instance, in cyber-attacks and cyber espionage
Power Asymmetries	Related to the fact that the knowledge offered by SIS and its practices, and how to regulate this knowledge is in the hands of a few powerful corporations
Privacy	Related to how much data is collected, where from (ie. public such as social media or privately directly from the person/home) and how well it is looked after
Responsibility	Related to the role of people themselves and to the capability of SIS to answer for one's decision and identify errors or unexpected results
Rights	As SIS, such as AI, gain more complexity and empowerment, then to what degree they should have rights and be protected, eg. digital personhood
Security	Related to the sensitivity of SIS given the amounts and kind of data that they hold which needs protection of the systems against hackers to ensure a positive impact and reduce risks
Sustainability	Related to a concern about the data centres needed to run SIS, as the demand for huge computing power along with greater resources and energy required for data collection, storage and analytics
Transparency	Related to the need to describe, inspect and reproduce the mechanisms through which SIS make decisions and learns to adapt to its environment, and to the governance of the data used created.
Trust	Related to using misrepresentative data or misrepresenting information (ie. predictions are only as good as the underlying data) and how that affects end user views on what decisions are made (ie. whether they trust the SIS and outcomes arising from it); also related to informed consent and that helps with trust
Unemployment	The worry that use of SIS will lead to significant drop in the need to employ people
Use of Personal Data	The concerns over how SIS might use your and anyone's personal data

Appendix B: Application Areas

- SIS Application areas
- Employee Monitoring and Administration
- Government
- Agriculture
- Sustainable Development
- Science
- Insurance
- Energy and Utilities
- Communications, Media and Entertainment
- Retail and Wholesale Trade
- Manufacturing and natural resources
- Predictive Policing
- Self-Driving Cars
- Mimicking Technologies
- Warfare
- Education

KEYWORDS: Ethics, Human Rights, AI, Big Data, Survey.

REFERENCES

Artificial Intelligence Select Committee Report of Session 2017-2019. *AI in the UK: ready, willing and able?* Volume 1. Report: House of Lords, HL 1 (2018).

Dignum, V. (2018). Ethics in artificial intelligence: introduction to the special issue, *Ethics and Information Technology,* 20(1), 1-3.

European Commission, (2019). *Ethics Guidelines for Trustworthy AI* Retrieved from https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai

12. Technology Meta-Ethics

Track chair: Wilhelm E. J. Klein, Researcher in Technology and Ethics, Hong Kong

9 HERMENEUTIC PRINCIPLES FOR RESPONSIBLE INNOVATION

Wessel Reijers

European University Institute

wesselreijers@gmail.com

EXTENDED ABSTRACT

The last decade has seen a great proliferation of methods for practising ethics in research and innovation (R&I) (Reijers et al., 2017). These methods assist both ethicists and practitioners (designers, engineers) to integrate ethical concerns in design practices and governance processes of research and innovation. For instance, value sensitive design (Friedman, Kahn Jr., & Borning, 2006) assists engineers in conceptualising which particular values to integrate in technology design, understanding how different stakeholders relate to these values, and investigating the technical possibilities for design interventions. Similarly, but focusing on the governance of R&I, the ethical impact assessment method (Wright, 2014) prescribes different steps for organising a project focused on generating innovations in a responsible way, by engaging with stakeholders in the process, and identifying and resolving ethical risks.

Methods for practising ethics frequently incorporate lists of ethical principles or values into their procedural structures, in order to ground decision making processes on a normative basis. This happens in different ways. Value sensitive design prescribes lists of values based on 'stakeholder preferences' (Borning, Friedman, & Kahn, 2004), which means that stakeholders are asked about what they find important to consider with regards to a particular technology. The ethical impact assessment approach merely mentions that it 'is possible to identify some ethical and/or social issues' (Wright, 2014, p.166) and presents an open-ended list that might be changed at will, depending on the R&I project. Other methods, such as the ethical matrix approach (Forsberg, 2007), take their cue from the 'principlism' approach that is widely used in biomedical ethics, and is based on four basic principles of autonomy, beneficence, non-maleficence, and justice.

The abovementioned approaches to formulating lists of principles or values for responsible innovation have frequently been criticised. Principlism, as the most coherent approach of the three, attempts to incorporate concerns from different strains of ethics (deontology, consequentialism, and virtue ethics). However, it has been criticised for lacking a normative basis (Clouser & Gert, 1990), meaning that one simultaneously has to support potentially conflicting theories in order to embrace all its four principles. The list of values offered by value sensitive design has been criticised for the reason that stakeholder preferences are highly contingent and no proper basis for ethical deliberation (Reijers, 2019). Open-ended lists like the one used in ethical impact assessment are most problematic, because they remain fully arbitrary and do not give people any grounds as to which principles to use and which to discard.

In short, existing lists of principles and values are problematic because they are at best expressions of stakeholder preferences and at worst arbitrary constructions. There are two good reasons for these problems to arise. First, while engaging in responsible innovation we have to deal with a plurality of values and principles: different stakeholders with different backgrounds have different ideas of what is important in life and for society as a whole. Therefore, committing to a 'fixed' list of principles seems to be fallacious, because it would exclude certain justified convictions. However, the issue of value pluralism is taken serious by philosophers who constructed comprehensive normative frameworks,

such as Habermas (1990) who argues that despite value pluralism we need to set universal principles for communicative action in order to achieve agreement on certain societal goods. Second, committing to a single ethical theory such as consequentialism seems fallacious because it would create potential conflicts in practice, by excluding for instance deontological considerations which are equally justified. However, this objection disregards the efforts that have been undertaken in normative ethics to integrate different ethical perspectives in comprehensive frameworks. I therefore argue that instead of rejecting the possibility of grounding a list of principles or values in normative ethics, philosophers working in technology ethics and responsible innovation should explore ways to formulate lists of principles that are grounded on a normative framework that both addresses the concerns of mere stakeholder preferences and arbitrariness, and of value pluralism and conflicting ethical theories.

Notwithstanding other valuable efforts in this direction, such as the development of discourse ethics in technology ethics (Rehg, 2015), I argue that Paul Ricoeur's 'Little Ethics' that he developed in Oneself as Another (1992) offers a valuable point of departure to fulfil the abovementioned ambition. Even though it is not framed in terms of a list of principles, 9 hermeneutic principles can be deduced from its structure: 1) self-esteem, 2) self-respect, 3) conviction, 4) care, 5) respect for persons, 6) critical solicitude, 7) equality, 8) rule of justice, and 9) sense of justice. A framework based on these principles addresses concerns raised against other lists in ethics of technology and responsible innovation, because principles are not based on mere stakeholder preferences and are also not arbitrarily listed but are justified by means of an underlying normative framework that integrates a philosophical anthropology with a theory of institutions; which culminates in the general normative aim that results from the framework: to live well, with a and for others, in just institutions. Furthermore, a list of principles based on Ricoeur's 'Little Ethics' addresses the concern of value pluralism because it commits to an ethics that accommodates rather than excludes a plurality of values (Ricoeur, 1992, p. 182); and it addresses the concern of conflicting theoretical perspectives by integrating the teleological perspective (through the notion of the good) and the deontological perspective (through the notion of the obligation).

To understand the structure of the 9 principles, two different axes have to be considered. The first axis refers to the process of ethical reasoning, the second to reasoning with regards to the self, the other, and the unknown other represented by the institution. The first axis conceptualises how a practitioner intending to engage in responsible innovation 1) forms a conception of the good (e.g., a care robot should help people be independent), 2) checks this conception of the good with a norm that limits it (e.g., the care robot should limit the independence of the patient to prevent safety risks), and 3) refines the conception of the good by reflecting on its limitations (e.g., the care robot should allow for human intervention when it generates a safety risk because of increased patient independence). Note that the first and third steps incorporate teleological reasoning while the second accommodates deontological reasoning. The second axis conceptualises how each of these three types of reasoning happen both at 1) the individual level, (e.g., an engineer setting aims for his work, limiting these aims by means of professional norms, and critically acting on these norms – sometimes overstepping them for the sake of situational necessity); 2) the interpersonal level (e.g., colleagues in a team setting product goals, limiting these goals by adhering to ethical rules, and critically reflecting on these rules by balancing technical objectives with ethical goals); 3) and the institutional level (e.g., companies instilling their technical personnel with certain political virtues, set hard, legal obligations for the development of technologies, and facilitate procedures for public debate to address conflicts in the company). Because these two axes intersect, they result in nine principles that set out the normative basis for ethical decision making in responsible innovation.

The full paper will further develop these nine principles in the context of a concrete case of a technology being developed. I would like to conclude by arguing that in addition to addressing the four

concerns with normative frameworks for principles in responsible innovation, the 9 principles that derive from Ricoeur's 'Little Ethics' additionally allow for proper ethical questioning and political questioning, which are often missing in methods in technology ethics and responsible innovation.

KEYWORDS: responsible innovation, principles, Ricoeur, critical hermeneutics, technology ethics.

REFERENCES

Association of Fundraising Professionals. (2017). International statement on ethical principles infundraising. http://www.afpnet.org/Ethic s/IntlA rticl eDeta il.cfm?ItemN umber =3681.

Borning, A., Friedman, B., & Kahn, P. H. (2004). Designing for Human Values in an Urban Simulation System : Value Sensitive Design and Participatory Design. In PDC-04 Proceedings of the Participatory Design Conference, Vol 2, Toronto, Canada (pp. 68–71). Palo Alto.

Clouser, K. D., & Gert, B. (1990). A Critique of Principlism. The Journal of Medicine and Philosophy, 15, 219–236.

