ICDS 2021

The Fifteenth International Conference on Digital Society

ISBN: 978-1-61208-869-3

July 18 – 22, 2021

Nice, France

ICDS 2021 Editors

Theo Lynn, Dublin City University Business School, Ireland
Lasse Berntzen, University of South-Eastern Norway, Norway
Pierangelo Rosati, Dublin City University, Ireland
Oana Dini, IARIA, EU/USA

ICDS 2021

Forward

The Fifteenth International Conference on Digital Society (ICDS 2021) was held in Nice, France, July 18 - 22, 2021.

Nowadays, most of the economic activities and business models are driven by the unprecedented evolution of theories and technologies. The impregnation of these achievements into our society is present everywhere, and it is only question of user education and business models optimization towards a digital society.

Progress in cognitive science, knowledge acquisition, representation, and processing helped to deal with imprecise, uncertain or incomplete information. Management of geographical and temporal information becomes a challenge, in terms of volume, speed, semantic, decision, and delivery.

Information technologies allow optimization in searching an interpreting data, yet special constraints imposed by the digital society require on-demand, ethics, and legal aspects, as well as user privacy and safety.

The event was very competitive in its selection process and very well perceived by the international scientific and industrial communities. As such, it is attracting excellent contributions and active participation from all over the world. We were very pleased to receive a large amount of top quality contributions.

The accepted papers covered a large spectrum of topics related to advanced networking, applications, social networking, security and protection, and systems technologies in a digital society. We believe that the ICDS 2021 contributions offered a panel of solutions to key problems in all areas of digital needs of today's society.

We take here the opportunity to warmly thank all the members of the ICDS 2021 technical program committee as well as the numerous reviewers. The creation of such a broad and high quality conference program would not have been possible without their involvement. We also kindly thank all the authors that dedicated much of their time and efforts to contribute to the ICDS 2021. We truly believe that thanks to all these efforts, the final conference program consists of top quality contributions.

This event could also not have been a reality without the support of many individuals, organizations and sponsors. In addition, we also gratefully thank the members of the ICDS 2021

organizing committee for their help in handling the logistics and for their work that is making this professional meeting a success.

We hope the ICDS 2021 was a successful international forum for the exchange of ideas and results between academia and industry and to promote further progress on the topics of digital society.

ICDS 2021 Chairs

ICDS Steering Committee
Lasse Berntzen, University of South-Eastern Norway, Norway
Claus-Peter Rückemann, Westfälische Wilhelms-Universität Münster (WWU) / DIMF / Leibniz Universität Hannover, Germany
Theo Lynn, Irish Institute of Digital Business, Dublin City University, Ireland

ICDS 2021 Advisory Committee
Åsa Smedberg, Stockholm University, Sweden
Jaroslaw Kozlak, AGH University of Science and Technology, Krakow, Poland

ICDS 2021 Publicity Chair
Alvaro Liebana, Universitat Politecnica de Valencia, Spain
Mar Parra, Universitat Politecnica de Valencia, Spain

ICDS 2021

COMMITTEE

ICDS Steering Committee
Lasse Berntzen, University of South-Eastern Norway, Norway
Claus-Peter Rückemann, Westfälische Wilhelms-Universität Münster (WWU) / DIMF / Leibniz Universität Hannover, Germany
Theo Lynn, Irish Institute of Digital Business, Dublin City University, Ireland

ICDS 2021 Advisory Committee

Åsa Smedberg, Stockholm University, Sweden
Jaroslaw Kozlak, AGH University of Science and Technology, Krakow, Poland

ICDS 2021 Publicity Chair

Alvaro Liebana, Universitat Politecnica de Valencia, Spain
Mar Parra, Universitat Politecnica de Valencia, Spain

ICDS 2021 Technical Program Committee

Chiniah Aatish, University of Mauritius, Mauritius
Mohamad Ibrahim Al Ladan, Rafik Hariri University, Lebanon
Laura Alcaide Muñoz, University of Granada, Spain
Ludivine Allienne, Université Picardie Jules Verne - laboratoire CURAPP-ESS, France
Kambiz Badie, ICT Research Institute, Iran
Ilija Basicevic, University of Novi Sad, Serbia
Najib Belkhayat, Cadi Ayyad University of Marrakech, Morocco
Lasse Berntzen, University of South-Eastern Norway, Norway
Aljosa Jerman Blazic, SETCCE Ltd. / IT association at Chamber of commerce, Slovenia
Mahmoud Brahimi, University of Msila, Algeria
Justin F. Brunelle, The MITRE Corporation, USA
Maria Chiara Caschera, CNR-IRPPS, Italy
Sunil Choenni, Dutch Ministry of Justice and Security / Rotterdam University of Applied Sciences, Netherlands
Yul Chu, University of Texas Rio Grande Valley (UTRGV), USA
Soon Ae Chun, City University of New York, USA
Kathy Clawson, University of Sunderland, UK
María E. Cortés-Cediel, Universidad Complutense de Madrid, Spain
Arthur Csetenyi, Budapest Corvinus University, Hungary
Fisnik Dalipi, Linnaeus University, Sweden
Monica De Martino, CNR-IMATI (National research Council, Institute of applied Mathematics and Information technology), Italy
Alexander Dekhtyar, California Polytechnic State University, USA
Joakim Dillner, Karolinska University Laboratory | Karolinska University Hospital - Center for Cervical Cancer Prevention, Sweden

Copyright Information

For your reference, this is the text governing the copyright release for material published by IARIA.

The copyright release is a transfer of publication rights, which allows IARIA and its partners to drive the dissemination of the published material. This allows IARIA to give articles increased visibility via distribution, inclusion in libraries, and arrangements for submission to indexes.

I, the undersigned, declare that the article is original, and that I represent the authors of this article in the copyright release matters. If this work has been done as work-for-hire, I have obtained all necessary clearances to execute a copyright release. I hereby irrevocably transfer exclusive copyright for this material to IARIA. I give IARIA permission or reproduce the work in any media format such as, but not limited to, print, digital, or electronic. I give IARIA permission to distribute the materials without restriction to any institutions or individuals. I give IARIA permission to submit the work for inclusion in article repositories as IARIA sees fit.

I, the undersigned, declare that to the best of my knowledge, the article is does not contain libelous or otherwise unlawful contents or invading the right of privacy or infringing on a proprietary right.

Following the copyright release, any circulated version of the article must bear the copyright notice and any header and footer information that IARIA applies to the published article.

IARIA grants royalty-free permission to the authors to disseminate the work, under the above provisions, for any academic, commercial, or industrial use. IARIA grants royalty-free permission to any individuals or institutions to make the article available electronically, online, or in print.

IARIA acknowledges that rights to any algorithm, process, procedure, apparatus, or articles of manufacture remain with the authors and their employers.

I, the undersigned, understand that IARIA will not be liable, in contract, tort (including, without limitation, negligence), pre-contract or other representations (other than fraudulent misrepresentations) or otherwise in connection with the publication of my work.

Exception to the above is made for work-for-hire performed while employed by the government. In that case, copyright to the material remains with the said government. The rightful owners (authors and government entity) grant unlimited and unrestricted permission to IARIA, IARIA's contractors, and IARIA's partners to further distribute the work.

Table of Contents

A Longitudinal Analysis of the Determinants of Citizen Acceptance of Contact Tracing Mobile Apps

Grace Fox
Irish Institute of Digital Business
Dublin City University
Dublin, Ireland
email: grace.fox@dcu.ie

Theo Lynn
Irish Institute of Digital Business
Dublin City University
Dublin, Ireland
email: theo.lynn@dcu.ie

Pierangelo Rosati
Irish Institute of Digital Business
Dublin City University
Dublin, Ireland
email: pierangelo.rosati@dcu.ie

Lisa van der Werff
Irish Institute of Digital Business
Dublin City University
Dublin, Ireland
email: lisa.vanderwerff@dcu.ie

***Abstract*—A significant number of governments worldwide have introduced mobile contact tracing apps in an attempt to contain the spread of COVID-19. The use of these government surveillance technologies provides clear benefits for citizens, health bodies and governments but also raises privacy concerns, which may ultimately undermine the adoption of such technologies. This research comprises two discrete longitudinal studies; the first study explores the influence of privacy perceptions and positive beliefs on citizen acceptance of mobile tracing apps while the second study focuses on the antecedents and behavioural intentions related to trust and privacy perceptions. Preliminary evidence suggests that perceived benefits, social influence and trust are positively related to adoption and disclosure intentions while privacy concerns have a negative effect. Both studies offer valuable theoretical contributions to the academic literature on privacy and technology adoption, and practical and timely contributions to government bodies and practitioners.**

***Keywords*—Privacy; Government Surveillance; COVID-19; Contact Tracing; Mobile Application.**

I. Introduction

It is well established that from the outset COVID-19 posed significant challenges for healthcare systems due to the uncertainty in transmission, pathogenicity and lack of strain-specific control options. As a result, governments focused on non-pharmaceutical interventions including social distancing and social isolation. For the first time, governments and health agencies were able to leverage the widespread adoption of mobile technologies and contact-tracing apps to support these efforts on a national scale. While contact tracing apps provide clear benefits in containing the spread of COVID-19, they constantly track users' location and social interactions and therefore can be seen as a potential form of government surveillance. As such, they generate understandable privacy concerns, which may ultimately undermine adoption, and consequently reduce the potential benefits of these apps in the fight against the virus [1]. In this research, we combine Privacy Calculus Theory (PCT) with Social Exchange Theory (SEC), and separately with Procedural Fairness Theory (PFT), to explore the influence of privacy perceptions, positive beliefs and trust on citizen acceptance of mobile tracing apps. These theoretical frameworks are briefly presented in Section II. Section III and Section IV present the sample used in the two studies and some preliminary results respectively. Finally, Section V presents some concluding remarks.

II. Theoretical Background

Technology adoption is one of the most developed streams of research within the information system literature comprising well known and widely-used theories [2]. Diffusion of Innovation theory, the Technology Acceptance Model, the Theory of Planned Behaviour, the Unified Theory of Technology Acceptance and Use have been used to investigate the determinants of individuals' acceptance of a wide range IT innovations and surveillance technologies [3]. Due to the relatively novel context and early stage of diffusion, factors in extant adoption theories may not be fully applicable to mobile contact tracing apps for pandemics. For this reason, our studies focus on contextually relevant factors, such as social influence and trust that are important factors in health and government surveillance research. More specifically, we combine PCT with SEC (Study 1) and PFT (Study 2).

PCT posits that individuals compare the costs and benefits associated with adopting a new technology or disclosing personal information before they engage in such a behaviour and are likely to do so as long as the benefits outweigh the costs [4]. Consistent with the core values of PCT, SET seeks

 ISBN: 978-1-61208-869-3

to explain human behaviour with an emphasis on social structures and norms, and posits that individuals expect reciprocal benefits when being required to adhere to social norms [5]. In our first study, we combine PCT and SEC to unravel the combinatory effect of privacy concerns and social factors on intention to adopt and use a mobile contact-tracing app.
Our second study focuses more on the role of trust as an antecedent of intention to adopt and use the app. In order to do so, we combine PCT with PFT, which posits that an individual's perception that a particular activity is conducted fairly is an important driver of risk appraisal [4]. In the context of contact tracing apps and privacy, the idea of fairness mostly refers to the perception that data are collected and used fairly.

III. Data Collection and Sample

The Irish Health Service Executive (HSE) launched a national contact tracing application called COVID Tracker on 6 July 2020. Within 48 hours of its initial launch, over 1 million people (almost 50% of adult smartphone users in Ireland) had downloaded the app, and 300,000 people had checked-in [6]. Data was collected using two surveys, one prior to the introduction of the app (T1) and one after the launch (T2). We used existing scales when developing our instrument with minor wording amendments to adapt items to the context. The T1 survey focused more on perceived privacy concerns, situational variables, and propensity to trust and adopt the app. The T2 survey focused more on future intention to use the app.
Responses were collected using an online panel of Irish residents provided by Qualtrics with age, gender and regional quotas to ensure the sample was representative of the population of Ireland. A total of 1,109 complete responses were collected at T1. All the respondents were then re-contacted at T2. The final sample consisted of 405 complete responses (37% response rate). Table I provides an overview of the characteristics of the final sample.

TABLE I
Sample Overview

Gender	**#**	**Age Range**	**#**
Male	189	18-20	4
Female	225	21-29	29
Rather Not Say	0	30-39	66
		30-49	79
		40-59	86
		60+	89
Employment Status	**#**	**Education Level**	**#**
Employed	186	Secondary School	157
Self-employed	26	Trade	5
Unemployed	36	Diploma	32
Student	11	Bachelor degree	133
Unavailable for work	42	Other Qualification	64
Retired	104	Doctorate degree	14

Responses were analysed using Confirmatory Factor Analysis and Partial Least Squares Structural Equation Modelling (PLS-SEM).

IV. Preliminary results

Preliminary results from Study 1 suggest that social influence, reciprocal benefits and perceived health benefits (i.e., positive beliefs) have a positive effect on individuals' intention to download the application. On the contrary, negative beliefs related to privacy concern such as risk does not have a significant influence on individuals' intention to download the application. Our results also show individuals' future intentions regarding the use of the application were influenced by their prior adoption intentions and reciprocal benefits, while privacy concern only has a negative influence (at the .10 level). Study 2 focused the role of trust as an antecedent of intention to adopt and use the contact tracing app. Preliminary findings suggest that propensity to trust technology, perceived government motives, perceived need for government surveillance through the contact tracing app, and perceived control have a positive effect on perceived trust in the mobile contact tracing app. In contrast, perceived intrusion through the mobile contact tracing app has a negative effect on perceived trust. Results also suggest that perceived trust and privacy have a positive effect on intention to adopt but do not directly influence usage intention, and that perceived trust and intention to use have a positive effect on users' willingness to disclose personal information.

V. Conclusion

By leveraging a longitudinal dataset, our research adopts a broader conceptualisation of user acceptance that includes both intention to adopt and use over time thus overcoming the typical limitations of cross-sectional samples. Furthermore, we extend PCT by combining it with SEC, and separately PFT, to capture the complexity of privacy consideration and the importance of fairness consideration in the context of government surveillance technologies. The findings of our studies provide useful insights for practitioners and policymakers as they clearly highlight the importance of transparency, perceived benefits, and reciprocity in fostering adoption. Policymakers and public health agencies need to carefully consider what communication and control mechanisms can be introduced to build (and repair) trust, remediate potential sources of distrust in the development and design of such apps, communicate the health and public benefits of using the app and how data will and will not be used, and ensure that collected data is deleted when no longer necessary.

References

[1] M. Zastrow, "Coronavirus contact-tracing apps: can they slow the spread of covid-19?" *Nature*, 2020.
[2] A. Jeyaraj, J. W. Rottman, and M. C. Lacity, "A review of the predictors, linkages, and biases in it innovation adoption research," *Journal of information technology*, vol. 21, no. 1, pp. 1–23, 2006.
[3] N. Thompson, T. McGill, A. Bunn, and R. Alexander, "Cultural factors and the role of privacy concerns in acceptance of government surveillance," *Journal of the Association for Information Science and Technology*, vol. 71, no. 9, pp. 1129–1142, 2020.
[4] Y. Li, "Theories in online information privacy research: A critical review and an integrated framework," *Decision support systems*, vol. 54, no. 1, pp. 471–481, 2012.
[5] E. M. Church, R. Thambusamy, and H. Nemati, "Privacy and pleasure: a paradox of the hedonic use of computer-mediated social networks," *Computers in Human Behavior*, vol. 77, pp. 121–131, 2017.
[6] C. Brennan. (2020) Hse's covid-19 tracing app passes 1m downloads. https://www.irishexaminer.com/news/arid-31010089.html (last accessed on 28 June 2021).