Forsberg, E. M. (2007). Pluralism, the ethical matrix, and coming to conclusions. Journal of Agricultural and Environmental Ethics, 20, 455–468. https://doi.org/10.1007/s10806-007-9050-0

Friedman, B., Kahn Jr., P. H., & Borning, A. (2006). Value Sensitive Design and Information Systems. In K. E. Himma & H. T. Tavani (Eds.), Human-Computer Interaction and Management Information Systems: Foundations (pp. 1–27). John Wiley & Sons, Inc. https://doi.org/10.1145/242485.242493

Habermas, J. (1990). Moral Consciousness and Communicative Action: Moral Conciousness and Communicative Action. Cambridge, Massachusetts: MIT Press.

Rehg, W. (2015). Discourse ethics for computer ethics: a heuristic for engaged dialogical reflection. Ethics and Information Technology, 17(1), 27–39. https://doi.org/10.1007/s10676-014-9359-0

Reijers, W. (2019). Moving From Value Sensitive Design to Virtuous Practice Design. Journal of Information, Communication and Ethics in Society, Forthcomin. https://doi.org/https://doi.org/10.1108/JICES-10-2018-0080

Reijers, W., Wright, D., Brey, P., Weber, K., Rodrigues, R., O'Sullivan, D., & Gordijn, B. (2017). Methods for Practising Ethics in Research and Innovation: A Literature Review, Critical Analysis and Recommendations. Science and Engineering Ethics. https://doi.org/10.1007/s11948-017-9961-8

Ricoeur, P. (1992). Oneself as Another. (K. Blamey, Ed.). Chicago: University of Chicago Press.

Wright, D. (2014). Ethical impact assessment. Ethics, Science, Technology, and Engineering, 163(c), 163–167. https://doi.org/10.1016/j.clsr.2011.11.007.Wright

DELIBERATING ALGORITHMS : A DESCRIPTIVE APPROACH TOWARDS ETHICAL DEBATES ON ALGORITHMS, BIG DATA, AND AI

Stef van Ool and Katleen Gabriels

Maastricht University (Netherlands)

stefvanool@gmail.com; k.gabriels@maastrichtuniversity.nl

EXTENDED ABSTRACT

This paper analyses moral argumentation on big data and algorithmic decision-making (ADM). The body of data consisted of ten expert interviews. The main arguments were analysed and categorized in order to study recurring patterns of argumentation. Findings reveal that consequentialist considerations were most dominant.

RESEARCH CONTEXT

Our society has become a data-driven 'algorithmic culture' in which we have come to approach and understand ourselves, others, and our surroundings increasingly through data and algorithms (Pasquale, 2015). The promises and pitfalls of this algorithmic society are widely discussed in academic and societal debates. Algorithms are helpful tools, especially in analysing complex datasets. Yet, several concerns have been raised, for instance concerning fairness, transparency, and non-neutrality. Critics warn that blind application of algorithmic decision-making (ADM) can amplify biases already present in data, which might particularly affect underrepresented groups in society (see e.g O'Neil, 2016; Noble, 2018).

Every time a new technology is introduced, there are recurring patterns of moral argumentation. Drawing upon the most common arguments, philosophers of technology Tsjalling Swierstra and Arie Rip (2007) developed a descriptive ethical framework of New and Emerging Science and Technology (NEST). Instead of arguing for or against a specific perspective, the framework explains the 'grammar' which is used by different sides in the debate and which defines the overall dynamics of how an ethical controversy over a new science or technology enfolds (Swierstra & Rip, 2007; Swierstra, 2016). Ethical arguments are not only about voicing criticism; the promotion of a new technology is equally important in ethical terms, as it is about articulating why something is 'good' (Swierstra & Rip, 2007).

Drawing upon Swierstra and Rip's framework, this paper analyses moral argumentation on ADM and, related to this, on big data and databased Artificial Intelligence (AI). Big data led to new breakthroughs in databased AI, especially in the domain of machine learning, as computer models are trained on huge data sets (Russell & Norvig, 2016, p. 29). For this study, we conducted ten in-depth interviews with experts working in research, the European Commission, NGOs, the private sector, politics, and the public domain (police). We subsequently analysed our dataset (46 000 words) through the analytical lens of the NEST-ethical framework.

NEST-ETHICAL PATTERNS OF MORAL ARGUMENTATION

This framework not only provides insight into different forms of ethical appraisal and criticism towards NEST, but also how different arguments call each other into existence in the form of argumentative

patterns. It illustrates which categories of arguments are most prominent, helps explain why this is the case, and supports decision-making when trying to find closure for ethical controversies.

Swierstra and Rip (2007) offer a typology that consists of four recurring patterns of moral argumentation. Most often, a new technology comes with promises regarding the positive (measurable) consequences that follow its introduction. These promises take the form of direct increases in effectivity, knowledge creation, and/or cost-reduction. This resonates with the ethical theory *consequentialism*, which states that an action is deemed morally right or wrong based on the overall consequences it produces for all actors involved. Critics point out the uncertain nature of such promises, in some cases even claiming there will be more negative than positive consequences associated with the technology in question.

Subsequently, *deontological* considerations about fundamental moral rights and duties are introduced mainly as a check on consequentialist reasoning. Such moral rules and principles are intrinsically right and should be adhered to, regardless of the direct consequences. Promoters also apply deontological justification for developing a technology, by claiming it helps to fulfil a fundamental duty towards society that an organization has. However, more often than not this category of arguments is used to voice criticism and alter, or even halt, the course of technological development.

The third and fourth categories of ethical arguments expected to be upfront when a new science or technology is introduced, concern issues of *distributive justice* and *conceptions of the good life*. The former deal with questions regarding a just distribution of costs and benefits. Are there privileged groups who reap all the benefits, while already marginalized groups are left to deal with the risks and negative effects? Good life ethical arguments are raised in relation to ideal images of the society we want to live in. The technology in question can either help us get there or be seen as an instrument that gives shape to a life we should morally resist.

Besides these normative arguments, Swierstra and Rip (2007) identify meta-ethical perspectives that deal with general ideas about the relationship between technological development and morality. Such meta-ethical perspectives centre around three main questions: 1) to what extent is technological development malleable; 2) are new science and technology really all that new or are they continuities and improvements of existing practices; and 3) how should the impact of new technology on our morality be characterized?

RESEARCH FINDINGS

We first looked for main themes in our body of data. Five main themes arose in our dataset: 1) big data and ADM as ‘game changing’ and unavoidable developments; 2) the novelty of ADM models; 3) importance of (societal) values, duties, and human rights; 4) conceptual muddles, policy vacuums, and regulation; and 5) the difference between public and private sector.

Then, we categorized our dataset according to four overarching ethical theories in order to study recurring patterns of moral argumentation: consequentialism, deontology, distributive justice, and good life ethics. Our empirical findings point at the underlying dynamics that drive ethical deliberation concerning ADM and big data. They show, firstly, that consequentialist considerations are dominant when promoting technologies based on algorithmic models. This so-called ‘efficiency paradigm’ receives substantial criticism based on equally consequentialist reasoning. General and vague promises are called into question. Experts point at a mysticism that surrounds big data and algorithms which informs a rather naïve belief in what these technologies are actually capable of.

Secondly, deontological principles such as transparency and fairness are invoked to counter the presumed negative effects of consequentialist reasoning. Critics and proponents alike stress the importance of ADM procedures being transparent and explainable in such a way that the overall model can be scrutinized, but by emphasizing the need for a human in-the-loop in those cases where highly impactful decisions need to be made. Yet, operationalizing these principles into concrete practices remains to be problematic.

Furthermore, there are arguments that focus on a *just* distribution of costs and benefits. How, if at all, can we guarantee that certain parts of the population are not disproportionately faced with the negative effects of ADM? This question is related to a fourth category of arguments, namely those that make reference to the ways in which general societal value structures guide our use of technology. The way in which we, as a society, define what is to be considered as problems in the first place determines how we use an algorithmic model to solve them. These definitions are mutually shaped by technologic capabilities. What can be processed by an algorithm becomes important, while the rest is at risk of losing relevance.

What Swierstra and Rip characterize as meta-ethical arguments can be found in the perspectives that discuss the novelty of (predictive) ADM and the presumed neutrality of computer models. Those who subscribe to this latter idea are more prone to giving arguments that characterize the use of algorithms, big data, and AI as inherently positive developments. When faced with ethical criticism, a common reaction was to downplay the novelty of algorithmic practices. In this meta-ethical category of arguments, we also see questions about how we should deal with changing moral frameworks of, for example, privacy. Where some informants argued we have to draw a line somewhere when it comes to data collection and automated decision-making, others claimed that we will inevitably get used to changing privacy norms and that this is not something that should naïvely halt technological development.

Other observations made during the interviews were the lack of clarity when it comes to technological and ethical concepts used when both promoting and voicing criticism. Also, it became apparent that the process of ethical reflection on technology is still seen by some to be inherently different from the seemingly neutral and objective process of developing computer models. Recent years have seen increased attention for the things that can go wrong if systematic reflection on ethical objections is not part of a design process from the start. Ethical reflection is then reduced to, sometimes literal, checklists that are used to tick pre-determined boxes after design choices have been made. In reality, ethics is a dynamic process of continuous reflection and deliberation.

Overall, our small-scale study is an exploratory study that allows us to highlight key arguments in current ethical discussions and their interrelatedness, as well as offer insights into the importance of conducting similar argumentative inquiries. In our final paper, we also give recommendations for future research that are directly relevant for actors that engage with the technologies at hand.

KEYWORDS: Algorithmic Decision-Making, Big Data, Artificial Intelligence, New and Emerging Science and Technology, Moral Argumentation.

REFERENCES

Noble, S. U. (2018). *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York: New York University Press.

O'Neil, C. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. London: Penguin Books.

Pasquale, F. (2015). The algorithmic self. *The hedgehog review: Critical reflections on contemporary culture* 17 (1).

Russell, S. J., & Norvig, P. (2016). *Artificial Intelligence: A Modern Approach (Third Edition)*. Essex: Pearson Education Limited.

Swierstra, T. (2016). The Ethics of New and Emerging Science and Technology. An Introduction. In R. Nakatsu, M. Rauterberg, & P. Ciancarini (eds.) (2016). *Handbook of Digital Games and Entertainment Technologies*. Dordrecht: Springer.

Swierstra, T., & Rip, A. (2007). Nano-ethics as NEST-Ethics: Patterns of Moral Argumentation About New and Emerging Science and Technology. *NanoEthics 1*: 3-20.

DIGITAL RECOGNITION OR DIGITAL ATTENTION: THE DIFFERENCE BETWEEN SKILLFULNESS AND TROLLING?

Bo Allesøe Christensen, Thessa Jensen

Aalborg University (Denmark), Aalborg University (Denmark)

boallesoe@hum.aau.dk; thessa@hum.aau.dk

EXTENDED ABSTRACT

In this article we propose a new understanding of Axel Honneth's theory on recognition placing it in relation to social media. Furthermore, we relate the notion of recognition to Løgstrup's ontological ethics emphasizing the importance of design and intent of use in social media (Løgstrup 1997; Jensen 2016; Jensen 2013).

Digital recognition depends on the creation of relationships between users who acknowledge each other and help develop the other's skills. The users connect and relate through an important third, together creating a place within the endless space of the internet – a place, the individual user feels they belong.

Digital attention, on the other hand, focuses on the strategy of following, provoking and interacting with different users to further one's own's reach and place within a certain hierarchy. The importance is on the 'Me', using whatever means are necessary to gather a higher status and prestige.

In short: through digital recognition, a user's goal is to create new content, relate and also interact with other users. The means to do this is by being creative, supportive, and affirmative. Through digital attention, the user's goal is to get a higher status within a group by means of creating reactions from opposing groups, and strategically use relationships with followers.

Contemporary cultural theorists often dismiss the whole idea of the "social" as merely a simplifying construction reducing the complexity of our social lives, and only necessary for sociology as a discipline to carry on (e.g. Latour 2005). They may even disregard the idea all together adopting either a neuroscientific explanation of human relations instead (e.g. Castells 2009), or a sole focus on algorithms as a structure influencing our digital lives (e.g. Floridi 2014). It is however also possible to consider the social as taking place (Christensen 2017). It takes place on and through media, the social appearing through new processes of creating communities and as contested in the same process by different actors trying to appropriate the social media space.

Social media is, in this sense, not a term of description but an appropriation of the social (Mejias 2013). As a consequence, any research into social media cannot settle on just describing the latest platforms, it must also focus on how social media generates and fails to generate ethical and normative concerns, and how this ethical side of our social life can be developed (Couldry and van Dijk 2015).

The question is: how will a theoretical framework capable of encompassing the complexity, contestation and ethical potentiality of digital social life look like, especially when focusing on the design and functionality of digital media?

Answering this question will take its point of departure in Axel Honneth's theory of recognition (Honneth 1996; 2008; 2012; 2015). Honneth's theory is very well suited as a basis since it understands the social as:

1) a struggle for recognition between people. Recognizing someone as having a certain feature, like being a good friend, implies treating this someone normatively in accordance with how friends are treated. Misrecognition, however, works in the opposite direction – treating people as to not have a genuine claim to what they do and are.
2) It works with a stratification of different institutional levels wherein this struggle of recognition takes place: the intersubjective relation of the family where primary upbringing and satisfaction of needs take place; relations in civil society like friendships, involving institutions like schools and sport clubs; and relations to the state where you are recognized as a citizen.
3) Honneth understands this struggle as essentially moved by a moral impetus towards establishing mutual confidence, respect and esteem in our social relations.

At each level of recognition there is also the possibility of misrecognition. These range from threatening the physical and social integrity, in the form of abuse and denial of rights, to attacking a person's dignity by denigration and humiliation. This enforces an evaluation of the behavior and norms presupposed in the act of misrecognition to restore mutual recognition: why did it take place, and how can we avoid it happens again?

Honneth, however, has never addressed the role of media or social media within his theory, probably because his theory was worked out before the global spread of the internet, hence before social media had any impact. This implies rethinking the theoretical basis of his theory, and therefore posing a second related research question: does social media and media generally work across the three institutional levels of struggles of recognition, or have social media and media generally become a new institutional level with a structure of recognition of its own?

It is this second question which gives rise to look at the difference between recognition and attention when being online. Taking Paßmann's (2018) impressive research of the German Twitter sphere as our starting point, it shows not recognition but attention as the main goal when tweeting. Paßmann claims that he and his fellow Tweeters are seeking recognition. However, in his interviews and his own autoethnographic approach, it becomes clear that the place in the very tangible hierarchy of Twitter Favs is a grab for attention, not recognition.

When being online, whether it is Twitter, Tumblr, or AO3 (archiveofourown.org), the one thing which separates the online from the offline world is the registration of every keystroke, every upload, every like or share or retweet. Every single online medium we have encountered has some kind of statistics provided for its user. In the following we avoid talking about privacy issues and the registration, which is not visible for any user.

However, Facebook, Instagram, Youtube etc. provide statistics for their users to measure their impact in one way or another. These statistics can be motivating. Especially when it is used to create a hierarchy of users. In the German Twitter sphere, this hierarchy was the talking point whenever Paßmann participated in offline Twitter meetings. He discussed strategies to gain more followers and larger impact, before talking about how he liked certain kinds of content. The latter was again talked about in a strategic way; how would he be able to learn from this way of writing tweets to gain a larger following? How could he use the ritual of gift giving (liking or faving tweets, retweeting) to get other, more successful Tweeters to follow him, to give him a higher status by their presence in his crowd of followers?

Paßmann is not seeking friendships or relationships, even if he ends up becoming friends with a few of the many Tweeters he meets during his research period. Also, some of the Twitter meetups make

him realize how little he has to talk about with certain other Tweeters, he had been following and acknowledged for their ironic tweets.

This experience is almost the opposite of our research into the fandom communities found on Twitter, Tumblr, and AO3. As we show in (Christensen & Jensen 2018; Jensen 2013; Jensen & Vistisen 2013), the main reason for participating in these communities comes from the interest in and love for a common third, that is the object for the fandom group in the first place. This means, that offline meetings always have a common subject of interest, something everybody in the room is able to talk about, contribute to, and discuss with the others.

How does the use of a platform like Twitter differ between the way Paßmann and the German Twittersphere uses it and the way, fandom uses the very same platform?

As we will show in this article, social media have changed over the past decade to ensure easy access, easy upload and distribution of its content. By doing so, it is our conclusion that this change of functionality not only gave rise to a change in uses and users, but it has also changed the possibility to create interactions which are meaningful and thought provoking. Instead, social media create a need for attention, almost a craving for the next like or retweet, ensuring a pressure to create offensive content which in turn creates easy answers to difficult questions. Design for attention instead of recognition.

KEYWORDS: Digital recognition, Functionality of platforms, Honneth's theory of recognition, Løgstrup's ontological ethics, Design ethics.

REFERENCES

Castells, M. (2009). Communication power. Oxford, UK: Oxford University Press.

Christensen, B. A. (2017). A place for space? Journal of dialogical science, 10(2), 35-43.

Christensen, B. A., & Jensen, T. (2018). The JohnLock Conspiracy, fandom eschatology, and longing to belong. Transformative Works and Cultures, 27, [2]. https://doi.org/10.3983/twc.2018.1222

Couldry, N.; van Dijck, J. (2015). Researching Social Media as if the Social Mattered. Social Media+Society, July-December, 1-7

Floridi, L. (2014). The Fourth Revolution. Oxford, UK: Oxford University Press

Honneth, A. (1996). The Struggle of Recognition. Cambridge, Massachusetts: MIT Press

Honneth, A. (2008) Reification. Oxford: Oxford University Press

Honneth, A. (2012). The I in We. Malden, MA: Polity Press

Honneth, A. (2015). Freedom's Right. New York: Columbia University Press

Jensen, T. (2013). Designing for relationship: Fan fiction sites on the Internet. In H. Nykänen, O. P. Riis, & J. Zeller (Eds.), Theoretical and Applied Ethics (1 ed., Vol. 5, pp. 241-255). Aalborg: Aalborg Universitetsforlag. Applied Philosophy / Anvendt Filosofi, No. 1, Vol. 5

Jensen, T. (2016). Let's make it personal! Ontological ethics in fan studies. Journal of Fandom Studies, 4(3), 255-275. [4]. https://doi.org/10.1386/jfs.4.3.255_1

Jensen, T., & Vistisen, P. (2013). Tent-Poles of the Bestseller: How Cross-media Storytelling can spin off a Mainstream Bestseller. Akademisk kvarter / Academic Quarter, 7, 237-248.

Latour, B. (2005). Reassembling the social: An introduction to actor-network-theory. Oxford, UK: Oxford University Press.

Løgstrup, K. E. (1997). The ethical demand. University of Notre Dame Pess.

Mejias, U. (2013). Off the network. Minneapolis: University of Minnesota Press.

Paßmann, J. (2018). Die soziale Logik des Likes: Eine Twitter-Ethnografie. Campus Verlag.