 ISBN: 978-1-61208-869-3

A Preliminary Analysis of the Determinants of Acceptance of Contact Tracing Apps in Brazil

Patricia Takako Endo
Caruaru Campus
Universidade de Pernambuco (UPE)
Recife, Brazil
email: patricia.endo@upe.br

Grace Fox
Irish Institute of Digital Business
Dublin City University
Dublin, Ireland
email: grace.fox@dcu.ie

Theo Lynn
Irish Institute of Digital Business
Dublin City University
Dublin, Ireland
email: theo.lynn@dcu.ie

Pierangelo Rosati
Irish Institute of Digital Business
Dublin City University
Dublin, Ireland
email: pierangelo.rosati@dcu.ie

Lisa van der Werff
Irish Institute of Digital Business
Dublin City University
Dublin, Ireland
email: lisa.vanderwerff@dcu.ie

***Abstract*—This research investigates the factors driving acceptance and use of mobile contact tracing apps in Brazil. The study leverages Theory of Planned Behaviour (TPB), Privacy Calculus Theory (PCT) and social contract theory, and a sample of 1,114 Brazilian residents to demonstrate the importance of perceived control and perceived surveillance in the formation of individuals' perceptions of privacy. Preliminary findings suggest that perceived privacy, reciprocal benefits, and social influence have a positive effect on individuals' intentions to download or continue use of contact tracing mobile apps. Similarly, intention to adopt, perceived privacy, and reciprocal benefits have a positive effect on intentions to disclose information.**

***Keywords*—Privacy; Social Contract; Social Influence; COVID-19; Mobile Apps; Contact Tracing.**

I. INTRODUCTION

The COVID-19 pandemic has radically changed all aspects of our lives. Given the initial lack of effective pharmaceutical countermeasures, a number of exceptional public health measures have been introduced by various governments in order to reduce the transmission of the virus. Most governments mainly focused on keeping the public informed, timely contact tracing, and isolation of individuals who have been exposed to the virus. Digital technologies have proven to be particularly useful in this context by enabling fast dissemination of information to citizens and location-based contact tracing. Contact tracing apps in particular have become a common feature of the battle against COVID-19 [1]. Some of these apps have been introduced by national governments while others have been introduced by application developers [1] .

Brazil is one of the countries that have been most affected by COVID-19 [2]. A number of mobile contact tracing apps have been introduced in the country including one funded by a local state government to monitor symptoms and self-isolation (Monitora), and a contact tracing app, which links information from healthcare organizations with geo-location data to infer users' risk of infection (Dycovid). While these apps require users to disclose different information, they both collect sensitive data, such as users' health information and location. As such, the use of these apps raise significant privacy concerns as they can be perceived as a form of government surveillance. Widespread adoption and users' willingness to disclose information are necessary conditions for the success of these apps [3], thus it is important to investigate how individuals' perceptions regarding privacy and benefits associated with these applications influence their acceptance and ultimate use of the apps. In this study, we combine Theory of Planned Behaviour (TPB), Privacy Calculus Theory (PCT), and Social Contract Theory (SCT) to answer the following research question: *what roles do individuals' perceptions of privacy and benefits play in determining their acceptance and use of contact tracing applications?* By doing so, we aim to contribute to the growing literature on privacy and the adoption of new online services, and to the emergent literature on the benefits and challenges of digital technologies during a pandemic. The remainder of this article is organised as follows. Section II introduces the theoretical framework; Section III provides a brief overview of the sample and data collection; Section IV presents a summary of some preliminary findings; and Section V presents some concluding remarks.

II. THEORETICAL BACKGROUND

Academic research on privacy and technology adoption has recently begun to focus on emerging online services and the privacy implications of such technologies. However, most studies follow the Antecedents-Privacy Concerns-Outcomes (APCO) framework presented by Smith et al. [4] and do not examine the relationship between antecedents, privacy and behavioral outcomes. In this study, we aim to the overcome such a limitation by investigating the role of privacy in the context of contact tracing mobile apps across the entire APCO framework. To this aim, we leverage TPB as our underlying framework. It posits that the behavioural intentions of an individual are a function of their attitudes towards the

 ISBN: 978-1-61208-869-3

behaviour, subjective norms, and perceived behavioural control [5]. While TPB has been the dominant theoretical approach for investigating health-related behaviour for decades, it has also attracted a number of criticisms mostly because of its low explanatory power in real-life experiments [6]. For these reasons, we combined TPB with SCT and PCT and as a result take into account a larger number of factors that might affect the behavioural intentions of a given user.

According to SCT, a social contract exists every time an exchange of information is required; this equally applies to the use of a mobile app. Such a contract is directly related to how information should be used. Individuals are more likely to engage in the exchange when they believe that the social contract will be honored [7]. SCT also states that organisations must provide the individual with control over how their information is used [8]. We argue that the same logic applies to contact tracing apps and that, if individuals' perceive that these apps can be used for government surveillance purposes, their perception of privacy will diminish. In turn, a lower perception of privacy would negatively affect their willing to accept and use the contact tracing app. In contrast, PCT posits that individuals compare positive and negative outcomes associated with a behavior prior to deciding whether or not to engage in that behavior [7], and that they will then engage in the behavior if the positive outcomes outweigh the negatives [9]. In this study, we adopt PCT to examine the influence of perceived privacy and several positive beliefs on users' acceptance of mobile contact tracing apps.

III. Data Collection and Sample

Data was collected using an online survey in Portuguese based on validated scales with minor wording amendments to adapt items to the context. Given the existence of several contact tracing apps in Brazil, participants were initially asked if they had downloaded any of the main apps. We only explored benefits and privacy perceptions of the users who had experience with this type of apps. Participants with no prior experience using contact tracing apps were presented with a neutrally-framed description of a hypothetical contact tracing app and asked questions on their intentions to (i) adopt such an app, (ii) disclose accurate data, and (iii) rely on the app for health advice. Responses were collected using stratified sampling with gender, age and regional quotas and an online panel provided by Qualtrics. 1,175 complete responses were collected however 50 responses were filtered out due to low standard deviation in the responses and other 11 responses due to a completion time that was too short. Table I provides a breakdown of the 1,114 responses included in our final sample. Responses were analysed using Confirmatory Factor Analysis and Partial Least Squares Structural Equation Modelling (PLS-SEM).

TABLE I
Sample Overview

Gender	**#**	**Age Range**	**#**
Male	524	18-20	79
Female	589	21-29	252
Rather Not Say	1	30-39	286
Other	0	40-49	214
		50-59	175
		60-69	96
		70+	12
Employment Status	**#**	**Education Level**	**#**
Employed	507	Primary School	9
Self-employed	272	High School	467
Unemployed	149	Technical College	14
Student	86	Bachelor degree	382
Unavailable for work	11	Masters degree	294
Retired	89	Doctorate degree	35
		Other	13

IV. Preliminary results

Preliminary findings suggest that perceived privacy, reciprocal benefits, and social influence have a positive effect on individuals' intentions to download or continue use of contact tracing applications. Similarly, intention to adopt, perceived privacy, and reciprocal benefits have a positive effect on intentions to disclose information. Interestingly, the results also show that perceived health benefits have a positive effect on intention to continue to use the app by current users while it has a negative effect on non-users' intention to adopt.

V. Conclusion

Our research aims to contribute to the academic literature on privacy and technology adoption in at least two ways. First, we combined PCT and SCT to investigate the role that privacy antecedents play in the context of mobile contact tracing app acceptance. Second, we compared the role of perceived privacy on acceptance of both users and non-users. The findings of our study have also practice implications as it demonstrates that the perception of privacy is a major driver of acceptance in the context of contacting tracing apps. It is therefore important for developers and organisations invested in these apps to foster these perceptions prior to the introduction of new applications.

References

[1] R. Raman, K. Achuthan, R. Vinuesa, and P. Nedungadi, "Covidtas covid-19 tracing app scale—an evaluation framework," *Sustainability*, vol. 13, no. 5, p. 2912, 2021.

[2] Worldometer, "Countries where covid-19 has spread," 2021, https://www.worldometers.info/coronavirus/countries-where-coronavirus-has-spread/ (last accessed on 28 June 2021).

[3] T. M. Yasaka, B. M. Lehrich, and R. Sahyouni, "Peer-to-peer contact tracing: development of a privacy-preserving smartphone app," *JMIR mHealth and uHealth*, vol. 8, no. 4, p. e18936, 2020.

[4] H. J. Smith, T. Dinev, and H. Xu, "Information privacy research: an interdisciplinary review," *MIS quarterly*, pp. 989–1015, 2011.

[5] I. Ajzen, "From intentions to actions: A theory of planned behavior," in *Action control*. Springer, 1985, pp. 11–39.

[6] F. F. Sniehotta, J. Presseau, and V. Araújo-Soares, "Time to retire the theory of planned behaviour," *Health psychology review*, vol. 8, no. 1, pp. 1–7, 2014.

[7] M. J. Culnan and P. K. Armstrong, "Information privacy concerns, procedural fairness, and impersonal trust: An empirical investigation," *Organization science*, vol. 10, no. 1, pp. 104–115, 1999.

[8] N. K. Malhotra, S. S. Kim, and J. Agarwal, "Internet users' information privacy concerns (iuipc): The construct, the scale, and a causal model," *Information systems research*, vol. 15, no. 4, pp. 336–355, 2004.

[9] T. Dinev and P. Hart, "An extended privacy calculus model for e-commerce transactions," *Information systems research*, vol. 17, no. 1, pp. 61–80, 2006.

 ISBN: 978-1-61208-869-3

Bad Robot: A Preliminary Exploration of the Prevalence of Automated Software Programmes and Social Bots in the COVID-19 #antivaxx Discourse on Twitter

Antonia Egli
Dublin City University Business School
Dublin City University
Dublin, Ireland
Email: antonia.egli@dcu.ie

Theo Lynn
Dublin City University Business School
Dublin City University
Dublin, Ireland
Email: theo.lynn@dcu.ie

Pierangelo Rosati
Dublin City University Business School
Dublin City University
Dublin, Ireland
Email: pierangelo.rosati@dcu.ie

Gary Sinclair
Dublin City University Business School
Dublin City University
Dublin, Ireland
Email: gary.sinclair@dcu.ie

***Abstract*—Health information is regularly sourced from social media platforms. However, health-related mis- and disinformation, particularly regarding vaccinations, has become increasingly prevalent on social networks since the spread of COVID-19. Automated attempts to manipulate or deceive the public by spreading false information on social media have adverse effects within the online vaccination discourse, for example by potentially converting vaccine hesitant individuals into vaccination deniers. 8,949 English-language tweets featuring the #antivaxx (i.e., anti-vaccination) hashtag generated by 7,721 discrete users in December 2020 were collected, a period when COVID-19 vaccines were first released in the United States. These were examined to determine (a) the prevalence of automated software and social bots in the #antivaxx discourse on Twitter during the focal period, (b) the prevalence of social bot use by active and visible users, and (c) the effectiveness of social network platforms to moderate misinformation. While there is evidence of use of automated software and social bots in the #antivaxx discourse during the period, such software is used by less than 1.5% of users and accounts for between 3.6% and 5.5% of the overall discourse. We also find that active users are more likely to be classified as bots than visible users. Furthermore, Twitter would seem to be effective in identifying and suspending highly active accounts associated with distributing potentially harmful information relating to vaccination.**

***Index Terms*—Twitter; Social Bots; Social Marketing; Social Media; Public Health Communications; Vaccination; #Antivaxx; #Antivaccination; Anti-vaccination.**

I. Introduction

Using social media to gather health information poses quality issues - despite its clear benefits in promoting public health information online, responding to health crises, and tracking disease outbreaks [1]–[4]. These issues arise specifically in regard to the accuracy of health information shared on social media, particularly due to a lack of regulation surrounding non-expert health sources and outdated or incomplete content [1] [5]. Research also suggests evidence of manipulative and deceptive practices [6] [7] which interfere with public health communication by creating a false sense of uniformity and validity, thus endangering public consensus and legitimising questionable or downright false information [7]–[9].

A lack of regulatory structure forces shifts in quality control responsibilities: what used to be monitored by content producers now lies in the hands of consumers' online information hygiene habits [10]. Psychological distance in virtual networks has further lowered norms of appropriate behaviour and increased the likelihood of malpractice [11]. These include tactics such as mimicking grassroots campaigns from higher-authority entities (*astroturfing*) or generating a high volume of content, replete with related hashtags and keywords, to de-emphasise or obscure some other type of activity or content (*smoke screening*) by humans or automated software programmes, i.e., *social bots* [6].

Social bots operate social media accounts and mimic human users with the aim of influencing specific online discussions. While their objective may be benign, there is substantial evidence of malicious use, for example with the aim of spreading rumors, spam, false information, slander, or noise [8]. Research suggests that social bots retweet more frequently than humans, while generating fewer replies, retweets, and mentions from human users [8]. Social bots may be used individually or in *social botnets*, which typically include large groups of bots under a single coordinator, or (*botmaster*), who coordinates their interactions to generate tweets independently of each other or within a retweet chain [13] [14]. In the context

 ISBN: 978-1-61208-869-3

of Twitter, Dickerson et al. [12] categorise such malicious bots into the following:

1) *Spambots*, which spread spam content only.
2) *Paybots*, which make money by tweeting content from accredited sources, but add links leading to sites that pay for traffic.
3) *Influence bots*, which try to sway conversations on Twitter in a specific direction.

In combination with users' increasing tendencies to turn to social media for health information, such malpractice impacts the validity of content shared and the prevalence of rumors or panic spread within a specific health-related online discourse [15]. Extant research suggests that the anti-vaccination movement has employed social bots in the past to influence narratives and decision making processes with respect to vaccination [16]. Repeated exposure to such content has been shown to result in increased hesitancy towards vaccinations, as well as a preference for finding information online rather than from accredited health care organisations [17]–[20].

Contemporaneous with the COVID-19 pandemic, the Director General of the World Health Organization (WHO) has labelled an emerging phenomenon of widespread false information surrounding the 2019 coronavirus as the "infodemic." At the time of writing, the WHO has identified over 30 discrete topics that are the subject of misinformation within the COVID-19 discourse [21]. In this short paper, we examine the prevalence and focus of automated software programmes in the COVID-19 anti-vaccination discourse. We specifically ask to which degree automated software programmes are used and examine those users who are highly active and highly visible within the December 2020 COVID-19 #antivaxx (i.e., anti-vaccination) conversation on Twitter.

The remainder of this short paper is organised as follows. Section II briefly describes the data gathered and methodology applied in three distinct analyses to examine levels of influence, types of software, and prevalence of moderation attempts within the data set. This is followed by an overview of preliminary findings in Section III, categorised by insights into generator software identified, user activity and visibility levels, and an examination of Twitter's platform moderation efforts. Finally, Section IV summarises the value of these findings and outlines elements of future research to be conducted on the subject.

II. Data & Methodology

Twitter is used widely for health surveillance and research [25], and is a popular channel within the anti-vaxx movement [9] [22]. The first release of COVID-19 vaccines took place at the beginning of December 2020. Using Twitter's enterprise application programming interface (API) platform, GNIP, we prepared a data set of 8,849 English language tweets generated in December 2020 featuring the #antivaxx hashtag. *Table 1* presents an overview of the data set.