DISTINGUISHABLITY, INDISTINGUISHABILITY AND ROBOT ETHICS: CALLING THINGS BY THEIR RIGHT NAMES

Alexis Elder

University Of Minnesota (United States)

alexis.elder@gmail.com

EXTENDED ABSTRACT

In Turing's (in)famous Turing Test, he proposes that entities should be considered intelligent when they turn out to be indistinguishable, in conversation, from human interlocutors (Turing, 1950). Since then, the anticipated eventual development of artificial entities of sufficient complexity so as to pass such a test have garnered a great deal of attention by robot ethicists. From LaBossiere's "Testing the Moral Status of Artificial Beings" (LaBossiere, 2017), to Neely's "Machines and the Moral Community" (Neely, 2014), to Klein's "Robots Make Ethics Honest - And Vice Versa" (Klein, 2016), to Danaher's "A Philosophical Case for Robot Friendship" (Danaher, 2019), scholars have explored questions about what, if anything, would constitute good reason to think that robots are in some relevant sense morally *indistinguishable* from their human counterparts and thus deserving of attributions of moral status. Some, such as Turkle (Turkle, 2011), Bryson (Bryson, 2010) and Sharkey (Sharkey, 2014) have even worried that robotic appearances that are (in practice) indistinguishable from human ones will undermine our existing moral practices, and others, like Gunkel (Gunkel, 2018a) (Gunkel, 2018b), Coeckelbergh (Coeckelbergh, 2010) and Darling (Darling, 2015) have argued that the emotional responses we human beings experience when confronted with human-like artificial intelligence are themselves grounds for moral consideration, *even when* our cognitive evaluations do not (yet) support the claim that these robots are *indistinguishable*. And it has become a commonplace consideration to remember Masahiro Mori's famous posit that we will find robotics *more* emotionally engaging when they are *less* realistic approximations of humans, thus avoiding what he calls the "Uncanny Valley".

Questions about the ethics of distinguishing are thus found throughout robot ethics. Topics debated include what to do once robots are indistinguishable from human beings, what grounds we would have for distinguishing (for example) organic from synthetic beings when they are experientially indistinguishable, the difference between something's being emotionally and cognitively (in)distinguishable, and what follows from treating different entities as indistinguishable, all of which lead to the question: what, if anything, should be the relevant sense (or what are the relevant senses) of (in)distinguishability? And that, in turn, raises questions about the *point* of distinguishing, especially its moral significance.

In the *Analects*, the following exchange is recounted: Confucius was asked what he would do if he were given a leadership position, and he responded that he would direct his attention to the *rectification of names*: making sure that words describe the world truthfully (Confucius, 2003, 13.3). He took this to be a central ethical concern and a remedy for the political and social chaos toward which the warring Chinese states were prone during his lifetime. Although to some ears this might sound as if he was attempting to make sure that our (constructed) language accurately describes the (mind-independent) objective state of affairs, carving nature at its joints, the idea of the rectification of names took on a much more philosophically sophisticated role in Chinese philosophy. In fact, even in Confucius' original version, when "naming" someone to a social role, this was not merely a descriptive project but one that brought along a variety of social and ethical norms. To call someone a father, a sister, or a child

was to assign them a social role with attendant obligations and expectations, and the harmonious function of a community depended on people occupying these various roles both faithfully and sincerely. When one is not confident in the correct name for something, one should be cautious about invoking the power of naming, he argued:

> A superior man, in regard to what he does not know, shows a cautious reserve. If names be not correct, language is not in accordance with the truth of things. If language be not in accordance with the truth of things, affairs cannot be carried on to success. When affairs cannot be carried on to success, proprieties and music do not flourish. When proprieties and music do not flourish, punishments will not be properly awarded. When punishments are not properly awarded, the people do not know how to move hand or foot. Therefore a superior man considers it necessary that the names he uses may be spoken appropriately, and also that what he speaks may be carried out appropriately. What the superior man requires is just that in his words there may be nothing incorrect. (Confucius, 13.3.4-7)

The Mohist school of thought, often diametrically opposed to Confucian philosophy, took up this idea of correcting standards for identifying things and argued, against the elitist Confucians, that what we need are uniform, publicly accessible, and widely useable-by-individuals standards, or *fa*, that they compare to carpenter's squares and levels, tools for helping us to coordinate our moral judgments (Mozi, 2001). The aggressively anti-realist Daoists resisted the idea that there *are* any mind-independent standards for how to divide up the world and hence what to call things, rejecting Mohist "chop-logic" and attempts to carve up the world in one definite way and distinguish one thing from another, and in a particularly striking passage, comparing those who insist on drawing distinctions without a (real) difference (which the Daoist thinks is in principle impossible) to monkeys that demand to be fed three nuts in the morning and four at night, rather than four in the morning and three at night, but at the same time suggesting that one might as well acquiesce to these illogical preferences since it makes no difference anyway (Zhuangzi, 2001). And the Confucian scholar Xunzi, arguing against these Daoists, grants that distinction-drawing is artificial but argues that morality itself is a useful artificial tool for overcoming innate but problematic human tendencies, and so distinctions are themselves tools to help us to achieve cooperative prosperity and avoid strife, giving us a standard for *which* artificial distinctions are defensible and which are problematic (Xunzi, 2014).

The arguments and distinctions that arise from this classical Chinese discussion can helpfully inform current debates around the ethical significance of various kinds of distinguishability and indistinguishability in robot ethics. I aim to use these historical debates to inform current controversies around distinguishing and distinguishability in robot ethics. For example, Neely argues that we ought to resist our innate tendencies to distinguish "us" from them and be permissive in granting moral patiency, a reason *not* to distinguish of which Xunzi would approve. At the same time, Mohist concerns about the public applicability of distinguishing standards seem pertinent: today, chatbots stand in for human interactions in ordinary contexts like scheduling appointments and therapeutic social robots with powerful emotional impact are developed for cognitively compromised populations, it matters that one have publicly accessible *and usable* standards. The point of this paper is not to once and for all settle which account(s) of distinguishability is significant, but to develop resources for evaluating and discussing which distinctions can and should be drawn, as well as understanding philosophically what else is at stake when we draw them.

KEYWORDS: Robot ethics; Asian philosophy; moral status; Confucian ethics.

REFERENCES

Bryson, J. J. (2010). Robots should be slaves. *Close Engagements with Artificial Companions: Key Social, Psychological, Ethical and Design Issues*, 63–74.

Coeckelbergh, M. (2010). Robot rights? Towards a social-relational justification of moral consideration. *Ethics and Information Technology*, *12*(3), 209–221. https://doi.org/10.1007/s10676-010-9235-5

Confucius. (2003). *Analects: With Selections from Traditional Commentaries:* (E. G. Slingerland, Trans.). Retrieved from https://books.google.com/books?id=6DseYHSfaagC

Danaher, J. (2019). The Philosophical Case for Robot Friendship. *Journal of Posthuman Studies*, *3*(1), 5–24.

Darling, K. (2015). *'Who's Johnny?'Anthropomorphic Framing in Human-Robot Interaction, Integration, and Policy*.

Gunkel, D. J. (2018a). *Robot rights*. MIT Press.

Gunkel, D. J. (2018b). The other question: Can and should robots have rights? *Ethics and Information Technology*, *20*(2), 87–99. https://doi.org/10.1007/s10676-017-9442-4

Klein, W. (2016). Robots make ethics honest: And vice versa. *ACM SIGCAS Computers and Society*, *45*(3), 261–269. https://doi.org/10.1145/2874239.2874276

LaBossiere, M. (2017). Testing the Moral Status of Artificial Beings; Or "I'm going to ask you some questions..." In P. Lin, K. Abney, & R. Jenkins (Eds.), *Robot Ethics 2.0: From Autonomous Cars to Artificial Intelligence* (pp. 293–306). Oxford: Oxford University Press.

Mozi. (2001). Mozi. In P. J. Ivanhoe (Trans.), *Readings in Classical Chinese Philosophy* (2nd ed., pp. 59–114). Indianapolis, Indiana: Hackett Publishing Company.

Neely, E. L. (2014). Machines and the moral community. *Philosophy & Technology*, *27*(1), 97–111.

Sharkey, A. (2014). Robots and human dignity: A consideration of the effects of robot care on the dignity of older people. *Ethics and Information Technology*, *16*(1), 63–75. https://doi.org/10.1007/s10676-014-9338-5

Turing, A. M. (1950). Computing Machinery and Intelligence. *Mind*, *49*, 433–460.

Turing, A. M. (2004). *The essential turing*. Oxford University Press.

Turkle, S. (2011). *Alone Together: Why We Expect More from Technology and Less from Each Other*. Retrieved from https://books.google.com/books?id=J2ine5sIIkgC

Xunzi. (2014). Correct Naming. In E. L. Hutton (Trans.), *Xunzi: The Conplete Text* (pp. 236–247). Princeton University Press.

Zhuangzi. (2001). *Chuang-tzŭ: The inner chapters* (A. C. Graham, Trans.). Indianapolis: Hackett Pub. Co.

FOR OR AGAINST PROGRESS?: INSTITUTIONAL AGENCY IN A TIME OF TECHNOLOGICAL EXCEPTIONALISM

You Jeen Ha

Smith College (U.S.A.)

yha@smith.edu

EXTENDED ABSTRACT

Popular press often touts the notion that legislation and governance will always be "playing catch up" to technology, that technology outpaces regulation, as if it were conventional wisdom. This occurrence is not merely limited to daily news headlines. In legal-ethical literature, a recurring critique of current American legal institutions and their response rate with respect to technological innovation has been referred to as "the pacing problem", especially by Marchant (2011). The critique assumes that the arenas of governance and technology have inherently fixed characteristics: the former can't help but be slower than the latter, and this difference will consistently hold unless new regulatory frameworks and approaches to emerging technologies are introduced. However, what this assumption implicitly does is establish a *status quo* in which at the intersection of governance and emerging technologies, the central concern is catching governance up to the technologies. Perhaps, this view is true to some degree if we observe the tech policy landscape today. As we head into the second decade of the twenty-first century, we have artificial intelligence systems developed and used by corporations and the military for various purposes. These purposes range from mass advertising and surveillance to international reconnaissance and warfare. Yet, we have virtually no laws nor ethical standards that actively regulate these systems, from their creation to their impacts on the public.

I argue that feeding into "the pacing problem" are subscriptions to three ideas: (a) institutional agency, which I understand as the agency of members in a collective group based on philosopher Kirk Ludwig's (2017) work, such that agency is event-causal, (b) Moore's Law, which states that the speed and capability of our computers are expected to increase every few years, suggesting that the rate of technological advancement will only increase, and (c) technological exceptionalism, which presumes that technological means can solve the world's complex sociopolitical and economic problems. These ideas together assume certain metaphysical and epistemological commitments, requiring a distinction between agent causation and event causation, with the former grounding my view of institutional agency, and consequently that the term *agency* denotes an epistemic capacity of human subjects. In the status quo, technology firms thus are not necessarily collectives with human agents of change, but instead are event-causal parts of a burgeoning collective industry capable of "uprooting" our existing institutions, as legal scholar Meg Leta Jones (2017) puts it.

Moreover, we may encounter in the language used within the technology policy and ethics sphere a different kind of agency: rather than express technology firms as collectives of human agents, one might be tempted to express the technologies themselves as agent-causal rather than event-causal. Event-causation is attributed more generally to inanimate objects since the causation of events by inanimate objects is always reducible to the causation of those events by other events involving those objects. This conflation in understandings of agency has the potential to occur more frequently as technologies such as artificial intelligence advance with the expectation that perhaps one day, they will exhibit human-level intelligence and rationality. By expressing these technologies as such, we may

be unintentionally subscribing to a more concerning commitment: material agency, or that material entities, ontologically and epistemologically, have the quality of agency, according to Kirchoff (2009).

Insofar as we take emerging technologies to be inanimate artifacts that are also material agents, we will detract from discussions concerning how we, as human members of a society, can correct or at the very least reevaluate our institutional practices with respect to the pacing problem. Instead, we will be entering a world of hypotheticals in which artifacts are metaphysically and epistemologically similar to human agents, especially if we take artificial intelligence as an example, an instance of technology that has not reached human-level cognitive capacities. It is my position that we avoid this world of hypotheticals and focus on immediate issues at hand. To do so, we would commit both to agent causation rather than to event causation and to the denotation of agency as a human epistemic capacity. As a result, we would view technological development as caused by human agents who constitute the technology firms and industry rather than as the result of a larger event-causal, material agent called “technology”.

Moreover, by making these commitments, we would be able to incorporate into conversations in light of the pacing problem our more explicitly agent-causal, human institutions in law and regulation. These very systems are indeed collectives of human agents as jurisprudence, laws, and court decisions are preserved, signed into existence, and upheld by scholars, councils, and governments. Thus, our legal institutions are institutional agents. So if we return to the pacing problem itself, which frames institutions of governance as inherently slower than developments in technology, attributing intrinsic qualities to each group puts us in a bind, implying that there is no way to solve a problem that is fundamentally fixed and predetermined. If those involved in technology ethics, policy, and governance as well as the public continue to accept the assumptions that underlie the pacing problem, then they may be tying themselves to a never-ending pursuit toward catching governance up with technological innovation. The “Move Fast and Break Things” mantra that pervaded the tech innovation sphere for years in the name of “progress” may have only exacerbated the sense of urgency associated with catching up. On the other hand, it may be the case that we work towards *slowing down* technological development rather than speed up the rate of governance by focusing on how companies develop and release their products. This approach however still assumes that innovation is innately faster than governance and that we should consequently fix the former. Moreover, this approach tends to be frowned upon as a hindrance to progress.

So, rather than treat the pacing problem as if it were merely a verifiable statement, I instead question the normative nature of the pacing problem and argue that we might be better off treating and viewing our institutions of governance and technology firms in industry as collectives of agent-causal members exercising agency. It would then follow that the pacing problem’s fixed presuppositions about governance and technological development disappear once we conceive of the problem as one involving dynamic institutional agents of change instead of one involving static actors in a rigid status quo. There is no pacing problem as we see it today if we do not accept its foundational assumptions. Although figures such as Moor (1985), who famously claimed that "computing will transform social institutions," have suggested an alternative hypothesis, namely that the pacing problem is not an issue caused by treating government and institutions as having inherently fixed characteristics, we perhaps still may be unable to dismiss the role of human agents in creating the very systems, structures, and traditions that sustain modern society. Ultimately, the onus is on the human individuals who comprise the institutional agents to recognize and stop their contributions to perpetuating what is conventionally believed to be a systemic conundrum. By challenging conventional wisdom, we can dispel anxieties about updating our legal and regulatory frameworks and define our relationship to technologies moving forward.

KEYWORDS: Pacing problem, technological exceptionalism, institutional agency, agent and event causation, material agency.

REFERENCES

Bratman, M. E. (2018, August 02). From Plural to Institutional Agency: Collective Action II // Reviews // Notre Dame Philosophical Reviews // University of Notre Dame. *Nortre Dame Philosophical Reviews.* Retrieved from https://ndpr.nd.edu/news/from-plural-to-institutional-agency-collective-action-ii/

Burn-Murdoch, J. (2013, April 12). Data protection law is in danger of lagging behind technological change. *The Guardian.* Retrieved from https://www.theguardian.com/news/datablog/2013/apr/12/data-protection-law-lagging-behind-technology

Hirschmann, D. (1971). Inanimate Agency. Proceedings of the Aristotelian Society, 72, 195-213. Retrieved from https://www.jstor.org/stable/4544824?seq=1#metadata_info_tab_contents

Jones, M. L. (2017). Does Technology Drive Law? The Dilemma of Technological Exceptionalism in Cyberlaw. Journal of Law, Technology & Policy, 2018(2), 102-137. Retrieved from https://ssrn.com/abstract=2981855.

Kirchoff, M. D. (2009). Material Agency: A Theoretical Framework for Ascribing Agency to Material Culture. Techné, 13(3), 205-219. Retrieved from https://scholar.lib.vt.edu/ejournals/SPT/v13n3/pdf/kirchhoff.pdf

Lowe, E. J. (2001). Event Causation and Agent Causation. Grazer Philosophische Studien, 61(1), 1-20.

Ludwig, K. (2017). *From Plural to Institutional Agency: Collective Action II.* Oxford: Oxford University Press.

Marchant, G. E. (2011). [Chapter 13: Addressing the Pacing Problem]. In G. E. Marchant, B. R. Allenby & J. R. Herkert (Eds.), *The Growing Gap Between Emerging Technologies and Legal-Ethical Oversight: The Pacing Problem* (pp. 199-205). Heidelberg, Germany: Springer Science+Business Media.

Moor, J. H. (1985). What is Computer Ethics? Metaphilosophy, 16(4), 266-275.

Taneja, H. (2019, January 22). The Era of “Move Fast and Break Things” is Over. *Harvard Business Review.* Retrieved from https://hbr.org/2019/01/the-era-of-move-fast-and-break-things-is-over

FROM JUST CONSEQUENTIALISM TO INTENTIONAL CONSEQUENTIALISM IN COMPUTING

Kimppa, Kai K.

University of Turku (Finland)

kai.kimppa@utu.fi

EXTENDED ABSTRACT

This paper starts with a misunderstanding: In early 2000s I found James Moor's paper from 1999 on Just consequentialism and computing, which I apparently did not read as well as I should have, and for over a decade misunderstood what Moor meant with deontological in this context. The paper refers to Rawls (1971), not to Kant (1785/1970), as I had thought, and I created in my head a picture on how the paper would approach deontological consequentialist model. The actual model in Moor's paper goes as follows: IF a system designed is both *Just*, in a Rawlsian sense, AND the consequences from the system are good, THEN the system is well designed – from an ethical perspective, and the designers have done morally right! Now, on the other hand, if the system is unjust and the consequences are bad, the system is definitely bad – again, from an ethical perspective, and again, the designers have not done well. Unfortunately, it is much harder to get know what kind of an answer the model gives if the system is either just but the consequences none the less are bad, or whether the system's consequences are good, but the system itself is unjust. Examples of these would be the system dealing results out justly, but making less in total, or the consequences being on average good, but spreading unjustly to those the system affects (see Table 1).

Now I had taken the paper to talk about *deontological consequentialism*, based on Kant (1785/1970) and Mill (1863), and not on Rawls and Mill, as it actually was. Thus, I came up with the idea, that what the model looks at is *intent*, and consequences, not *justness* and consequences. Thus, practically accidentally, I came up with a different model (Table 2), in which if the intent is good and the consequences are good, the system is morally good and the designers have put in the effort they should. Now, if the intent is bad (or evil), the purpose and thus the designers of the system are always bad – no matter what the consequences, as with evil intent the *accidentally* good consequences do not justify the goodness of the system. Of course, an accidentally good system can still be used by others for good purposes, but the designers have in any case not done their professional duty. Examples of these kinds of systems would be for example total surveillance systems of your own citizens – or those of your allies. It does not really matter what the consequences of these systems are (preventing terrorism, for example – although very little evidence for this exists), the *intent* alone is *evil*, and thus the system is a bad system. Now, my proposed system solves one more square in the two-by-two, but the good intent bad consequences still remains at least not easy to answer.