We performed three discrete analyses in order to assess the prevalence of automated software programmes. First, we identify the most active and visible users in the data set as a proxy for influence [23]. In this case, activity is measured as the sum of tweets, retweets, and replies posted by a user, while visibility is measured as the number of retweets and replies received by a user [24]. Second, to explore the sophistication of technologies used in the discourse, we examine the type of software used to generate tweets. We use the generator metadata available from GNIP to identify the software utility that was used to post the Tweet. This metadata includes the name and a link for the source application that generated the tweet. The general public typically use official Twitter clients or other social networking platforms for cross-posting (e.g., Instagram, Facebook, etc.), while commercial actors are more likely to use social automation and other marketing automation software. The generator metadata can also provide evidence of bot applications. Third, a machine learning algorithm designed specifically for detecting social bots on Twitter, the Indiana University Network Science Institute (IUNI) Botometer, is used to identify the use of social bots by the data set's most active and most visible users. The IUNI Botometer leverages a thousand features from a Twitter account and its activity (such as astroturfing, spamming, potential self-declaration as a bot, or the account's number of 'fake followers') in order to evaluate the similarity of that account to the known features of social bots. In doing so, the IUNI Botometer reports a social bot detection accuracy in excess of 95% [8] [26].

TABLE I
Descriptive Analysis

Metric	Count	Percentage
No. of Distinct Users	7721	-
Total Tweets	8949	100%
No. of Original Tweets	2301	25.71%
No. of Replies	336	3.74%
No. of Retweets	6312	70.53%
No. of Tweets with URLs	1680	18.77%
Avg. Tweets per User	0.3	-

III. Findings

A. Generator Software

Our analysis of the generator metadata confirms our expectations in that the main software used by users in the #antivaxx discourse are official Twitter client software (97%); between 1% and 2% of end users used identifiable bots or automated software to generate tweets. However, these accounts generated between 3.5% and 5% of tweets. Approximately 45 generators (54% of all generators) were self-identified as bots or exhibited bot behaviour.

It is important to note that generator software is not a highly accurate predictor of black hat techniques. For example, some automated social botnets may use official Twitter clients, while others may be human operated, often *en masse* in click farms. The human social botnet analog, referred to as 'meat puppetry,' typically involves paid networks of real Twitter users operating under the direction of a single user who sells the network's reach for a price [13]. Such networks are extremely difficult to

 ISBN: 978-1-61208-869-3

TABLE II
GENERATOR OVERVIEW

Generator	No. of Tweets	No. of Users
Twitter Client	94.68%	97.71%
Bot	3.36%	0.86%
Third Party Twitter Client	1.21%	0.89%
Social Network	0.20%	0.18%
Other	0.54%	0.36%
Total	100.00%	100.00%

identify as the network comprises real users. These networks can be used to generate spam tweets independently of each other or as a single retweeting tree or retweet chain [14].

B. Activity & Visibility

Table 2 offers a deeper breakdown of original tweets (i.e., non-replies and non-retweets), showing the number of those posted by the data set's most active and most visible users. The percentages of original tweets published by the data set's most active and most visible users remain comparably low, indicating a widely spread online conversation with only few highly influential actors functioning as information sources.

TABLE III
ORIGINAL TWEET ANALYSIS

Metric	Count	Percentage
No. of OTs by Most Active Users	387	16.82%
No. of OTs by Most Visible Users	100	4.35%

As discussed, we analysed the most active and visible users using Botometer. Firstly, 22% of the Top 100 active users and 17% of the Top 100 visible users were either suspended or no longer available via Twitter. This typically, although not exclusively, means that the user has deleted their account or their account has contravened Twitter guidelines. Of the remaining Top 100 accounts, 21% of the most active accounts and 6% of the most visible accounts were rated as exhibiting social bot behaviour. This was largely driven by (i) self-declaration that the account was a bot, (ii) high volumes of retweeting (*echo chamber* behaviour), and (iii) a higher number of fake followers than average users. Presenting bot behaviour does not necessarily mean that the account is a bot or has malicious or malign intent. Many of the most visible and active users in our data set, while exhibiting bot-like behaviour e.g., retweeting in high volumes, are either benign, e.g., media outlets or automated news feeds, or represent high-volume Twitter users who may be passionate about countering anti-vaccination messaging through their own form of smoke screening. These include doctors and other health advocates. While outside the scope of this particular paper, the lack of anti-vaccination proponents may suggest a shy anti-vaccination supporter hypothesis and is worthy of future research.

It is important to note that the use of automated software and/or social bots should not be taken to mean that the user or their messages are anti-vaccination. Manual analysis of the most visible and active users suggests that in both cases the majority of accounts identified as behaving like bots were pro-vaccination. Of those rated with a medium to high probability of bot behaviour across both the most active (n=21) and most visible (n=5) users, only one of the accounts was actually anti-vaccination. A greater proportion of the accounts designated unavailable were true anti-vaccination supporters - seven of the 22 most active unavailable accounts were anti-vaccination and only two of the most visible unavailable accounts were anti-vaccination.

TABLE IV
BOT SCORE OVERVIEW

Bot Score	Active Users	Visible Users
Very High	9	1
High	12	5
Medium	2	6
Low	19	21
Very Low	36	51
Suspended/No Longer Accessible	22	17

C. Platform Moderation

Given the higher proportion of true anti-vaccination accounts in the unavailable accounts, and the relatively low number of anti-vaccination promoters in the most visible and most active accounts, one could reasonably posit that Twitter as a platform has been effective at moderating potentially harmful anti-vaccination messaging.

IV. CONCLUSIONS

Vaccine hesitancy is a significant contributor to avoidable deaths and disease burden worldwide. The availability of COVID-19 vaccinations presents an opportunity to control a highly transmissible disease that has resulted in a significant number of deaths as well as an unprecedented health, social, and economic burden on society worldwide. False information on social media can result in individual consumers becoming vaccine hesitant or, *in extremis*, vaccination deniers, and thereby reduce the effectiveness of COVID-19 vaccination programmes.

Few studies consider the computer as a social actor in the context of the anti-vaxx movement and the use of black hat techniques, which may influence the associated discourse. This short summary gives preliminary insights into the prevalence of automated software programmes in the online COVID-19 #antivaxx discourse on Twitter. It forms part of a wider research project (i) analysing over 24.5 million tweets generated on COVID-19 vaccination during December 2020, and (ii) comparing the wider anti-vaccination discourse on Twitter pre-COVID-19 (2018) and during COVID-19 (December 2020 onwards).

Countering the anti-vaccination movement is a significant multi-stakeholder challenge that requires active interventions by public health agencies, policy makers and professionals, pharmaceutical companies, and social media platforms themselves. Greater understanding of the different mechanisms being used by anti-vaccination promoters can help

 ISBN: 978-1-61208-869-3

pro-vaccination stakeholders mitigate the adverse effects of the anti-vaxx movement and restore faith in vaccines and vaccination programmes.

REFERENCES

[1] S. A. Moorhead et al., "A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication," Journal of Medical Internet Research, vol. 15/4, e85, 2013.

[2] C. Chew and G. Eysenbach "Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1120 outbreak," PloS One, vol. 5/11, e14118, 2010.

[3] Y. Zhao and J. Zhang, "Consumer health information seeking in social media: a literature review," Health Information & Libraries Journal, vol. 34/4, pp.268-283, 2017.

[4] P. Rutsaert, Z. Pieniak, A. Regan, A. McConnon, and W. Verbeke, "Consumer interest in receiving information through social media about the risks of pesticide residues," Food Control, vol. 34, pp. 386-392, 2013.

[5] I. C. Sinapuelas and F. N. Ho, "Information exchange in social networks for health care," Journal of Consumer Marketing, vol. 36/5, pp. 692–702, 2019.

[6] M. Kovic, A. Rauchfleisch, M. Sele, and C. Caspar, "Digital astroturfing in politics: Definition, typology, and countermeasures," Studies in Communication Sciences, vol. 18/1, pp.69-85, 2018.

[7] T. Lynn, P. Rosati, G. L. Santos, and P. T. Endo, "Sorting the Healthy Diet Signal from the Social Media Expert Noise: Preliminary Evidence from the Healthy Diet Discourse on Twitter," International Journal of Environmental Research and Public Health, vol. 17/22, 8557, 2020.

[8] E. Ferrara, O. Varol, C. Davis, F. Menczer, and A. Flammini, "The rise of social bots," Commun ACM, vol. 59/7, pp. 96–104, 2016.

[9] D. A. Broniatowski et al., "Weaponized health communication: Twitter bots and Russian trolls amplify the vaccine debate," American Journal of Public Health, vol. 108/10, pp.1378-1384, 2018.

[10] X. Lin, P. R. Spence, and K. A. Lachlan, "Social media and credibility indicators: The effect of influence cues," Computers in human behavior, vol. 63, pp. 264–271, 2016.

[11] C. E. Naquin, T. R. Kurtzberg, and L. Y. Belkin, "The finer points of lying online: Email versus pen and paper," Journal of Applied Psychology, vol. 95/2, pp. 387-394, 2016.

[12] J. P. Dickerson, V. Kagan, and V. S. Subrahmanian, "Using sentiment to detect bots on Twitter: Are humans more opinionated than bots?", Advances in Social Networks Analysis and Mining, IEEE/ACM International Conference, pp. 620-627, 2014.

[13] D. M. Cook, B. Waugh, M. Abdipanah, O. Hashemi, and S. A. Rahman, "Twitter deception and influence: Issues of identity, slacktivism, and puppetry," Journal of Information Warfare, vol. 13, pp. 58–71, 2013.

[14] S. Stieglitz, F. Brachten, B. Ross, and A. K. Jung, "Do social bots dream of electric sheep? A categorisation of 147 social media bot accounts," Proceedings of Australasian Conference on Information Systems, pp. 1-11, 2017.

[15] J. P. Allem and E. Ferrara, "Could social bots pose a threat to public health?", American Journal of Public Health, vol. 108/8, pp. 1005-1006, 2018.

[16] V. S. Subrahmanian et al., "The DARPA Twitter bot challenge," Computer, vol. 49/6, pp. 38-46, 2016.

[17] M. J. Smith and G. S. Marshall, "Navigating parental vaccine hesitancy," Pediatric annals, vol. 39/8, pp. 476-482, 2010.

[18] A. M. Jones et al., "Parents' source of vaccine information and impact on vaccine attitudes, beliefs, and nonmedical exemptions", Advances in Preventive Medicine, pp. 1-8, 2012.

[19] D. Jolley and K. M. Douglas, "The effects of anti-vaccine conspiracy theories on vaccination intentions," PLoS One, vol. 9/2, e89177, 2014.

[20] E. Dubé, M. Vivion, and N. E. MacDonald, "Vaccine hesitancy, vaccine refusal and the anti-vaccine movement: influence, impact and implications," Expert Review of Vaccines. vol. 14/1, pp. 99–117, 2015.

[21] World Health Organisation, "Coronavirus disease (COVID-19) advice for the public: Mythbusters", 2020. Available at: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public/myth-busters [retrieved: June, 2021].

[22] J. Ruiz, J. D. Featherstone, and G. A. Barnett, "Identifying Vaccine Hesitant Communities on Twitter and their Geolocations: A Network Approach." Proceedings of the 54th Hawaii International Conference on System Sciences, p. 3964-3969, 2021.

[23] M. Cha, H. Haddadi, F. Benevenuto, and K. P. Gummadi, "Measuring user influence in twitter: The million follower fallacy." Fourth international AAAI conference on weblogs and social media, pp. 1-8, 2010.

[24] B. K. Chae, "A complexity theory approach to IT-enabled services (IESs) and service innovation: Business analytics as an illustration of IES," Decision Support Systems, vol. 57, pp. 1–10, 2014.

[25] L. Sinnenberg et al., "Twitter as a tool for health research: a systematic review," American Journal of Public Health, vol. 107/1, e1-e8, 2017.

[26] C. A. Davis, O. Varol, O. E. Ferrara, A. Flammini, and F. Menczer, "Botornot: A system to evaluate social bots," Proceedings of the 25th international conference companion on world wide web, pp. 273–274, 2016.

[27] T. Anderson and H. Kanuka, E-research: Methods, strategies, and issues. Boston, MA: Pearson Education, 2003.

[28] D. L. Olson and G. Lauhoff, "Descriptive data mining," In: Descriptive Data Mining (pp. 129-130). Springer, Singapore, 2019.

[29] L. Silver, C. Huang, and K. Taylor, "In emerging economies, smartphone and social media users have broader social networks." Pew Research Center, 2019.

[30] Royal Society of Public Health, "MOVING THE NEEDLE: Promoting vaccination uptake across the life course," 2018. Available at: https://www.rsph.org.uk/static/uploaded/3b82db00-a7ef-494c-85451e78ce18a779.pdf [retrieved: June, 2021].

[31] S. Blume, "Anti-vaccination movements and their interpretations," Social Science & Medicine, vol. 62/3, pp. 628-642, 2006.

 ISBN: 978-1-61208-869-3

RINNO: Transforming Deep Renovation through an Open Renovation Platform

Omar Doukari*, Theo Lynn†, Pierangelo Rosati†, Antonia Egli†, Stelios Krinidis‡, Komninos Angelakoglou‡, Vasileios Sougkakis‡, Dimitrios Tzovaras‡, Mohamad Kassem*, David Greenwood*

*Faculty of Engineering & Environment, Northumbria University, Newcastle, United Kingdom;
Email: {omar.doukari;mohamad.kassem;david.greenwood}@northumbria.ac.uk
†Dublin City University, Dublin, Ireland;
Email: {theo.lynn;pierangelo.rosati;antonia.egli}@dcu.ie
‡Centre for Research and Technology Hellas, Thessaloniki, Greece;
Email: {krinidis;dimitrios.tzovaras}@iti.gr; {angelakoglou;sougakis}@certh.gr

***Abstract*—Building stock accounts for a major portion of worldwide energy consumption and greenhouse gas emissions. Integrating resources, information and automation systems in a proper manner to achieve the required outcomes and meet the relevant regulatory standards for deep renovation and emerging energy efficiency is a significant challenge. RINNO is a Horizon 2020 project that aims to deliver a set of processes that, when working together, provide a system, repository, marketplace, and enabling workflow process for managing deep renovation projects from inception to implementation. This paper presents an overview of the RINNO project and, specifically, RINNO's design for an open renovation platform for managing and delivering deep renovation projects for residential buildings based on seven design principles. The proposed platform will be developed, implemented and evaluated at four pilot sites with varying construction, regulatory, market and climate contexts.**

***Keywords*—Deep renovation; open renovation; energy efficiency; residential buildings; renovation management platform.**

I. Introduction

Despite pledges and public sector efforts to curb global warming, carbon dioxide emissions from the energy sector and industry have increased by 60% since the signing of the UN Framework Convention on Climate Change in 1992 [1]. As such, de-carbonising energy sources and implementing sustainable development and circular economy principles in industry has become a key international effort as evidenced in the UN 2030 Agenda for Sustainable Development, the 2015 Paris Climate Agreement, the EU climate action, and the European Green Deal.

Meeting 2030 targets requires reduction in greenhouse gas emissions in excess of 50% [2]. The EU building stock currently accounts for 40% of the EU's energy consumption and 36% of greenhouse gas emissions [2]. Reducing the energy impact of these building is a significant challenge. Most buildings in Europe have substantially longer lifetimes than those built in less developed countries, many of which were built before thermal standards were introduced or required [3]. By 2050, more than 85% of the EU building stock will still be in use, albeit in need of significant renovation [3].

Deep renovation is defined as a form of renovation that "captures the full economic energy efficiency potential of improvement works, with a main focus on the building shell, of existing buildings that leads to a very high energy performance" [4]. This includes improving heating and cooling demand through a more efficient building envelope, the use of low-energy consuming equipment, increasing renewable energy sources and implementing *smart building* principles [2]. RINNO, a four-year Horizon 2020 project, aims to introduce significant improvements in (i) the efficiency of the deep renovation process, both in terms of cost and time, and (ii) the energy performance and stakeholder satisfaction of deep renovation projects [5]. This will be achieved through a novel open renovation platform for the deep renovation of residential buildings - ranging from planning and design, to retrofitting, operation and monitoring - and implementing emerging technologies including new materials, prefabrication, artificial intelligence, amongst others. In Section II, we introduce the RINNO Open Renovation Platform Renovation Framework. Section III provides a brief overview of how the RINNO platform will be tested in real-life implementations. Section IV presents some concluding remarks.

II. The RINNO Open Renovation Platform & Renovation Framework

RINNO aims to develop a new open renovation software platform. It is designed against seven principles [5]:

1) Full-Lifecycle: implementing the best available technological solutions and construction practices across the renovation project lifecycle;
2) Multi-Stakeholder: taking into account multiple concurrent stakeholder perspectives and requirements;
3) Modularity: capitalising on independently designed systems which, when integrated, function as a whole;
4) Open & interoperable: making data, processes and renovation-specific functionalities available in an ecosystem of building stakeholders;

 ISBN: 978-1-61208-869-3

5) Algorithmic: using machine learning and other, more generic algorithmic systems based on data generated from buildings to enable *augmented building intelligence* in support of planning, design, and building operations;
6) Scalable: using (green) cloud computing to accommodate the massive range of data structure types, algorithms and communication mechanisms inherent in smart buildings; and,
7) Secure: integrating security considerations into all aspects of the open renovation platform from the cloud to the edge.

The RINNO Operational Platform (Figure 1) comprises five layers building on cloud computing Infrastructure-as-a-Service (IaaS) to enable scalability, algorithmic approaches, and data management [5]. The Core RINNO Operational Platform comprises the platform's primary architecture, middleware, orchestration, and analytics systems. The RINNO API (Application Programming Interface) Management Platform supports design, security, publishing, monitoring, analysis, consumption, and monetisation APIs in support of open collaboration standards and data exchange [5]. On top of the underlying infrastructure sits a number of renovation-specific components, which may comprise one or more modules. These include lifecycle-specific modules, e.g., a renovation repository, a planning & design assistant, a retrofitting manager, a building lifecycle renovation manager, and a renovation workflow & transactions manager, finance, and training, as well as more general user administration & support modules. On-site data about the building and its use are gathered through a multi-sensor network. Finally, users and devices can access the platform and data through a variety of interfaces, facilitating multi-stakeholder interactions.

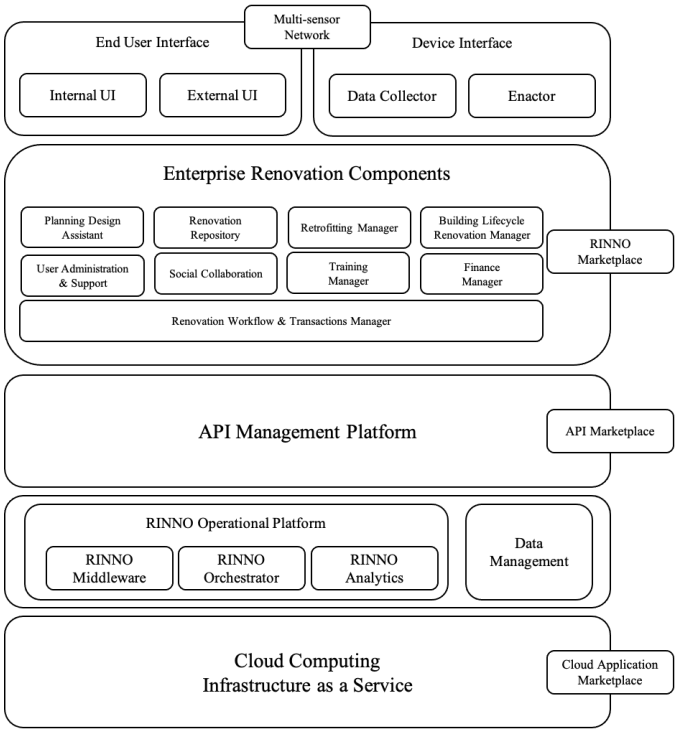

Fig. 1. High level conceptual RINNO Open Renovation Platform with illustrative components [5].

III. RINNO Pilot Sites and Evaluation

The RINNO platform will be used in the renovation of buildings at four pilot sites in Greece, Poland, France and Denmark. These buildings are all multi-unit social housing buildings situated in different climatic regions, with diverse local building codes and regulations. Furthermore, each building features different construction materials and elements and is equipped with different heating, ventilation, and air conditioning (HVAC) systems and other related building systems and features. The project will be evaluated based on a wide range of key performance indicators including reduced energy consumption, the adoption and use of renewable energy sources (RES), thermal performance, renovation time and effort and comparative cost, as well as stakeholder satisfaction measures [5].

IV. Conclusions

This extended abstract provides a short overview of the Horizon 2020 RINNO project. Upon delivery, all processes will work as a system, repository and integrated workflow for the initiation and implementation of deep renovation ventures. This will result in optimised renovation design, reduced renovation time, effort, energy consumption and cost, increased energy efficiency, improved thermal performance, and increased stakeholder satisfaction [5].

Acknowledgment

This work was funded by the European Union's Horizon 2020 Research and Innovation Programme through the RINNO project (https://RINNO-h2020.eu/) under Grant Agreement Number 892071.

References

[1] S. Flemming, "This chart shows the energy milestones we need to reach by 2050 if we are to achieve net-zero carbon emissions," World Economic Forum [online], 2021. Available at: https://www.weforum.org/agenda/2021/06/net-zero-emissions-2050-milestones/ [retrieved: June, 2021]

[2] European Commission, "Impact Assessment-Stepping up Europe's 2030 Climate Ambition: Investing in a Climate-Neutral Future for the Benefit of Our People," [online], 2020. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52020SC0176&from=EN [retrieved: June, 2021].

[3] European Commission, "Renovation Wave" [online], 2021. Available online: https://ec.europa.eu/energy/topics/energy-efficiency/energy-efficient-buildings/renovation-wave_en [retrieved: June, 2021].

[4] S. Shnapp, R. Sitha, and J. Laustsen. "What Is Deep Renovation?" 2013. Available at: https://www.gbpn.org/sites/default/files/08.DR_TechRep.low_.pdf [retrieved: June, 2021].

[5] T. Lynn, P. Rosati, A. Egli, S. Krinidis, K. Angelakoglou, V. Sougkakis, D. Tzovaras, M. Kassem, D. Greenwood, and O. Doukari, "RINNO: Towards an Open Renovation Platform for Integrated Design and Delivery of Deep Renovation Projects," Sustainability, vol. 13/11, p. 6018, 2021.

 ISBN: 978-1-61208-869-3

Strategic Management, Public Value and Co-Creation in the Renewal of Public Agencies across Europe

Keith Shaw
Department of Social Science
Northumbria University
Newcastle upon Tyne, United Kingdom
e-mail: keith.shaw@northumbria.ac.uk

Andreja Pegan
Department of Social Sciences
Northumbria University
Newcastle upon Tyne, United Kingdom
e-mail: andreja.pegan@northumbria.ac.uk

***Abstract*—This is document present the work programme and results from the Horizon 2020 COGOV project. Here we introduce the key research questions of the project on the role of strategic management in co-creation. We then provide an overview the main findings arising from the project and provide a set of recommendations.**

Keywords: Co-creation, Strategic Management, Public Value

I. INTRODUCTION

The COGOV project aims to understand the transformation of European public administrations following post-managerial trends. These trends advocate for a move from bureaucratic organisations and de-emphasise efficiency and effectives to include a broader set of public value such as public legitimacy. The project's focus is on co-creation and the role of strategic management in implementing co-creation. This extended abstract present and defines the main concepts and results.

II. CO-CREATION

We define co-creation as the process through which two or more public and private actors attempt to solve a shared problem, challenge, or task through a constructive exchange of different kinds of knowledge, resources, competences, and ideas that enhance the production of public value in terms of visions, plans, policies, strategies, regulatory frameworks, or services [1]. This can either be through a continuous improvement of outputs or outcomes or through innovative step-changes that transform the understanding of the problem and leads to new ways of solving it. Co-creation is on the public sector agenda for three reasons [2]:

- The public sector is caught in a cross-pressure between growing expectations and scarce public resources;
- The public sector has a limited reach and needs to involve societal actors to solved wicked problems;
- The public sector produce its own distinct public value that many different actors can help produce.

While increasingly popular in contemporary public policy discourses, research shows that co-creation has not yet fully achieved its transformative impact. In this sense, co-creation is still in its infancy and needs to mature to achieve real impact.

III. RESEARCH PROBLEM

The COGOV project responds to the problem of how strategic management can best enable managers and professionals in the public sector to exploit the drivers - and overcome the barriers - to the co-creation of innovative public value outcomes at both organizational and project levels, and which lessons can be shared on undertaking strategically managed co-creation. A strategic management approach means that planning is linked with implementation on an ongoing basis [5]. Because of its long-term and organisation-wide approaches, strategic management support major organizational transitions towards co-creation.

IV. METHODS

COGOV's intent is to produce evidence to answer this policy problem by:

- Administering a survey across 6 national sites;
- Establishing an archive of innovative practices, based on 205 interviews in 15 case studies [6];
- Examining the co-creation of territorial cultural strategies in the UK and France;
- Implementing a design-experimental approach to co-creation in 5 countries to produce a 'CO-CREATOR' board game to aid policy learning;
- Identifying relevant skills for professionals engaged co-creation through focus groups;
- Developing an e-toolkit for managers to aid the development of co-creation.

V. FINDINGS

Our findings indicate that co-creation has several benefits as well as pitfalls, which nonetheless strategic management can help overcome.

A. Co-creation Balance Sheet

Co-creation has many advantages as it:

- Mobilizes experiences and resources of relevant and affected actors that enhance service quality;
- Strengthens social cohesion by bringing together diverse groups of citizens;
- Stimulates innovation and builds common ownership for new solutions;
- Enhances citizens' democratic influence at the output side of the political system.

 ISBN: 978-1-61208-869-3

When public authorities decide to implement co-creation, they should consider the following pitfalls:

- The co-destruction of value due to negligence, incompetence or the abuse of power;
- Arena capture, allowing access to powerful groups at the expense of marginalized groups;
- The stigmatization of citizens;
- Suppression of disagreement in the public pursuit of consensus and quick wins;
- Increasing public costs of facilitation and add-on services.

B. Emerging Lessons on Co-creation

Several lessons on co-creation have emerged:

- Irrespective of administrative traditions, co-creation is widely implemented across countries.
- The use of co-creation to explore new issues and unmet needs is increasingly more likely than to improve already existing services ('co-production').
- Public Value (with an emphasis on stakeholder engagement and improved service performance and satisfaction) as opposed to resources and capacity, is increasingly important in shaping the response of managers.
- Culture shift - managers/professionals need to learn to invite and accept (experiential) knowledge of clients and others - a challenging task, whereas clients and citizens are by definition seen as 'objects of care' rather than equal partners.

C. Emerging Lessons on Strategic Management

Our analysis of strategic management efforts indicates that successful co-creation is based upon the following elements:

- **Values:** A central part of the organizational vision
- **Leadership:** Embraced by leading managers, but infrequently by elected politicians
- **Platforms**: Underpinned by digital platforms but physical platforms are vital for building trust, spurring mutual learning and crafting new solutions
- **Roles and Perceptions**: Supported by a tentative and yet incomplete movement from traditional command and control leadership and management to a more facilitative forms of leadership and management that aims to empower employees and citizens to co-creation innovative public value outcomes.

VI. RECOMMENDATIONS FOR MANAGERS

Based on the findings, COGOV's recommendation are:

- Managers, politicians, and professionals should be aware that a hybrid/mixed approach to strategy may be beneficial when attempting co-creative or co-productive projects or services
- Greater attention needs to be paid to understanding the processes involved with co-creation and co-production outside of simple impact evaluations;
- Managers should set aside time and resources for training professionals in how to co-create;
- Politicians, professionals, and managers should be aware of the necessity of funding when beginning a co-creative project or service
- Ensure consistency and long-term planning. Despite experimentation, co-creation has suffered from the fragmentation caused by one-off and temporary initiatives.

VII. CONCLUSION AND FUTURE WORK

In the past, public and private organizations have focused on using their own money and employees to create results. In the future, platforms and arenas could be created that: enable relevant and affected actors to co-create public value; support these co-creation processes with new digital tools, institutional designs and forms of leadership; while government funding could be made conditional on adopting co-creation in key service areas.

In general, future research on co-creation should emphasize extensive and sustained empirical work including: longitudinally orientated case study work to track how traditional public agencies move to different co-creation approaches and large-scale quantitative research allowing for international benchmarking.

ACKNOWLEDGMENT

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 770591.

REFERENCES

[1] Torfing, J., Sørensen, E., & Røiseland, A. (2019). Transforming the public sector into an arena for co-creation. Administration & Society, 51(5).

[2] Sørensen, E. & Torfing, J. (2015) Enhancing Public Innovation through Collaboration, Leadership and New Public Governance. In A. Nicholas, J. Simon & M. Gabriel (Eds.) New Frontiers in Social Innovation Research. Basingstoke: Palgrave Macmillan, pp. 145-169.

[3] Ferlie, E. & Ongaro, E. (2015). Strategic Management in Public Services Organizations. Abingdon: Routledge

[4] Joyce, P. (2015). Strategic Management in the Public Sector. Abindgon: Routledge.

[5] Bryson, J. M. (2011). Strategic Planning for Public and Nonprofit Organizations. Hoboken: John Wiley & Sons.

[6] Regal, B. & Ferlie, E. (2020). Patterns of strategic renewals. Available at: http://cogov.eu/wp-content/uploads/2020/10/D2-3.pdf (14/7/2020)

 ISBN: 978-1-61208-869-3

Autonomous Weapon Systems, Public Opinion, and the Moral Equality of Combatants

Daniel Lim
daniel.lim672@dukekunshan.edu.cn
Arts & Humanities Division
Duke Kunshan University
Kunshan, China

Philip Santoso
lie.santoso@dukekunshan.edu.cn
Social Sciences Division
Duke Kunshan University
Kunshan, China

***Abstract*—A novel deontological objection to Autonomous Weapon Systems (AWS) based on the moral equality of combatants has recently been developed. While this is an interesting objection, there is reason to believe that it: (1) fails to maintain a moral distinction between AWS and long-distance human-guided weaponry and (2) fails to show that AWS are truly independent in terms of targeting and engaging enemies. Moreover, based on a random sampling of U.S. citizens, public opinion in the U.S. suggests that popular assessment of AWS is more sensitive to relative effectiveness rather than deontological reason. Consequently, this deontological objection to AWS fails and deontological objections to AWS, more generally, can be overridden by consequentialist considerations.**

Keywords-autonomous weapon systems; moral equality of combatants; deontology.

I. INTRODUCTION

In current debates over the development and use of Autonomous Weapon Systems (AWS), deontological arguments have been brought to bear against AWS [1][2][3][4]. A deontological argument against AWS can be characterized as an argument against AWS even if AWS were able to provide the best possible outcomes in terms of legitimate military aims (e.g., minimizing casualties, both military and civilian). In this paper, we aim to show that a specific kind of deontological argument fails and that deontological arguments, more generally, can be overridden by consequentialist considerations.