In the full paper I plan to delve deeper in these topics and give more examples, but the gist of the paper is already presented in this extended abstract.

Table 1 Just consequentialism

		Consequences	
		Good	Bad
Justice	Yes	**Ok!**	**?**
	No	**?**	**Not ok!**

Modified from Moor (1999)

Table 2 Intentional consequentialism

		Consequences	
		Good	Bad
Intent	Good	**Ok!**	**?**
	Bad	**Not ok!**	**Not ok!**

KEYWORDS: Justice, Deontology, Consequentialism, Model, System Design.

REFERENCES

Kant I. (1785/2002) *Groundwork for the Metaphysics of Morals* (translated by Wood, A. W), New Have and London, Yale University Press.

Mill, J. S. (1863) Utilitarianism, https://www.utilitarianism.com/mill1.htm, accessed 21.10.2019.

Moor, J. H. (1999) Just consequentialism and computing, *Ethics and Information Technology*, **1:** 65-69.

Rawls J. (1971). *A Theory of Justice,* Cambridge, Massachusetts: Belknap Press of Harvard University Press.

SELF-RELIANCE: THE NEGLECTED VIRTUE TO HEAL WHAT AILS US

Richard Volkman

Southern Connecticut State University (USA)

volkmanr1@southernct.edu

EXTENDED ABSTRACT

When the internet was young, many scholars expressed hope that a "global information ethics" might emerge to make sense of our new cultural reality in "cyberspace," much as liberal and individualist ethical theories of the early modern era emerged to makes sense of the new cultural reality in the wake of the printing press revolution (Gorniak, 1996). Over two decades later, we observe instead an online world fractured and fractious with tribal rage (e.g., see Nagle, 2017; Phillips, 2015; Soave, 2019). Sunstein (2019) identifies an important source of all this in conformity: "We live in an era of tribalism, polarization, and intense social division…the key to making sense of living in this fractured world lies in understanding the idea of conformity." In his celebrated essay of 1841, Emerson declares, "The virtue in most request is conformity. Self-reliance is its aversion." The paper I propose for Ethicomp2020 will argue that fostering Emersonian self-reliance as a core virtue of the global information ethic we aspire to is the best way to reduce the acrimonious divisions that characterize online life today, whether those divisions are ultimately a consequence of social pressure, pluralism with respect to genuine objective values, or a looming nihilism empowering petty tyrants and trolls and their shallow, conformist followers.

Contemporary online discord, increasingly spilling into the offline world (e.g., see Iyengar and Westwood, 2015), suggests profound implications for meta-ethics. Insofar as our online world establishes a truly vast and open marketplace of ideas, we should expect it will serve as a powerful information processing technology that increases our ability to discern truth. The arguments for this view famously go back at least to Mill (1859), and they have been extended, amplified, and applied directly to our contemporary circumstance in terms of "Wisdom of Crowds" technologies (Suroweicki, 2005). If there is some single ethical reality for us to converge upon, then the aggregation of our diverse, independent, and decentralized judgments should lead to a greater convergence of ethical opinion with the advent of the internet and social media. Instead, we get greater polarization, tribalism, filter bubbles, echo chambers, and the like. This suggests either there is no single ethical reality upon which we might converge, or our present Wisdom of Crowds technologies (e.g., the online marketplace of ideas) are somehow broken or processing bad inputs. Encouraging self-reliance with respect to belief and discourse as a core virtue addresses both possibilities. If our failure to converge on some single set of ethical truths is a consequence of psychological pressures to conform to our ingroups, subverting the independence, diversity, and decentralization of the judgments we contribute to discussion, then self-reliance directly reestablishes the conditions of a successful Wisdom of Crowds technology. On the other hand, if the truth in meta-ethics is that some deep pluralism or error theory makes convergence impossible, since there is no single ethical reality upon which to converge, then self-reliance remains as the ultimate ground of an ethical individualism that affirms multiple ethical realities, leaving the number and diversity of those realities an empirical matter.

Self-reliance is a neglected and misunderstood virtue, so a considerable part of the paper will be devoted to saying what self-reliance is and especially what it is not. Self-reliance sometimes refers to human activity generally in the spirit of a "do-it-yourself" ethic, but I will follow Emerson in advocating

merely for self-reliance with respect to belief formation and discourse. The self-reliance I am discussing has nothing to do with "pulling one's self up by the bootstraps" or eschewing assistance from others. It is rather the virtue associated with thinking for one's self and boldly proclaiming before others one's own actual opinions and reasons in the utmost sincerity and with the utmost commitment to speaking the truth as one sees it. Since my own actual opinions and reasons are inevitably and by my own lights justly informed in discourse with others past and present, living and dead, there is nothing in self-reliance that eschews such discourse. To the contrary, my commitment to truth leads me to seek out diverse opinions to edify and sometimes revise my own honest assessments. There is nothing intrinsically anti-social or dismissive of community in the affirmation of self-reliance. As Emerson writes, "It is easy in the world to live after the world's opinion; it is easy in solitude to live after our own. But the great man is he who in the midst of the crowd keeps with perfect sweetness the independence of solitude." Self-reliance does not mean one is automatically right; it just means one remains in contact with that perspective on reality which it is one's unique responsibility to contribute to the conversation. Failures of self-reliance impoverish our understanding of the world, since information about the world is distributed across diverse individuals with diverse perspectives. At the same time, a community of self-reliant individuals may or may not come to agree about a great many things and possibly even about everything. Our agreement indicates no base conformity unless its motive is one's desire to flatter or submit to the authority of others. Self-reliance is not about being different for the sake of being different, which would itself be a failure of self-reliance insofar as its motive is one's reputation before others. We can all be non-conformists without thereby being conformist in our non-conformity. If we each come to our own views in honesty and diligence, we are each conforming to our own visions of the truth, even if it should turn out that our independent visions are in perfect agreement. This account of self-reliance will be further clarified by conceiving it in an Aristotelian frame as the virtue with respect to confident assertiveness regarding one's own opinion that lies at the mean between the deficiency of conformity and the excess of closed-minded stubbornness.

Considered in this light, self-reliance is already widely accepted as a fundamental virtue among scholars. The AAUP Statement on Professional Ethics begins, "[Professors] primary responsibility to their subject is to seek and to state the truth as they see it." But self-reliance is extremely hard. We are profoundly social animals with an oversized concern for our reputations and standing amongst our peers. For most of us, self-reliance does not come naturally. It has to be nurtured and promoted in the culture. Fortunately, as Sunstein documents, it only takes a few self-reliant individuals to make an enormous difference for how self-reliant others in the group are. This suggests we can indeed have a more self-reliant culture if only a few of us will manifest the courage to set the example. If we can convince our peers that it is safe and wise to manifest self-reliance by their own lights, we can significantly upgrade the functioning of our best information processing technology for discovering ethical truth (viz., the marketplace of ideas).

Perhaps a more self-reliant culture would issue in a global information ethics that finally discovers for us a single set of ethical truths for all humans. It seems more likely instead that it will reveal a great diversity in human aspirations and temperaments speaking to a deep pluralism perhaps extending all the way to ethical individualism. In that case, I will argue that each individual's self-reliant vision of his or her highest aspirations serves as the sole and ultimate ground of ethical authority, which seems to have been Emerson's own view. This view is sometimes wrongly dismissed as a shallow subjectivism, but that is a criticism Emerson anticipated and addressed, declaring, "If any one imagines that this law is lax, let him keep its commandment one day." There is a fact about who I most aspire to become in light of my own assessment of the evidence, and self-reliance requires one to honor that fact despite

temptations to base conformity, mindless pleasure, selfish indulgence, or whatever else would distract one from the fact of one's highest aspirations.

Self-reliance is hard for social animals like us, but there is reason to believe it is the virtue that can heal what ails us in the information age, whether the source of our troubles lies in our base psychology or is a symptom of a deeper meta-ethical anxiety.

KEYWORDS: wisdom of crowds, individualism, self-reliance, tribalism, conformity, marketplace of ideas.

REFERENCES

American Association of University Professors (2009). Statement on Professional Ethics. Retrieved from https://www.aaup.org/reports-publications/aaup-policies-reports/standing-committee/ethics.

Emerson, R. W. (1841) Self-Reliance. Retrieved from http://www.emersoncentral.com/selfreliance.htm.

Gorniak-Kocikowska, K. (1996). The computer revolution and the problem of global ethics. *Science and Engineering Ethics*, *2*(2), 177–190.

Iyengar, S., Westwood, S. (2015). Fear and Loathing across Party Lines: New Evidence on Group Polarization. *American Journal of Political Science,* 59(3), 690-707.

Mill, J. S. (1859) *On Liberty*. Retrieved from http://www.utilitarianism.com/ol/one.html.

Nagle, A. (2017). *Kill All Normies*. Alresford, UK: Zero Books.

Phillips, W. (2015). *This is Why We Can't Have Nice Things*. Cambridge, MA: MIT Press.

Sunstein, C. R. (2019). *Conformity: the power of social influences*. New York: New York University Press.

Soave, R. (2019). *Panic Attack: Young Radicals in the Age of Trump*. New York: All Points Books.

Suroweicki, J. (2005) *The Wisdom of Crowds*. New York: Anchor.