This paper is organized as follows. In Section II we will briefly summarize several prominent deontological arguments that have been raised against AWS. In Section III we will examine a recent, novel deontological argument based on the moral equality of combatants developed by Skerker et al. [5] that purports to fill gaps in existing arguments. In Section IV we will argue that Skerker, et al.'s argument fails because the concept of 'independent targeting' cannot be fleshed out in a way that (i) draws a clear distinction between AWS and conventional weapons and (ii) can be practically implemented. In Section V we will share results from a study that suggests deontological arguments based on the distinction between independent and dependent targeting can be overridden by considerations of the effectiveness of the weapons involved. We will end with concluding remarks in Section VI.

II. DEONTOLOGICAL ARGUMENTS

Here, we will briefly detail some of the prominent deontological arguments that have been raised against AWS. Some have argued that AWS create responsibility gaps [1]. If AWS were to violate the international laws that govern war, there would be no one to hold morally responsible for such crimes so it would be wrong to deploy AWS. Others have argued that AWS disrespect human combatants because they fail to establish 'interpersonal relationships' with the enemy or fail to acknowledge the humanity of the enemy [2][3]. Consequently, it would be wrong to deploy AWS. Finally, some have argued that combatants must act for the proper reasons in order to justly engage in war [4], but because AWS cannot act for reasons *simpliciter* (let alone the proper reasons) it would be wrong to deploy AWS.

These deontological arguments all have shortcomings. The argument from responsibility gaps would fail if AWS could flawlessly replace human combatants. If AWS could ensure adherence to the international laws that govern war, then the need to locate morally responsible parties for breaches of international laws seems to disappear. The argument from disrespect based on the failure to establish interpersonal relationships fails because it is unable to distinguish AWS from commonly used long-distance, human-guided weaponry (e.g., drones). Consequently, if this argument succeeds against AWS it would also succeed against many weapons that are currently in use. The argument from disrespect based on a failure to acknowledge the humanity of the enemy fails because AWS could, in a sense, be more respectful to the enemy than human combatants. If AWS were to become superior to human combatants in terms of discriminating between legitimate and illegitimate targets and providing assurance that illegitimate targets are not unjustly treated, then AWS would be more respectful. Finally, the argument from acting based on proper reasons fails because human combatants do not always act based on proper reasons. Consequently, this would invalidate much that currently goes on in war. Moreover, if most human combatants were to act from improper reasons, then we would have a legitimate reason to replace human combatants with AWS.

 ISBN: 978-1-61208-869-3

III. THE MORAL EQUALITY OF COMBATANTS

In a recent paper, Skerker et al. have developed a novel deontological argument against AWS that they believe will overcome the shortcomings of the alleged deontological reasons introduced above [5]. They argue that all military personnel at war enter into a martial contract with enemy combatants. Through this contract military personnel:

> "… cede a right not to be directly targeted with lethal violence to… agents able to reason about moral considerations (including whether to exercise their rights at others' expense). Such agents would also need to have an understanding of the authority of moral reasons, reasons grounded in the value of human life." [5]

In other words, military personnel cede a right not to be directly targeted with lethal violence to agents who are *moral equivalents*. Moral equivalents of human combatants have, among other things, "the capacity for moral responsibility" and "moral maturity."

Since AWS do not have the capacity for moral responsibility and are not morally mature, they cannot enter into the martial contract that all military personnel are required to enter. So, human combatants are unable to cede a right not to be harmed to AWS. It would therefore be wrong to deploy AWS against human targets because those targets are unable to cede the right not to be harmed to AWS. Let us call this the Moral Equality of Combatants (MEC) argument against AWS.

One might be tempted to think that this objection to AWS has an obvious flaw: it is too strong since the same objection can easily be raised against conventional weapons. When a human combatant uses a gun to shoot a human enemy, one might argue that it was the bullet that harmed the enemy. However, because the enemy could not cede the right not to be harmed to the bullet (because bullets are not moral equivalents), it would be wrong to use guns against human targets.

This response, however, fails to take seriously the critical distinction between AWS and conventional weapons. Conventional weapons do not *target* human enemies on their own. It is the human combatant who targets and ultimately harms the enemy. While it might be natural to say, for example, that "the bullet killed him," technically speaking, the bullet is nothing more than the means by which the human combatant kills the enemy. So the enemy, by entering the martial contract with opposing human combatants, is not violated because the right not to be harmed was *properly ceded* to the opposing human combatant, not the bullet.

Contrary to conventional weapons, AWS target enemies *on their own*. Because of this, it is not possible for human combatants "to count as targeting combatants targeted by the AWS they deploy." [5]. As a result, the enemy cannot cede the right not to be harmed to the human combatant who deploys the AWS. The enemy harmed by an AWS is disrespected because such an enemy never successfully ceded the right not to be harmed to anyone in this situation.

Not only does the MEC argument enjoy prima facie plausibility, it is strengthened by the fact that it addresses the shortcomings of the deontological arguments introduced above. The MEC argument, unlike the argument from responsibility gaps, is not contingent on AWS that fail to flawlessly replace human combatants. Even if AWS could flawlessly replace human combatants, deploying AWS would still be wrong because humans targeted by AWS would not be able to cede the right not to be harmed.

The MEC argument, unlike the arguments from disrespect, does not depend on the inability of AWS to establish interpersonal relationships nor does it depend on the inability of AWS to respect the humanity of human enemies. Moreover, the MEC argument, unlike the argument from acting based on proper reasons is not contingent on the inability of human combatants to wage war based on proper reasons. The MEC argument requires only that AWS are not moral equivalents for the argument to succeed in showing that deploying AWS is morally wrong.

IV. INDEPENDENT TARGETING

The success of the MEC argument critically depends on the distinction between existing weapons and AWS. This distinction, according to Skerker, Purves, and Jenkins, lies squarely on the claim that AWS target their enemies 'on their own'. We might call this the ability to *independently* target enemies. No other weapons have this ability.

It is unclear, however, whether the concept of independent targeting can be clarified with sufficient detail in a way that vindicates the MEC argument. If the concept cannot be applied to AWS or if the concept can be applied to existing weapons, the MEC argument would be undermined. Let us begin with an extremely strong definition of independent targeting.

INDEPENDENCE$_1$: An entity *S* independently targets an enemy only if *S* does not receive any initial guidance or constraints regarding who or what to target.

According to this definition, it is clear that AWS fail to independently$_1$ target enemies because they are programmed (and possibly trained on datasets) in ways that provide guidance and constraints regarding who or what they target. Since AWS fail to independently$_1$ target it can be argued that it is the human combatant who programmed the AWS that did the targeting. In this sense, AWS are merely tools and moral responsibility ultimately falls on its human programmer. So deploying AWS still makes it possible for AWS victims to cede their right not to be harmed.

INDEPENDENCE$_2$: An entity *S* independently targets an enemy only if *S* makes the *final* targeting decision.

According to this definition, it is clear that AWS independently$_2$ target enemies. Since independence$_2$ does not require the total absence of guidance or constraints, once AWS are deployed, like human combatants, they are continually making decisions and negotiating who or what to target up until harm is inflicted. So, we can say that AWS,

 ISBN: 978-1-61208-869-3

though given initial guidance and constraints, make the final targeting decisions.

It is equally clear that bullets do not independently$_2$ target enemies. All the targeting is done by human combatants wielding guns. Once the gun is fired and the bullet is propelled, the bullet makes no decisions about who or what to target. So there is a principled distinction between human combatants and AWS on the one hand and conventional weapons like guns on the other.

The problem for this definition, however, is that it fails to distinguish AWS from existing weapons like Raytheon's AIM-120 AMRAAM (advanced medium range air-to-air missile). The AIM-120 AMRAAM is considered a 'fire-and-forget' weapon because, though its target is pre-selected, once it reaches a certain proximity to its target, its on board radar-guidance system allows it to autonomously track the target. It continues to negotiate the location of the target up to the point that harm is inflicted. Because it makes the final targeting decision, we should also say that the AIM-120 AMRAAM independently$_2$ targets enemies. The deployment of this weapon along with many other existing 'fire-and-forget' weapons (developed in at least a dozen nations) would fall prey to the MEC argument.

INDEPENDENCE$_3$: An entity *S* independently targets an enemy only if *S* makes the *final* targeting decision and it is false that *S* receives initial guidance or constraints that targets a *single* individual.

Perhaps the difference between existing 'fire-and-forget' weapons is that these weapons, unlike AWS, are always targeted at specific, individually identifiable targets. The AIM-120 AMRAAM, though it makes the final targeting decision, is not independent$_3$ because this weapon receives initial guidance or constraints that targets a single individual. Though this may show that the AIM-120 AMRAAM does not independently$_3$ target enemies, it is unclear that the same can be said for the AGM-114 Hellfire Longbow Variant (or simply Hellfire). This is also a 'fire-and-forget' weapon but, unlike the AIM-120 AMRAAM, the Hellfire can lock onto its target *after* being launched. It seems that it does *not* receive initial guidance or constraints that targets a single individual. It targets *any* member of a set of individuals that match the specifications programmed into its radar guidance system. So it is arguable that the Hellfire, an existing weapon, may be classified as a weapon that independently$_3$ targets enemies and therefore also falls prey to the MEC argument.

Perhaps, there is a deeper issue lurking. Though the distinction between targeting a single individual and targeting a set of individuals may seem theoretically clear, the distinction may prove untenable in practice. To see this we might consider the task of translating *singular* propositions (i.e., propositions that make claims about specific individuals) into *categorical* propositions (i.e., propositions that relate two categories). Categorical propositions in standard form come in four possible formats: (1) All *S* are *P*, (2) No *S* are *P*, (3) Some *S* are *P*, and (4) Some *S* are not *P* – where *S* and *P* are placeholders for categories. Here is a categorical proposition in the first format:

All weapons are dangerous things.

Two categories are being related: weapons (*S*) and dangerous things (*P*). This proposition is asserting that everything in the category of weapons belongs in the category of dangerous things. Without getting into details, the value of translating propositions into categorical propositions in standard form is that arguments based on such propositions can easily be assessed through formal logic.

Consider singular propositions like the following:

Philip is British.

Is it possible to translate this into a categorical proposition? One way is to treat Philip as a criterion for membership into a category.

All people identical to Philip are people who are British.

What this might show is that the distinction between singular propositions and categorical propositions is not rigid. Singular propositions can be treated as categorical propositions.

How does the non-rigid distinction between singular and categorical propositions relate to the distinction between targeting a single individual and targeting a set of individuals? Perhaps we want to use a weapon, like the AIM-120 AMRAAM, to kill a specific individual *X*. It may be tempting to say that this weapon fails to independently$_3$ target its enemy because it received guidance that targets a single individual. As just shown, we can easily construe the targeting of a single individual in terms of a category: all targets identical to *X*. Put in this way, we might argue that it is false that the AIM-120 AMRAAM receives initial guidance or constraints that targets a single individual. Rather it was targeted at a set of individuals.

This would seem to be a pyrrhic victory since the meaning behind the locution 'all targets identical to *X*' is, for all intents and purposes, a way of referring to a single individual. It may, however, only be pyrrhic in a purely theoretical sense. Consider how one might operationalize 'being identical to *X*'. How could this provide concrete guidance in identifying *X*? There seems to be something of a tautology at work here. If I were to tell you to target all people identical to, say, Max Eisenhardt (i.e., Magneto), then, assuming you did not know who Max Eisenhardt is, to tell you to target all and only those individuals identical to Max Eisenhardt would do you no good.

Presumably, one would have to rely on a variety of characteristics that are putatively unique to Max Eisenhardt - characteristics that would be readily available for a human combatant engaged in war to asses. What might such characteristics be? Most likely they would be physical attributes (e.g., height, skin color, facial features). Then what would be meant by 'identical to Max Eisenhardt' would essentially be a list of descriptions. Perhaps something like the following:

 ISBN: 978-1-61208-869-3

All people 190 cm tall with light brown skin color and facial features f_1, f_2, … are targets to be engaged.

Practically speaking, given the number of combatants in a given military, it will be possible that more than one individual will satisfy these descriptions (with some threshold accuracy).

While there might be a neat theoretical distinction between targeting a single individual and a set of individuals, it can reasonably be argued that the practical distinction between targeting a single individual and a set of individuals is rather blurry. As such one might reasonably argue that AWS, when given actual guidance and constraints on who or what to target, do not engage in independent$_3$ targeting.

V. PUBLIC OPINION

The debate over the ethics of AWS will not, however, be decided solely on theoretical grounds. A critical factor in how this debate evolves is public opinion. And this is not merely a descriptive fact about the nature of societies. Apparently, public opinion is encoded into the very laws that govern international warfare. The Martens Clause states:

> "Until a more complete code of the laws of war is issued, the High Contracting Parties think it right to declare that in cases not included in the Regulations adopted by them, populations and belligerents remain under the protection and empire of the principles of international law, as they result from the usages established between civilized nations, from the laws of humanity and the requirements of the public conscience." [6]

Here, we highlight the term 'public conscience' which emphasizes the importance of aligning international law with public opinion.

To probe 'public conscience' we ran a survey of the U.S. public to see, among other things, what their views are regarding AWS [7]. In particular, we wanted to see whether their support for or against AWS would be affected by the distinction between weapon systems that independently target enemies and weapon systems that rely on human controllers to determine who or what to target. Moreover, we wanted to see whether their support for or against AWS would be affected by their 'effectiveness'. That is, whether the weapon systems would be better than human combatants using conventional weapons in terms of reducing the number of overall casualties.

1,600 respondents took our survey. 52% were female and age was evenly distributed. Half the respondents were given a scenario describing weapon systems that rely on human controllers to determine who or what to target (dependent condition) and the other half were given a scenario describing weapon systems that independently target enemies (independent condition). Moreover, with each of these conditions we varied the way the weapon systems were described in terms of relative effectiveness. The weapon systems either performed worse, the same as, or better than human soldiers using conventional weapons. Moreover, we added a control condition that did not comment on their effectiveness. Respondents were then asked whether they "agree that weapons like these should be developed" based on a 5-point Likert Scale ('strongly disagree', 'somewhat disagree', 'neither disagree nor agree', 'somewhat agree', and 'strongly agree').

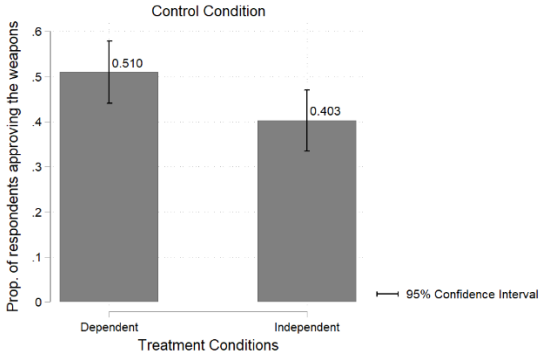

Figure 1. 51% of respondents in the dependent condition and 40% in the independent condition supported the development of the relevant weapons.

We see in the control condition, as shown in Figure 1, weapons that independently target enemies enjoy less approval than their dependent counterparts (where approval stands for a 'somewhat agree' or 'strongly agree' response). Not only is the public less familiar with AWS with the capacity for independent targeting, there is legitimate concern over whether or not such weapons can perform with the same level of effectiveness as human soldiers with conventional weapons.

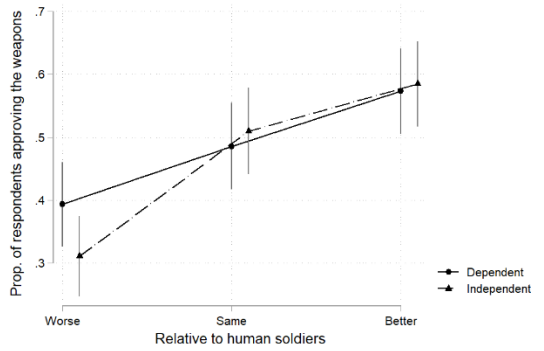

Figure 2. The proportion of respondents steadily increased as the weapons improved in their effectiveness across both dependent and indepent conditions.