THREE ARGUMENTS FOR "RESPONSIBLE USERS". AI ETHICS FOR ORDINARY PEOPLE

Pak-Hang Wong

Department Of Informatics, Universität Hamburg (Germany)

Wong@Informatik.Uni-Hamburg.De

EXTENDED ABSTRACT

Our daily life is increasingly mediated—or, even structured—by Artificial Intelligence (AI) systems, by which I refer to the combination of various machine learning algorithms with (big) data collection and analysis techniques. AI systems have been used in predictive policing, in credit scoring, and in everyday mundane decisions, e.g. recommendations on books, films, music, etc. Whatever benefits AI systems may offer to individuals, numerous events in the past few years have demonstrated the possible grave consequences from its (mis)uses (see, e.g. O'Neil 2016). For example, the COMPAS recidivism algorithm, along with other algorithmic applications, are condemned as discriminatory, and they have led to serious individual and collective harm (see, e.g. Angwin et al. 2016). Similarly, Cambridge Analytica, and other disseminators of misinformation, are blamed for the demise of knowledge and democracy. Responding to these potentially harmful effects and consequences, researchers, practitioners, and policymakers have sought to critically examine the design and application of AI systems and explore various social, technological, and regulatory solutions to ensure AI systems are created and used for human good. The existing discourse, however, is either directed at the AI companies (in the form of legislation and regulation) or at AI researchers and developers (in the form of code of ethics and best practice policy). Legislation, regulation, code of ethics, and best practice policy are certainly important instruments to ensure *better* and *more responsible* design and use of AI, but these endeavors often neglect—or, even exclude—the consideration of the *downstream* of AI systems, i.e. the *users*. Particularly, the discourse neglects *individual* users (not *corporate* users) of AI systems, who are often the subjects of data collection/analysis or of algorithms.

The existing discussion ignores users of AI and only focuses on AI companies and AI researchers and developers, as it is centered on a "moral harm paradigm", in which users are often construed as *passive victims* of AI and thus are not responsible for the harm created by and related to AI systems (Magalhães 2018). The difficulty to include users in critical reflection is compounded by the problem of responsibility gap, which states that we are rarely in control of many machine(-mediated) actions, and thus these *autonomous* actions by machines pose a threat to our ordinary understanding of moral responsibility because no human being seems to be (morally) responsible for them (Matthias 2004). It is already difficult, according to the problem of responsibility gap, to ascribe moral responsibility to developers and operators of machines, it would be even more difficult to ascribe moral responsibility to users of AI. Moreover, the inclusion of users of AI in critical reflection may extend the "moral crumple zone", where moral and legal responsibility arising in the context of new and complex AI systems are misattributed to them to protect the integrity of AI systems (Elish 2019).

The danger of extending the "moral crumple zone" is certainly real, and the problem of responsibility gap does indeed present a serious challenge to the inclusion of AI users in our critical reflection, but they should *not* stop us from considering the role and significance of AI users in creating and maintaining better AI systems. The role and significance of AI users become obvious once we begin to take AI systems as a set of continuing processes of input-encoded procedures-output, as Mike Annany

notes, "[AI systems] are embedded within the sociotechnical structures; they are shaped by communities of practice, embodied in standards [...], the relevance, quality, and stability of algorithms depend upon end users" (2015, 98). Users of AI at least *partially* determine the outputs of AI systems by *being* inputs of the systems. If it is the case, users are not—and should not—be viewed merely as *passive victims* of AI systems, and they might be said to be *responsible* for the negative effects and consequences of AI systems.

Moreover, AI systems are increasingly being used in decision-making that has serious impact on *resource distribution*, e.g., crediting scoring, college admission, etc., which, in turn, creating 'winners' and 'losers' in the society. The patterns of winners and losers, however, are expected to resemble the *existing* patterns of social injustice, as AI systems depends on social and historical data that biased by social and historical circumstances. In short, the winners often benefit *unjustly* from the losers *via* AI systems; and, the users in continuing to use biased AI systems, they can be viewed as complicit in causing and perpetuating injustice. It is in this sense users of AI should be held responsible for the injustice of AI systems.

I elaborate on *two* arguments for the moral responsibility of users of AI systems based on (a) the structural role of AI system's users in determining the outcome and (b) the wrongful benefits users may receive from biased AI systems. For (a), I discuss the works of philosopher and political theorist Iris M. Young (2011), who developed and defended the idea of structural injustice and the social connection model of (political) responsibility. Young argues that structural injustice exists insofar as individuals (or groups of individuals) are systematically disadvantaged by a set of prevalent norms and/or circumstances; and, precisely because the problem is structural, no individual could be blamed for it. She then proposes an alternative model of responsibility, namely the social connection (political) model of responsibility to account for structural injustice. Here, I argue that negative effects and consequences of AI systems are forms of structural injustice, and I illustrate why users of AI systems have political responsibility based on Young's social connection model of responsibility. For (b), I draw on Garrath Willams' (2019) recent argument for collective harm based on individuals' complicity to further strengthen the case that users of AI are morally responsible for the negative effects and consequences of AI systems.

Finally, and additionally, since AI systems are, in Luciano Floridi's (2014) term, *re-ontologizing technologies* as they radically transform our self-understanding, I develop a *third* argument for user responsibility based on the new understanding of the self from the introduction and implementation of AI systems. This (c) ontological argument for user responsibility defends the view that the self is inherently *relational* and *interdependent with other users*, and so one's decisions, as mediated or structured by AI systems, are *always* at the same time self-regarding and other-regarding, and therefore one ought to *take responsibility* for the effects and consequences of AI systems. Together, the three arguments ground the need to reflect more critically the role of users of AI systems.

KEYWORDS: Responsibility, Artificial Intelligence, Users, Justice, Algorithmic Harm, Ethics of Technology.

REFERENCES

Ananny, M. (2015). Toward an Ethics of Algorithms: Convening, Observation, Probability, and Timeliness. *Science, Technology, & Human Values*, 41(1), 93-117.

Angwin, J. Larson, J. Mattu, S. Kirchner, L. (2016) Machine Bias. *ProPublica*, May 23, 2016. Retrieved from https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Elish, M. C. (2019). Moral Crumple Zones: Cautionary Tales in Human-Robot Interaction. *Engaging Science, Technology, and Society*, 5, 40-60.

Floridi, L. (2014). *The Fourth Revolution: How the Infosphere is Reshaping Human Reality*. New York: Oxford

Matthias, A. (2004). The Responsibility Gap: Ascribing Responsibility for the Actions of Learning Automata. *Ethics and Information Technology*, 6(3), 175–183.

Magalhães, J. C. (2018). Do Algorithms Shape Character? Considering Algorithmic Ethical Subjectivation. Social Media+Society, 4(2). Retrieved from https://doi.org/10.1177/2056305118768301

O'Neil, C. (2016). Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. New York: Crown.

Williams, G. (2019). The Social Creation of Morality and Complicity in Collective Harms: A Kantian Account, *Journal of Applied Philosophy*, 36(3), 457-470.

Young, I. M. (2011). *Responsibility for Justic*e. New York: Oxford University Press.

VIRTUE, CAPABILITY AND CARE: BEYOND THE CONSEQUENTIALIST IMAGINARY

Alison Powell, Funda Ustek-Spilda, Irina Shklovski

London School of Economics, London School of Economics, IT Unviersity of Copenhagen

a.powell@lse.ac.uk; f.ustek-spilda@lse.ac.uk; irsh@itu.dk

EXTENDED ABSTRACT

The consequentialist ethical tradition, where the 'goodness' of decisions is assessed in relation to their measurable arguments, is often applied to reflections by technologists on the responsibility for creating new technologies such as artificial intelligence (AI), machine learning (ML) or connected systems (i.e. Internet of Things (IoT)). MIT Media Lab's *Moral Machine* experiment, for example, approached concerns about ethics of connected vehicle systems by listing a series of moral conundrums that these systems are likely to encounter, and propose a 'universal machine ethics'. This experiment, not only illustrates the challenges in universalising ethics through a consequentialist imaginary where the focus is solely on maximising utility whilst minimising harm; but also serves to stresses the rational-individual as the key decision-maker in a solution-oriented mode of technology development. As such, the individual is expected to make structured decisions which would ultimately favour sparing the lives of humans over animals; a higher number of people over fewer; young people over the elderly and fit people over the un-fit and so on. Such thinking however leaves little room if any to attend to the contexts, structures and conditions that enable certain outcomes while removing others from being viable options; and rests on the fundamental assumption that the value of any decision is determined vis-à-vis its outcome.

In this paper, we intend to shift the focus from approaches that assess the morality of decisions made in the context of technology based on their actual or potential outcomes; and instead explore ethical theories that look at individual decision-making within the particularities of contexts, technologies used and relationships individuals are part of. Our approach is a result of our ongoing work with technology developers in the Internet of Things space in Europe where we have been quantitatively and qualitatively following the networks of developers, designers and entrepreneurs in order to understand the values that guide their decision-making during the design, development and marketing of their IoT products. As part of a large consortium of researchers, we have conducted fieldwork in Europe, followed hackathons, accelerator programmes, software and hardware showcases as well as immersed ourselves in meetup groups, and conducted interviews. We have also followed networks of IoT developers through online platforms such as Slack channels, Twitter and meetup. It is based on our ongoing research into the field of IoT that we developed the practical ethical framework that we present in this paper. In this extended abstract, we provide a short description of the ethical approaches that we are drawing upon for a broadly applicable ethical framework for practical ethics in technology design.