That, being said, we also found that the effectiveness of the relevant weapon system affected approval, as shown in Figure 2. It is interesting to note that though weapons which independently target enemies start with a lower proportion of approval (in the worse condition), they eventually end up slightly eclipsing the weapons requiring human controllers for targeting (in the better condition). What this suggests is that whatever deontological worries the public might harbor regarding AWS, these worries are not static. Were AWS able

 ISBN: 978-1-61208-869-3

to provide the best possible outcomes in terms of legitimate military aims (e.g., minimizing casualties, both military and civilian), it seems the public conscience would not be so deeply affected with the deployment of AWS.

VI. Conclusion and Future Work

Though there may be other kinds of deontological arguments raised against AWS in the future, the widely discussed arguments discussed today hinge on the distinction between independent and dependent targeting. This distinction, while theoretically interesting, either cannot be appropriately fleshed out or cannot serve a useful ethical purpose. Moreover, our survey results suggest that whatever deontological reasons people might have for rejecting AWS, the weight of these reasons can be overridden by consequentialist considerations.

To make progress on our results we would like to expand our study to probe military personnel. It may be that experience in military contexts (especially live combat) may affect the way people understand the ethics of AWS. Moreover, it would be interesting to see if the trends discovered with a U.S. population will remain consistent in other cultures. It would be interesting to extend this work to probe the 'public conscience' in other key nations developing AWS like China and Russia.

References

[1] R. Sparrow, "Killer Robots," Journal of Applied Philosophy, vol. 24, pp. 62-77, 2007.

[2] R. Sparrow, "Robots and Respect: Assessing the Case against Autonomous Weapons Systems," Ethics and International Affairs, vol. 30, pp. 93-116, 2016.

[3] T. Nagel, "War and Massacre," Philosophy and Public Affairs, vol. 1, pp. 123-144, 1972.

[4] D. Purves, R. Jenkins, and B. J. Strawser, "Autonomous Machines, Moral Judgment, and Acting for the Right Reasons," Ethical Theory and Moral Practice, vol. 18, pp. 851-872, 2015.

[5] M. Skerker, D. Purves, and R. Jenkins, "Autonomous Weapons Systems and the Moral Equality of Combatants," Ethics and Information Technology, vol. 22, pp. 197-209. 2020.

[6] International Committee of the Red Cross. "The Martens Clause and the Laws of Armed Conflict" [Online]. Available from: https://www.icrc.org/en/doc/resources/documents/article/other/57jnhy.htm 2021.06.26

[7] M. Horowitz, "Public Opinion and the Politics of the Killer Robots Debate," Research and Politics, vol. 3, pp. 1-8. 2016.

 ISBN: 978-1-61208-869-3

Enhancing the Experience of the Digitally Conscious Customer

Viktorija Babica
Deniss Sceulovs
Institute of Business Engineering and Management
Riga Technical University
Riga, Latvia
e-mail: viktorija.babica@rtu.lv

Abstract— **Digital innovations have caused a shift in customer expectations, resulting in a new kind of modern buyer – the digitally conscious customer. The customer experience plays a crucial role in digital transformation and vice versa – a lot of digital transformation initiatives arise from customer's pain points, business innovation needs and growth imperatives on understanding patterns of digital purchasing culture, eliminating information seeking barriers, enhancing virtual social interactions, and overcoming challenges of digital literacy. The focus of the research is on the transformation of purchase process. The research aims to understand the new, digital customer journey and behaviour changes patterns, and map the transformation process of value-based purchase and retail processes. For scholars and practitioners, it is necessary to understand who is driving the change in the modern world – the digital technology phenomena or the customer value-push approach. The research contributes to the practical understanding of how digital innovations have affected the online information seeking process and what barriers in communication have arisen, such as how online social interaction and information filter bubbles affect customers and at the same time how the Zero Moment of Truth (ZMOT) and "dark social" phenomenon affect companies' strategies. Customer journeys are becoming more extensive and versatile, but also challenging. Thus, the result of the research contributes to future mapping the compromise between purchase & retail processes for the Modern Customer Journey in the digital age.**

Keywords - digital journey; customer journey; digital innovation; digitalisation; customer experince.

I. INTRODUCTION

The dynamic world we live in is continuously and rapidly changing. The ubiquity of the information communication technologies has transformed the customer experience. The influence of current digital technologies on the lives of individuals and communities is evident, and one may claim, unprecedented. Mark Weiser [1] 30 years ago in his essay "The Computer for the 21st Century" wrote that "the most profound technologies are those that disappear...", they "will be so ubiquitous that no one will notice their presence" because these technologies will become inseparable and indistinguishable from everyday life. The aim of the present paper and the ongoing research conducted by the authors is to investigate how digital innovations affect user behaviour.

Digital innovations have caused a shift in customer expectations [2], resulting in a new kind of modern buyer culture – the digitally conscious customer. The customer experience plays a crucial role in digital transformation and vice versa – a lot of transformation initiatives arise from customer's pain points (such as purchase convenience, online research, website navigation, checkout process, etc.), business innovation needs and growth imperatives on understanding patterns of digital purchasing culture, [3] eliminating information seeking barriers, [4] enhancing virtual interactions, and overcoming challenges of digital literacy. The digital age, introduced to customers via emerging technologies and digital innovations, is inevitably changing patterns of societal progress, consumption culture, human interaction, values, [5] and even (virtual) identity [6].

Embracing digital transformation requires companies to identify technologies, that are applicable to their strategy and how those may be enacted in the business offerings. As digital technology burrows deeper into the organizational processes and market offerings, it will inevitably affect business strategies, as firms revaluate their perceptions of themselves, as well as their relationships with partners and customers [7]. Firms are feeling the pressure not just to alter their existing business models, but also to operate a portfolio of different business models to cope with increasingly volatile customers preferences, who demand both, flexibility and personalization of products and services [8].

Culture is the informational basis of human society, a vital condition for its existence. As such, culture is inseparable from the information and information technologies. Levin and Mamlok [9] argue that the widespread use of digital technologies constructs and arranges virtual cyberspace that has become an integral part of people's existence. Yet, cyberspace does not replace the familiar reality, but rather complements it and becomes its integral part. Perhaps, the most decisive change relates to the replacement of the traditional concept of the human being as a separate entity by a new ontological self-perception of human beings as an information organism, interconnected with the entire world.

Communication has a tremendous effect on our lives; transformation of the way how people interact with each other, gain information, and learn new things has also been affected by digital innovations. Cultural changes or modification of society values and habits are influenced by

 ISBN: 978-1-61208-869-3

innovation and invention, globalization and cooperation, and the digital innovations are the drivers behind this.

The societal progress would be impossible without digital innovations and interaction between society and technology. Practitioners and scholars state that the customer is the one who is driving the changes which shape the process of digitalization. The challenge is that companies' strategies remain behind the speed of transformation [10]. New technologies have also created novel challenges for companies to understand the changing behaviour of customer.

For practitioners and scholars, it is necessary to investigate emerging digital innovation in the consumption culture and value-based retail process. Understanding how customers react to personalized ads and how those affect buyers' behaviour and priorities may ease the digital innovation adoption process for companies [11].

The rest of the paper is structured as follows. In Section II, we describe the planned methods that are going to be implemented to gather data; in Section III discusses the key aspects of digital transformation and digital innovations; in Section IV the digital customer journey challenges are discussed; in section V the profile of the digitally conscious customer is discussed; section VI presents the further research objectives.

II. METHODS

The main methodology of the further research is analysis of consumer data, literature review, consumer interviews, customer surveys, social media monitoring, survey data from an online panel, secondary data analysis (previous research, statistics, etc.), questionnaires and case studies analysis. The empirical approach is addressed through experiments of purchase and retail processes. The customer perspective we plan to investigate via surveys, questionnaires, interviews, focus group discussion, primary and secondary data review. Customer data we aim to collect with the help of commercial providers and to analyse it themselves. After data review it is planned to test out theories and hypotheses via dialogues and surveys. To address company's perspective, we aim to conduct dialogues, questionnaires, secondary data reviews. Case studies are going to be analysed. Based on the results, there will be developed strategies for practitioners to understand the opportunities to reach the satisfaction equilibrium between customer and retailer. Via qualitative research, there will be investigated the cultural characteristics of the customer journey and retail process. The developed strategies are going to be piloted and then conducted the second round of collecting the customer and retailer vision and feedback.

Data will be collected fully anonymised with use of a mixed method approach for the triangulation of data. We aim at carrying cross-European empirical studies to understand the comparative perspectives of digital customer journeys. We aim to develop the field of purchase and retail processes for modern customer journey conceptually and empirically; establish collaboration with the stakeholders (digital customers and online retailers; practitioners and scholars). The present paper is based on literature review.

III. DIGITAL TRANSFORMATION

Literature does not provide a generally accepted definition of digitalization nor digital transformation; however, some scholars argue that digitalization as a technology-induced transformation process improves organizational flexibility, agility, and responsiveness by simultaneously aligning its operations, strategy, business processes, and organizational and IT structures to technological advancements [12]. Digital transformation cannot be exclusively focused on the technologies, it includes alignment of organizational leadership, relationships with partners and customers, resources, technology, processes, structure, and culture with a strategy that supports them and serves as a basis for achieving a competitive advantage. According to Denner [13], digitalization of products and services is a fast-moving, global megatrend that transforms value networks across all industries. Legner et al. [14] describe digitalization as a pervasive force, that involves transformations of key business operations and affects products and processes, as well as organizational structures and management concepts.

There is no commonly agreed definition and set of differences between digitalization and digital transformation, and scholars continue to use the terms interchangeably, without clear statement what is understood by the certain terms. Henriette et al. [15] define digital transformation as changes in the ways of working, roles, and business offerings caused by adoption of digital technologies in an organization, or in the operation environment of the organization, including process level, organization level, business domain level, society level. It affects organizational strategies and challenges existing business models to be reconsidered and disrupted [16]. Existing business models are being optimized, transformed by reconfiguration and extension, but some devoured by more adjusting competitor.

Digital transformation provides companies with an opportunity to improve their performance, increase reach and ensure better results by using digital innovations [17]. Those bring new actors, structures, practices, values, and beliefs that complement or disrupt existing practices within organizations and industries by affecting all key processes in the organization [18]. New business model typology based on the smartization of products and services has been introduced. In the world of Industry 4.0, sustainability of business models is being questioned. Digitalization has its impact on collection of operational data through sensors with low human intervention, enhances efficiency of data sharing among digital units through connected devices [19].

Innovation concept has been popularized and studied by Schumpeter [20]. He described innovations as creative destruction or a process of industrial mutation [21] that continually changes the existing, destroys old and creates new. Innovation, is traditionally perceived as a creative disruption thus, presenting a threat to traditional business models, current organizational structures, and well-established business operations [22]. Innovation is the core of transformation, and the concept of digital transformation is closely related to the definition of innovation. Innovation

 ISBN: 978-1-61208-869-3

is understood as a process of a new or substantially improved product, idea, method, or business process being implemented or commercialized in the market or organization, creating a new value for the consumer, competitive advantage new industries and job positions and forms along with catalysing the economic growth [23]. Digital innovation is understood as implementation of the new digital technologies either to the existing practices with an aim of their optimization or introducing new process. According to [6] the digital concerns arisen from digital transformation are lack of clear vision, customer expectation, outcome ambiguity, familiarity over innovation, socio-technical misalignment, and cultural inertia.

The traditional ways of doing business have changed dramatically with the emergence of new information technologies. The Internet of things, machine learning, artificial intelligence, cyber-physical systems, cloud computing, wireless networks, robotics, augmented reality, big data analytics, and simulations are some of these key digital technologies. Nowadays, all physical devices can connect to the Internet. Cyber-physical systems incorporate the functions of computing, communications, precision control, coordination, and autonomy. Cloud technologies can be used to increase data sharing across company boundaries, improve agility and flexibility of system performance, and reduce costs by bringing systems online. Simulation optimization-based tools are used for complex systems and automation technologies. As a result, systems are becoming smarter by using these technologies. These means provide opportunities for enhancing customer journey and improving customer experience. Moreover, big data and analytics are used for scaling and evolving information technology. The data with larger volumes and speeds can be analysed more precisely and faster decisions can be made with these technologies. The ability to adopt digital innovation is becoming a vital mechanism for organizations [24]. The capabilities of digital innovations bring great potential for companies to enhance customer experience, improve communication with purchasers and better understand clients' needs.

In the frames of the present paper and further research, the focus is on the digitalization of consumption and its social, cultural, ethical, political, and economical implications. It, thus, answers the call for more research on how customers react to personalized ads and how it affects buyer behaviour and priorities, consumer activities such as the purchase, comparison, and examination of goods [25]. The rate at which digital technology can spawn new the "smart" products and services is matched only by its ability to extend the reach and range of social interactions via ubiquitous infra-structure and malleable platforms [26].

TABLE I. DIGITAL TECHNOLOGIES AND CUSTOMER JOURNEY

The technologies used within customer journey		
Technology	***Description***	***Impact***
Data mining (DM)	Companies use DM capabilities for building the customer-centric approach that focuses on client's pain points in digital customer journey (classification, regression, clustering and association) [33].	Companies can improve digital customer engagement and the related key performance metrics.
Eye Tracking	Contemporary eye tracking technology offers an objective way to document digital users' activities on retailers' websites and mobile app [27].	Companies can reveal consumers' visual attention and see how they evaluate the product or service.
Google Analytics	Web analytics allow companies to track the behaviour of customers visiting their websites, measure web traffic and analyse commercial activity [28].	Application of Google Analytics and AI benefits to optimization of the customer experience and enhance the customer journey.
Internet of Things (IoT)	Internet-based structure for remote locating, sensing, and/or operating the components with real-time data flows between them. IoT contributes to information collection, automating transaction, purchase maintaining and servicing [29]	Companies obtain massive data on consumer behaviour. Arising privacy issues include identification of features unknown to the user, localization and tracking, profiling, making private information public and linking separate information that the user does not like to be linked.
Augmented Reality (AR)	AR creates an add-on and interactive experience of a real-world environment through computer-generated displays, thereby creating more interactive, vivid, and richer experiences for consumers [30].	Companies are able to enhance customer engagement and customer journey by implementing AR, VR and MR. AR is being commercialized actively, meanwhile VR and MR are held behind due to the lack of suitable devices. These technologies facilitate imagination, complement to the physical world, appropriate visualization [32].
Virtual Reality (VR)	Entire environment simulation.	
Mixed Reality (MR)	MR is used for combination of the real and virtual worlds in order to produce new visual environments where physical and digital elements co-exist and interact in real time.	
Artificial Intelligence (AI)	Computer programs that understand user queries and complete a limited set of tasks asked by the users [31].	Companies automate process of advising, and customizing, giving feedback, and recommending additional consumption. Users obtain enhanced decision making.

Table 1 presents non-exhaustive list of the emerging digital technologies used by retailers to better understand and enhance customer journey as well as their engagement.

By integrating Cloud services with its AI tools, companies can realize dynamic landing pages with personalised content based on users' past behaviour or project their behaviour. Additionally, companies, that invest in digital technologies, can achieve in-depth insights into the

 ISBN: 978-1-61208-869-3

effectiveness of its social media marketing efforts through certain cloud-enabled functions. These insights allow to create more relevant and personalised offers for users. In the age of social media, businesses have lost quite a lot of control over how their brand is perceived. However, on the other hand, they have gained a lot of third-party voices that can help them promote their brand. Social media is a much cheaper and more accessible form of advertising for businesses than traditional media, representing a huge opportunity. What businesses need to consider in their engagement strategy in social media is how they can make their messaging contextual and consistent to their target audience. This is especially if their main buying segment is millennials, who tend to be extremely active and vocal on social media and look for brands that align with their values and beliefs.

IV. DIGITAL CUSTOMER JOURNEY

Digital customer journey and digital customer behaviour change patterns are not yet clearly defined. The research, on which this paper is based, aims to offer a better understanding of these concepts and an integrate technological (embodiment), psychological (presence), and behavioural (interactivity) perspectives to propose a new taxonomy of purchase and retail process in the digital age, namely the complementary models, presenting the value-based compromise between purchase & retail processes for modern customer journey in digital age. The models and the equilibrium will assist in enhancing digital customer journey and provide better opportunities for retailers for economic growth. The research aims to provide a better understanding of the cultural differences in customer behaviour change patterns and digitalization of customer journey; analysis of the digital environment complementary elements' effect on the purchaser and retailer; map of the social and cultural aspects interactions with digital innovations.

There is no one unique, or even one-industry comprehensive strategy approach for the customer journey digitalization. Digital technologies have affected consumption practices – changes in customer behaviour, values and loyalty, increased influence of judgements from social media, [33] personalized ads, filter bubbles [34] [35], information asymmetry and social virtual inclusion affect customer's digital journey. Due to the information filter bubbles, barriers in communication and the non-transparent nature of algorithmic filtering digital users are not aware of being in an information asymmetry. The patterns of purchase process transformation arise from the customers' changing priorities, barriers in communication and social interaction, information seeking challenges and habits. Digital literacy of customers also is an important aspect in the transformation process, it is affected by the filter bubbles, dark social (non-traceable source of feedback or way information sharing) and "zero moment of truth" (ZMOT) [36] phenomena. ZMOT is related to the digital value: the degree to which one actively searches for and trusts other users' feedbacks on their shared experiences of a product. Aside with the purchase process, changes arise with the effect on digital innovation on social anxieties, feelings of (in)security, customer loyalty and emergence of such phenomena as "virtual identity" [37].

In the transformation process the focus of companies' (retailer) perspective lies on the value-based retail chain. Is not clear, how do the global interaction, the necessity of implementing sustainable approach and meeting the customers' needs affect the development of the value-based retail chain. Practitioners are aware, that nowadays, it is the customer, who has the power of setting the values, thus the customer value-based proposition is crucial for companies' livability. Digital innovation has provided new possibilities for retailers, but simultaneously have arisen challenges that affect companies' data analysis and strategy [38].

With the present paper and further research, we aim to understand the transformation process of the purchase process, and what are the drivers behind (virtual identity, customer changing values, customer loyalty factors, etc.), explore what are the modern barriers in the information seeking and communication processes (information asymmetry, social online interactions, dark social, filter bubbles, ZMOT, etc.), understand, what are the challenges of digital literacy. We aim to map the change patterns brought by digital innovation, that have affected the consumption culture, user's identity formation, social interactions in the online world and understand the possible future scenarios.

Despite the popularity of digital innovations among customers and retailers, knowledge of their impact on the customer journey remains limited. The main priority for retailers is to better understand the customers behaviour and their values and then to enhance the journey of digital purchase [39]. No prior research has analysed in the depth digital innovation affection on customer values, identity, behaviour, and journey. Processes, data, agility, prioritization, technology, integration, information, business and IT alignment, digitization etc. all matter but are conditions for better customer experiences.

V. DIGITALLY CONSCIOUS CUSTOMER

Smartphones, mobile apps, machine learning, and artificial intelligence provided customers the opportunity to get what they want and when they want it. This has contributed to a shift in customer expectations and created a new kind of consumer – one that has changed the way business is done in specific industries. The concept of digital customer engagement refers to encouraging users to obtain digital channels across various industries remotely [27]. According to [27], there's lack of studies concerning the factors affected by digital customer engagement. Thus, it is crucial to understand digital customer's profile.

Disruptive technologies that reduce friction for the customer journey tend to be the ones that enhance the customers' experience. There is a strong correlation between the effort score and the customer satisfaction levels — the lower the effort required by the customer to conduct business with a company, the higher their satisfaction level. The customer experience plays a crucial role in the digital transformation, many digital transformation initiatives arise from customers pain points. These are caused by the increasing importance of an end-to-end customer experience

 ISBN: 978-1-61208-869-3

improvement approach, which in turn is, among others, caused by changing customer expectations and customer experiences.

According to [40], the digital customer experience and electronic word of mouth (social media) also played positive role on promoting brand image. Digital customer experience through e-commerce and electronic word of mouth played positive role to boost brand image, brad image had positive role to boost supply chain image and finally promoted the performance of the sustainable supply chain.

VI. Conclusion and Future Work

In the digital consumption process on one hand is the customer, on the other is the retailer. In the future we aim to continue the research by approaching this phenomenon from the perspective of business interactions and social consequences to map the state of current situation of barriers in information seeking, lacking digital capabilities, changing patterns of customer loyalty, patterns of customer behaviour changes. Further research requires to relate these occurrences from historical and geographical perspectives' differences and provide scenarios on future changes. It is necessary to test empirically the findings of the present research. The added value from interdisciplinarity is the comprehensive understanding on the research questions, it provides understanding for stakeholders on how deep the business, society and technology are connected.

Besides adeptly performance of high-quality academic research, meticulous examination of practitioners' case studies, and big data analysis of digital customers and online users, it is a crucial need to have a real dialogue with the society, respectively with representatives of different age, gender, and cultures. The public engagement is needed to address the research questions of digital customer priorities, barriers in information seeking, social interaction, social anxieties, feelings of insecurity and customer values. To better understand the current state of digital customers and online users' feeling of inclusion and level of social media judgement affection is not enough to analyse raw data and or perform single questionnaires. The data obtained from digital innovation agents about digital consumption is needed, but it is limited by the information asymmetry impact, filter bubbles and other restrictions and inequalities put by digital technologies. For making research as close to the real situation are needed participatory dialogue events such as focus groups and workshops. We plan to engage society in panel debates, focus groups dialogues, workshops, media engagement and webinars first to inform society (to assist them in understanding the problem the project addresses), to hear out their feedback on the project analysis and research results, involve society to ensure that the research is on the right path in understanding and considering societal challenges in the usage of digital technology in purchasing process. The societal engagement aims to obtain the clear perspective of the research issues and enhance the result quality.

Enhancing customer digital journey requires creative approach, in the future research we aim to investigate if the human personnel can be replaced by the emerging innovations. It is assumed that historical and geographical comparison will help understand the process and address challenges in developing Modern Customer Journey. The future research aims to provide better understatement of digital literacy in consumption, map of digital innovation effect on the digital user and clear definition of cultural differences. For stakeholder communities the benefit of the present research is the awareness of digital innovation potential (positive and negative), ways of overcoming the digital challenges, enhanced customer journey and better value proposition from retailers.

References

[1] M. Weiser, "The Computer for the 21 St Century", Scientific American, vol. 265, Sept. 1991, pp. 94-105.

[2] M. Weber and C. Chatzopoulos, "Digital customer experience: the risk of ignoring the non-digital experience", International Journal of Industrial Engineering and Management, vol. 1, pp. 201-210, 2019, doi:10.24867/IJIEM-2019-3-240.

[3] G. Dovzhik, V. Dovzhik and O. Kurasova, "Empathy and Identification as an Online Technology of Blogger's Communication in Digital Marketing", Studies in Systems, Decision and Control, vol. 314, pp. 1291 – 1301, 2021,doi: 10.1007/978-3-030-56433-9_135.

[4] S. Sarkar, Y. Wang and C. Shah, "Investigating relations of information seeking outcomes to the selection and use of information sources", Proceedings of the Association for Information Science and Technology, vol. 54, pp. 347-356, 2017, doi:10.1002/pra2.2017.14505401038.

[5] C. Flavián, S. Ibáñez-Sánchez and C. Orús, "The impact of virtual, augmented and mixed reality technologies on the customer experience", Journal of Business Research, pp. 547-560, 2019, doi:10.1016/j.jbusres.2018.10.050.

[6] C. Fuentes and E. Brenes Peña, "Internet Conversation: the New Challenges of Digital Dialogue", Dialog Systems: A Perspective from Language, Logic and Computation, pp. 1-25, 2020, doi:10.1007/978-3-030-61438-6_12.

[7] T. Saarikko, U. Westergren and T. Blomquist, "Digital transformation: Five recommendations for the digitally conscious firm", Business Horizons, pp. 825-839, 2020, doi:10.1016/j.bushor.2020.07.005.

[8] F. Li, "The digital transformation of business models inthe creative industries: A holistic framework and emergingtrends", Technovation, vol. 92-93, pp. 102012, 2018, doi: 10.1016/j.technovation.2017.12.004.

[9] I. Levin and D. Mamlok, "Culture and Society in the Digital Age", Information, 12:68, 2021, doi:10.3390/info12020068.

[10] L. Ethan, "The Transformation Of Consumer Behaviors In The Digital Era", Forbes Business Council. [Online]. Available from: https://www.forbes.com/sites/forbesbusinesscouncil/2020/03/13/the-transformation-of-consumer-behaviors-in-the-digital-era/?sh=3753a24a7887/ 2021.05.01

[11] J. F. Klein, Y. Zhang, T. Falk, J. Aspara and X. Luo, "Customer journey analyses in digital media: exploring the impact of cross-media exposure on customers' purchase decisions", Journal of Service Management, pp. 489-508, 2020, doi:10.1108/josm-11-2018-0360.

[12] F. Holotiuk and D. Beimborn, "Critical Success Factors of Digital Business Strategy", International Conference on Wirtschaftsinformatik St. Gallen, 2017, pp. 991-1005. [Online]. Available from: https://aisel.aisnet.org/cgi/viewcontent.cgi?article=1099&context=wi2017 2021.05.03

[13] M-S. Denner, L. Püschel and M. Röglinger, "How to Exploit the Digitalization Potential of Business Processes", Business

& Information Systems Engineering, pp. 331–349, 2017, doi:10.1007/s12599-017-0509-x.

[14] C. Legner et al., “Digitalization: Opportunity and Challenge for the Business and Information Systems Engineering Community” Business & Information Systems Engineering, 59(4), pp. 301–308, 2017, doi:10.1007/s12599-017-0484-2.

[15] E. Henriette, M. Feki and I. Boughzala, “The Shape of Digital Transformation: A Systematic Literature Review”, Mediterranean Conference on Information Systems (MCIS), 2015. [Online]. Available from: https://aisel.aisnet.org/mcis2015/10/ 2021.05.03

[16] C. Linz, G. Müller-Stewens and A. Zimmermann, “Radical Business Model Transformation: Gaining the Competitive Edge in a Disruptive World”, Kogan Page, London, Philadelphia, PA and New Delhi, 2017, pp. 1 - 25, ISBN: 978-0749480455.

[17] J. Brocke et al., “A journey of digital innovation and transformation: the case of Hilti”, G. Oswald & M. Kleinemeier (Eds.), Shaping the Digital Enterprise, Cham: Springer International Publishing, pp. 237–251, 2017, doi: 10.1007/978-3-319-40967-2_12

[18] B. Hinings, T. Gegenhuber and R. Greenwood, “Digital innovation and transformation: An institutional perspective”, Information and Organization, vol. 28(1), pp. 52–61, 2018. doi:10.1016/j.infoandorg.2018.02.004

[19] V. Parida, D. Sjödin and W. Reim, “Reviewing Literature on Digitalization, Business Model Innovation, and Sustainable Industry: Past Achievements and Future Promises”, Sustainability, vol. 11(2), 391, 2019, doi:10.3390/su11020391

[20] S. Breschi, F. Malerba and L. Orsenigo, “Technological regimes and schumpeterian patterns of innovation”, Economic Journal, 110 (463), pp. 388-410, 2000, doi: 10.1111/1468-0297.00530

[21] J.A. Schumpeter, “The Theory of Economic Development: An Inquiry into Profits, Capital, Credit, Interest and the Business Cycle”, Harvard Economic Studies, Vol. 46, Harvard College, Cambridge, MA, 1934.

[22] A. Bharadwaj, O.A. El Sawy, P.A. Pavlou and N. Venkatraman, “Digital business strategy: toward a next generation of insights”, MIS Quarterly, vol. 37(2), pp. 471–482, 2013, doi: 10.25300/MISQ/2013/37:2.3.

[23] V. Babica and D. Sceulovs, “Business Opportunities Unclaimed: Digital Disruption of Business Models”, Proceedings of The 11th International Multi-Conference on Complexity, Informatics and Cybernetics: IMCIC 2020, pp. 88-92, ISBN 978-1-950492-32-9

[24] C-S. Abide and T. Zuhal, “Digital innovations-driven business model regeneration: A process model”, Technology in Society, 2021, pp. 101461, doi:10.1016/j.techsoc.2020.

[25] C. Franck, L. Christian, P. Magdalena and N. McIntyre, “Digitalizing consumer society: equipment and devices of digital consumption”, Journalof Cultural Economy, vol. 13, 2020, pp. 1-11, doi:10.1080/17530350.2019.1702576

[26] S. Nambisan, M. Wright and M. Feldman, “The digital transformation of innovation and entrepreneurship: Progress, challenges, and key themes”, Research Policy, vol. 48(8), 2019, pp. 103773, doi: 10.1016/j.respol.2019.03.018.

[27] M. Mona, N. Agami, G. Elkhayat, and M. Kholief, “A Literature Review of Data Mining Techniques for Enhancing Digital Customer Engagement”, International Journal of Enterprise Information Systems, 16(4), 2020, pp, 80–100, doi:10.4018/IJEIS.2020100105

[28] S. Białowąs, and A, Szyszka, “Eye-tracking in Marketing Research”, Managing Economic Innovations – Methods and Instruments, pp. 91-104, doi:10.12657/9788379862771-6.

[29] Z. Tupikovskaja-Omovie and D. Tyler, “Eye tracking technology to audit google analytics: Analysing digital consumer shopping journey in fashion m-retail”, International Journal of Information Management, vol. 59, 2021, pp. 0268-4012, doi.org/10.1016/j.ijinfomgt.2020.102294.

[30] S. Bayer, H. Gimpel, and D. Rau, “IoT-commerce - opportunities for customers through an affordance lens”, Electron Markets 31, pp. 27–50, 2021, doi.org/10.1007/s12525-020-00405-8

[31] J. Whang, J. Song, B. Choi, and J. Lee, "The effect of Augmented Reality on purchase intention of beauty products: The roles of consumers’ control," Journal of Business Research, vol. 133(C), 2021, pp. 275-284.

[32] L. Mihardjo, W. Wasono, Sasmoko, F. Alamsjah and E. Elidjen, “The influence of digital customer experience and electronic word of mouth on brand image and supply chain sustainable performance”, Uncertain Supply Chain Management, 2019, pp. 691–702, doi: 10.5267/j.uscm.2019.4.001.

[33] C. Kuehnl, D. Jozic, C. Homburg, “Effective customer journey design: consumers’ conception, measurement, and consequences”, Journal of the Academy of Marketing Science, 2019, pp. 551–568, doi:10.1007/s11747-018-00625-7.

[34] F. Seth, G. Sharad and M. Justin, “Filter Bubbles, Echo Chambers, and Online News Consumption”, Public Opinion Quarterly, Volume 80, Issue S1, 2016, pp. 298–320, doi:10.1093/poq/nfw006.