Virtue ethics claims that there is a kind of 'final good' which represents the desirable aims of someone's life, and against which these aims can be evaluated. All questions attached to right action are assessed against this final good - known as eudaimonia. This means focusing on excellence, virtue, and eudaimonia, instead of duty, rights, and obligations, which were the typical concerns of popular consequentialist and deontological approaches. More recently Vallor (2016) applied a version of virtue ethics to the problems of technology, calling for a concerted collective effort to develop "technomoral

virtues" that can guide the nature and direction of technical innovation in this rapidly changing world to ensure human flourishing. Virtue ethics draws with significant concern on the moral action of the individual and the role of community. Such an approach also offers a methodological opportunity to justify engagement with individuals and their articulations of values and principles as a legitimate pursuit. Yet in terms of identifying values, virtue ethics presents an interesting challenge. We have identified that the social milieu of (especially commercial) IoT development provides many constraints to people's ability to act in ways that they might think of as ethical (Author's paper). In particular, the idea of competing in a market or being subject to market pressures provides a particular constraint, which some people talk about transcending through their own personal work or actions or through the creation of alternative organizational structures such as technology trustmarks or manifestos. Part of the difficulty with virtue ethics however, is precisely its tendency to individualize the responsibility for virtuous action even if there is a role for communities in this process. According to MacIntyre (2007), a virtuous agent knows the correct way to act in various contexts while also desiring to act in such a way. This, however, is easier said than done, as several developers we have interviewed told us. They have also indicated that when pressed with immediate challenges, it is not always straightforward to foresee what is to come and what kind of implications their decisions might have in the long run; or whether their decisions align with the values they hold as individuals.

In trying to understand how ethics manifest as values in action in the contexts of hierarchy and power, we have been increasingly concerned with the questions of what leads some individuals/groups to choose to act in a certain way and what might shape or constrain that choice of action. One important attempt to elaborate on this question has been provided by Amartya Sen in his capabilities approach (Hesmondhalgh 2017). Sen (1999) explains that "a person's 'capability' refers to the alternative combinations of functionings that are feasible for her to achieve. Capability is thus a kind of freedom to achieve alternative functioning combinations." This means that paying to attention to individual's internal capabilities is insufficient and we must also consider the possibilities created by a combination of internal capabilities and the structural conditions defined by the particular social, economic and political environment within which the individual attempts to act. This recognition that personal principles may need to be compromised to cope with structural constraints point to the importance of understanding what these constraints are and what influence they might exert. Furthermore, technology developers are in a curious position of both having to make decisions within the structural constraints of their context and having to acknowledge that the design decisions they make will result in producing structural constraints and possibilities for their users. Thus for developers to "do good" it is important to not only evaluate how existing constraints affect design but also to consider how these constraints are translated into the design and how these might be mitigated to offer more or different possibilities to the users.

The capabilities framework augments the internally oriented focus of virtue ethics on the moral capacities of the individual, by adding the importance of structural constraints. However, in both of these philosophical approaches decisions are made by individuals (even if within a social milieu) and it is individuals that must take responsibility, accounting for the constraints imposed by the broader social, political and economic contexts. Developers and designers of IoT technologies, just like everyone else, are certainly not alone in making decisions and in facing the consequences. Thus, we bring in *care ethics* to account for the value stemming from relational practice in considering different points of view as well as the possibilities of negotiating conflicts that arise between them. This has the enables including different points of view than the dominant discourses; such as those made by women or marginalized people who have not been part of the ethical discussion otherwise, and also for considering the ethics of practices such as caring which may have been absent from other readings.

Joan Tronto (1993), for example, rejects essentialisms in gender and moral thought and advocates for contingent and historically situated definitions.

Individuals are always entangled in a diversity of relations that hold contradictory values and conflicting demands. In this paper, we bring these differing and at times conflicting demands in focus to illustrate both the complexity of the contexts in which decisions about emerging technologies are made and acted upon; but also how rather than the consequences, the infrastructures, relations and individual and community values shape the way these decisions come to be made. As such, we hope to provide an actionable and practical ethical framework for technology design and development that brings an alternative to the overly-dominant discussions of emerging technologies based on their potential [dystopian] outcomes.

KEYWORDS: ethics & technology, virtue, capability, care, consequentialism.

REFERENCES

Annas, J., 1993. The morality of happiness. Oxford University Press. New York; Oxford.

MacIntyre, A., 2007. After Virtue: A Study in Moral Theory, 3rd ed. University of Notre Dame Press, Notre Dame, Indiana.

MIT Media Lab. (2017). Moral Machine. http://moralmachine.mit.edu/ Accessed 10 October 2019.

Tronto, J.C., 1993. Moral boundaries: A political argument for an ethic of care. Routledge, Chapman and Hall Inc., London.

Vallor, S., 2016. Technology and the virtues: A philosophical guide to a future worth wanting. Oxford University Press, United States of America.

WHAT IS VECTOR UTILITARIANISM

Wilhelm E. J. Klein

Zeta Motion Ltd. (Germany)

mail@wilhelmklein.net

EXTENDED ABSTRACT

The aim of the proposed paper is to introduce and outline Vector Utilitarianism and explain why I believe it represents a valuable contribution to the existing landscape of ethical frameworks. In previous papers and presentations I have very strongly argued against exceptionalisms in the ethics of humans, animals and beyond. Further I have advocated for yet another expansion of the circle of ethical consideration. The idea I have put forward previously was that any a priori notion of "specialness' and exceptionalism should be reflected before the background of what we have learned about ourselves and the way our moral judgement has evolved and works today. And what we have learned here - which is a very uncomfortable realisation to come to (Floridi, 2014, ch.4) - is that there is nothing inherently special about us. Like all the other animals, plants and other living organisms, we are simply the walking, talking vehicles for mindless and 'selfish' replicators (Dawkins, 2006). One dimension deeper, we're nothing but elaborate, temporary assemblages of various particles and their respective natural properties as dictated by the laws of nature (Carroll, 2016). From this perspective of scientific naturalism, the aim of Vector Utilitarianism is to operate without any of the notions of ethical significance which are incompatible with this ontological view and to present a scaffolding to construct an alternative version of utilitarianism which is able to accommodate any entity known and yet to be discovered within our natural world.

To introduce the general argument, the paper will introduce a variation of a short story by Elezier Yudkowsky titled Three Worlds Collide (Yudkowsky, 2009). My version will be shortened on one hand and augmented with one more colliding 'world'. In this thought experiment, we find ourselves in a future where humanity has progressed dramatically in terms of technology and is now roaming the galaxy and exploring new worlds. Socially, it still adheres to more or less the same moral standards as contemporary humanity and has not (yet) altered itself dramatically in terms of both biology and culture. What these future humans have not yet encountered, however, is aliens – a fact about to change in the course of this thought experiment. These alternative life forms which I will introduce, exhibit radically different bodies, 'minds', values and ways of 'living' than humans do. Nevertheless, they need to find some sort of common ground, or in other words an ethical approach fit to deal with all their differences. In many ways, I will argue, this mirrors our current situation, where in our technological present and future, we need to be able to deal with entities of radical difference, possessing and exhibiting types of life, consciousness, and communication completely outside the scope of everything we have biologically and culturally evolved to intuit or relate to (Greene, 2013; De Waal, 2014).

The most important next barrier to transcend in terms of the current boundaries of the circle of ethical concern, I will argue, is "familiar consciousness", as this seems to be at the basis even of some of the most progressive and otherwise quite naturalistic approaches. Consider Peter Singer's position, for example, who argues that to anchor the lowest common denominator for ethical significance on consciousness or is defensible because it rests on "familiar ground" (Singer, 2011, p.248). He states: "The question 'What is it like to be a possum drowning?' at least makes sense, even if it is impossible

for us to give a more precise answer than 'It must be horrible'. [. . .] There is, however, nothing that corresponds to what it is like to be a tree dying because its roots have been flooded", he states (ibid., p.248).

To demarcate one kind behaviour from another one like this, however, I believe represents a mere act of Dennett'ian 'argument from intuition' and familiarity (Dennett, 1999). From the perspective I am putting forward everything is just entities; Some of these entities are constructed in one way, with one specific pattern of information, e.g. by evolution. Others are constructed in another, e.g. by the assembly lines in a factory. But neither process makes the resulting entity 'special' per se. And whether what one may call its "directed behaviour" is the result of a familiar type of consciousness, or not does not matter. This is the basis for 'Vector Utilitarianism' – which I will further outline in the proposed paper.

KEYWORDS: Ethics, Moral agents & patients, Meta-Ethics, AI Ethics, Consciousness.

REFERENCES

(examples)

Carroll, S. (2016). The Big Picture: On the Origins of Life, Meaning, and the Universe Itself. Penguin.

Dawkins, R. (2006). The Selfish Gene: 30th Anniversary Edition - with a New Introduction by the Author. Oxford University Press, Oxford.

Dennet, D. (November 1999). The zombic hunch: Extinction of an intuition? Royal Institute of Philosophy Millennial Lecture.

De Waal, F. (February 2014). Evolved Morality: The Biology and Philosophy of Human Conscience. Brill Academic Pub, Leiden, Boston.

Greene, J. (2013). Moral Tribes: Emotion, Reason, and the Gap between Us and Them. The Penguin Press, New York.

Floridi, L. (2014). The Fourth Revolution: How the Infosphere Is Reshaping Human Reality. Oxford University Press, Oxford.

Singer, P. (2011). Practical Ethics. Third Edition. Cambridge University Press, Cambridge.

Yudkowsky, E. (2009). Three Worlds Collide. Rerieved from http://lesswrong.com/lw/y4/three worlds collide 08/.

Information technologies are transforming our lives, becoming a key resource that makes our day to day activities inconceivable without their use. The degree of dependence on ICT is growing every day, making it necessary to reshape the ethical role of technology in order to balance society's 'techno-welfare' with the ethical use of technologies. Ethical paradigms should be adapted to societal needs, shifting from traditional non-technological ethical principles to ethical paradigms aligned with current challenges in the smart society.

UNIVERSIDAD DE LA RIOJA

UNIVERSITAT ROVIRA i VIRGILI