[35] F. Borgesius, D. Trilling, J. Moeller, B. Bodo, Balazs C. de Vreese, N. Helberger, “Should we worry about filter bubbles?” Internet Policy Review, Vol.5, 2016, pp. 1-16, doi: 10.14763/2016.1.401.

[36] S. Han and K. Kihyung, "Role of consumption values in the luxury brand experience: Moderating effects of category and the generation gap," Journal of Retailing and Consumer Services, vol. 57(C), 102249, 2020, doi.org/10.1016/j.jretconser.2020.102249

[37] P. Nagy and K. Bernadet, “The Digital Transformation of Human Identity: Towards a Conceptual Model of Virtual Identity in Virtual Worlds.” Convergence, vol. 20, no. 3, Aug. 2014, pp. 276–292, doi:10.1177/1354856514531532.

[38] R. Mancha and G. Shankaranarayanan, “Making a digital innovator: antecedents of innovativeness with digital technologies”, Information Technology & People, 2020, pp. 318-335, doi:10.1108/itp-12-2018-0577.

[39] P. Praveen, V. Kumar and M. Subramaniam, “How legacy firms can embrace the digital ecosystem via digital customer orientation”, Journal of the Academy of Marketing Science, 2019, pp. 114-131, doi:10.1007/s11747-019-00694-2.

[40] S. Nambisan, M. Wright and M. Feldman, “The digital transformation of innovation and entrepreneurship: Progress, challenges, and key themes”, Research Policy, vol. 48(8), 2019, pp. 103773, doi: 10.1016/j.respol.2019.03.018.

 ISBN: 978-1-61208-869-3

The Design of a Framework for the Detection of Web-Based Dark Patterns

Andrea Curley, Dympna O'Sullivan, Damian Gordon, Brendan Tierney, Ioannis Stavrakakis
ASCNet Research Group, School of Computer Science, Technological University Dublin, Dublin, Ireland
Email: Andrea.F.Curley@TUDublin.ie, Dympna.OSullivan@TUDublin.ie, Damian.X.Gordon@TUDublin.ie,
Brendan.Tierney@TUDublin.ie, Ioannis.Stavrakakis@TUDublin.ie

***Abstract*— In the theories of User Interfaces (UI) and User Experience (UX), the goal is generally to help understand the needs of users and how software can be best configured to optimize how the users can interact with it by removing any unnecessary barriers. However, some systems are designed to make people unwillingly agree to share more data than they intend to, or to spend more money than they plan to, using deception or other psychological nudges. User Interface experts have categorized a number of these tricks that are commonly used and have called them Dark Patterns. Dark Patterns are varied in their form and what they do, and the goal of this research is to design and develop a framework for automated detection of potential instances of web-based dark patterns. To achieve this we explore each of the many canonical dark patterns and identify whether or not it is technically possible to automatically detect that particular pattern. Some patterns are easier to detect than others, and there others that are impossible to detect in an automated fashion. For example, some patterns are straightforward and use confusing terminology to flummox the users, e.g. "Click here if you do not wish to opt out of our mailing list", and these are reasonably simple to detect, whereas others, for example, sites that prevent users from doing a price comparison with similar products might not be readily detectable. This paper presents a framework to automatically detect dark patterns. We present and analyze known dark patterns in terms of whether they can be: (1) detected in an automated way (either partially or fully), (2) detected in a manual way (either partially or fully) and (3) cannot be detected at all. We present the results of our analysis and outline a proposed software tool to detect dark patterns on websites, social media platforms and mobile applications.**

Keywords-Dark Patterns; User Experience; Digital Ethics; Privacy.

I. INTRODUCTION

Computers and technological applications are now central to many aspects of life and society, from industry and commerce, government, research, education, medicine, communication, and entertainment systems. Computer scientists and professionals from related disciplines who design and develop computer applications have a significant responsibility as the systems they develop can have wide ranging impacts on society where those impacts can be beneficial but may also at times be negative. Grosz et al. [17] argue that modern technology cannot be considered "value-neutral" (p. 54); as it can have unplanned negative consequences.

In this paper, we outline and explore the ethical limits of a technology design phenomenon known as "dark patterns". Dark patterns are user interfaces that benefit an online service by leading users into making decisions they might not otherwise make. At best, dark patterns annoy and frustrate users. At worst, they can mislead and deceive users, e.g., by causing financial loss, tricking users into giving up vast amounts of personal data or inducing compulsive and addictive behavior in adults and children. They are an increasingly common occurrence on digital platforms including social media sites, shopping websites, mobile apps, and video games. Although they are gaining more mainstream awareness in the research community, dark patterns are the result of three decades-long trends: one from the world of retail (deceptive practices), one from research and public policy (nudging), and the third from the design community (growth hacking) [26].

The aim of our work is the development of a framework for detecting web-based dark patterns. The framework forms the basis of a software tool that can automatically alert users to the presence of dark patterns on websites, social media platforms and mobile applications. In developing the framework we analysed common documented types of data patterns. We present these dark patterns to the reader and classify each dark pattern using the following taxonomy: (1) A pattern that can be detected in an automated way (either partially or fully); (2) A pattern that can be detected in a manual way (either partially or fully); and (3) A pattern that cannot be detected. These classifications dictate the type of automated software. In this paper we outline the features and functionality of the proposed tool. This research is part of a larger research project (called Ethics4EU) whose goal is develop a repository of teaching and assessment resources to support the teaching of ethics in computer science courses, supported by the Erasmus+ programme [28].

In Section 2, a review of some of the key literature focussing on what dark patterns are, and why they are so successful. Section 3 looks at the specific collection of dark patterns that will be explored in this research. Section 4 presents the initial framework for the detect of dark patterns, looking at which patterns can be detected automatically, which manually, and which cannot be detected at all. Section 5 outlines some other dark patterns that should also be

 ISBN: 978-1-61208-869-3

looked at, and finally, Section 6 presents some conclusions and future work about this research.

II. LITERATURE REVIEW

Since the early 1980s computer programmers have used the concept of patterns in software engineering as a useful way of categorizing different types of computer programs. The term dark patterns has been used since 2010 to refer to interface design solutions that intend to deceive users into carrying out undesirable actions [9]. Gray, et al. [16] defined dark patterns as "instances where designers use their knowledge of human behavior (e.g., psychology) and the desires of end users to implement deceptive functionality that is not in the user's best interest".

There has been significant research done on dark patterns from the fields of Cognitive Psychology, Usability, Marketing, Behavioural Economics, Design and Digital Media. All this research has led to the abandonment of the rational choice theories for explaining decision making, particularly for matters of privacy [2] and has prompted new examinations that attribute the effectiveness of dark patterns on human cognitive limitations. However, there is still not a universal theoretical explanation of the 'whys' and 'hows' of the effectiveness of dark patterns. For example, Maier [22] argues that manipulation is closely linked to decision making and the latter can be easily influenced through one's emotions and mood leading to decisions lacking rational thought [24].

What is more, according to Kahneman [19] humans are more intuitive than rational thinkers and most of their daily reasoning is performed by their intuition. Below are the main human psychological mechanisms being targeted or exploited by Dark Patterns [10]:

- Nudging, which is based on soft paternalism, positive reinforcement and compliance [1]. Nudging can be and has been used with good intentions in mind and has been proved effective [7][29]. However, because of its proven efficiency, nudging is one of the most common digital manipulation strategies used to mislead users into bad decisions privacy-wise.
- Persuasion techniques built on what Cialdini [12] identifies as the "six basic tendencies of human behaviour" [13] (p. 76). These tendencies namely are: reciprocation, consistency, social validation, liking, authority and scarcity.
- Cognitive biases that fundamentally are information processing limitations of the human mind and are rooted in cognitive heuristic systems [19]. According to Waldman [33] the five most pervasive are: anchoring [8], framing [5], hyperbolic discounting [4][30][35], overchoice [11][14][18][25] and metacognitive processes such as cognitive scarcity [33] and cognitive absorption [6].
- Cognitive dissonance, an uncomfortable state of mind where one's beliefs and actions are contradictory. Bosch et al. [10] (p. 247) mention "[i]n terms of privacy dark patterns, this process can be exploited by inconspicuously providing justification arguments for sugar-coating user decisions that have negatively affected their privacy".

Although, so far, it appears that the cognitive and psychological factors play a significantly important role on users' failure to protect their privacy when dealing with Dark Patterns, some researchers argue that contextual and social factors are important too. For example, Acquisti et al. [2] claim that incomplete or asymmetric access to information between two agents in a transaction can significantly disadvantage one party leading to problematic decisions. Furthermore, users are not always certain of what they are agreeing to share as the collection of personal data is not always apparent and therefore people remain unaware of what information is collected about them by both private and public organisations [3]. This is usually the norm in digital environments where the user has no control over the design and information processing they are being shown.

On the other hand, research has shown that users, care about their privacy [20], however, the contextual, social and cognitive aspects mentioned earlier lead users to a set of behaviours that are inconsistent to their attitudes towards privacy [27][34]. Norberg et al. [27] have called this the 'privacy paradox'.

In today's digital environment most digital platforms' provide services seemingly for free. In order for these services to generate revenue they have become dependent on accumulating and processing users' data, oftentimes personal data [15]. According to Zuboff [37] user data is the raw material that produces, what she calls, 'behavioural surplus' which has become a valuable commodity for companies. Behavioural surplus is a powerful tool for predicting user behaviour and many companies use it to influence users into providing more data which leads into a vicious cycle of user data, influence, prediction and so on [32].

Mathur et al. [23] did a meta-analysis of 11,286 shopping websites, and found that 11.1% (1254 websites) of the sites had dark patterns, and recommend the development of plug-ins for browsers to help detect these patterns.

Dark patterns are only just beginning to emerge as a topic in the software development literature. In 2021 Kollnig et al. [21] reported in the development of a functional prototype that allows users to disable dark patterns in apps selectively. This differs from our approach where we are developing a comprehensive framework for identifying dark patterns across a range of platforms, from apps to websites.

III. PATTERN DESCRIPTIONS

A vital step in developing the web-based Dark Patterns Framework is to clearly define each pattern and to categorize the patterns into themes. In the research literature previously discussed there is some variance as to the exact meaning of each pattern, therefore below we present definitions that attempt to be as inclusive as possible to the range of definitions for each pattern, but always prioritising the original canonical definitions developed by the pioneer of dark patterns - user experience designer Harry Brignull [9].

 ISBN: 978-1-61208-869-3

A. Sneaking

Sneak into Basket: When purchasing a product, an additional item is added into the basket, usually the new product is added in because of an obscured opt-out button or checkbox on a previous page.

Hidden Costs: When reaching the last step of the checkout process, some unexpected charges have appeared in the basket, e.g. delivery charges, tax, etc.

B. Misdiretion

Trick Questions: Often found when registering for a new service. Typically, a series of checkboxes are shown, and the meaning of checkboxes is alternated so that ticking the first one means "opt out" and the second means "opt in"

Misdirection: When the design purposefully focuses users' attention on one thing in order to distract their attention from another, for example, a website may have already undertaken a function and added a cost to it, and the opt out button is small.

Confirmshaming: This involves guilting the user into opting into something. The option to decline is worded in such a way as to shame the user into compliance, for example, "No thanks, I don't want to have unlimited free deliveries".

Disguised Ads: Advertisements that are disguised as other kinds of content or navigation, in order to get you to click on them, for example, advertisements that look like a "download" button or a "Next >" button

C. Obstruction

Roach Motel: When users find it easy to subscribe to a service (for example, a premium service), and find it is hard to get out of it, like trying to cancel a shopping account.

D. Forced Action

Forced Continuity: When a user gets a free trial with a service comes to an end and their credit card silently starts getting charged without any warning, and there isn't an easy way to cancel the automatic renewal.

E. Variegations

Privacy Zuckering: Tricking users into sharing more information than they intended to, for example, Facebook privacy settings were historically difficult to control.

Price Comparison Prevention: The retailer makes it hard for you to compare the price of an item with another item, so you cannot make an informed decision. Retailers typically achieve this by creating different bundles where it is not easy to work out the unit price of the items within the bundles.

Bait and Switch: The user sets out to do one thing, but a different, undesirable thing happens instead, for example, Microsoft's strategy to get users to upgrade their computers to Windows 10.

Friend Spam: The product asks for users for their email or social media permissions to spam all their contacts.

IV. Developing The Framework

With these definitions established, it becomes possible to categorize the patterns into one of the following three classifications: (1) A suspected pattern that can be detected in an automated way (partially or fully) based on the text, images or HTML in a webpage or website. (2) A suspected pattern that can be detected in a manual way (partially or fully) based on the text, images or HTML in a webpage or website.(3) A suspected pattern that cannot be detected, based on the fact that there is so much variation in either how the pattern is defined or in how the pattern is implemented.

As all of the researchers involved in this project are teaching on an MSc in Data Science, they have knowledge of a wide range of detection techniques, therefore, a Morphological Matrix approach [36] was undertaken, whereby a table was created listing all of the pattern types on the Y-axis, and listing a range of detection techniques on the X-axis (HTML Parsing, Computational Linguistics, Image Processing, Machine Learning, Data Mining, Compiler Design, Regular Expressions) and a series of three online brainstorming sessions were held to identify which patterns might be detectable using which techniques (if any). To help reach a shared understanding of the patterns, not only were definitions of each pattern shared and discussed, but also images from over 100 websites with dark patterns from the Mathur et al. [23] dataset were presented and discussed. Of all patterns discussed, there was general consesus as to which aspects of patterns could be detected, how they could be detected and to what extent that detection was possible. The full framework is presented below in Table 1 where each of the patterns presented in Section III is classified as to how it can be detected (automated, manually or cannot be detected), as well as some detail as to how such a pattern can be detected (or, in fact, if it cannot be detected) as shown in the *Rationale* column.

Patterns that can be detected automatically will typically have terms in them such as "opt-in", "activate", or "subscribe". These, and other indicators such as the placement or configuration of images, or in the formulation of the HTML tags, allow for the automated detection of dark patterns. In contrast, there are some web-based activities or transactions that cannot, in and of themselves, be automatically detected, but are sufficiently indicative to suggest the presence of a dark pattern. In these cases the framework proposes the development of an ancillary (or appurtenant) window to highlight to the users that there may be something suspicious occurring in the transaction that they are undertaking. Finally, it is worth noting that, there are some patterns that cannot readily be detected, but may be reported using the reporting feature of the system.

 ISBN: 978-1-61208-869-3

TABLE I. DARK PATTERNS DETECTION FRAMEWORK

Category	*Pattern*	*Detection*	*Rationale*
Sneaking	Sneak into Basket	Manual (fully)	Highlight changes in cost
	Hidden Costs	Manual (fully)	Highlight changes in cost
Misdirection	Trick Questions	Automated (partially)	Look for phrases like "opt-in" and "opt-out", as well as pre-ticked checkboxes
	Misdirection	Cannot be detected	There is too much variation in how this pattern is implemented.
	Confirmshaming	Cannot be detected	There is too much variation in how this pattern is implemented.
	Disguised Ads	Automated (partially)	Look for buttons (noting colour and size) and see which ones link to external sites.
Obstruction	Roach Motel	Automated (fully)	Look for sites with "activate" or "subscribe" links or buttons but with no "deactivate" or "unsubscribe"
Forced Action	Forced Continuity	Cannot be detected	There is too much variation in how this pattern is implemented.
Variegations	Privacy Zuckering	Cannot be detected	There is too much variation in how this pattern is implemented.
	Price Comparison Prevention	Manual (fully)	Highlight if products are displayed with different units of the product
	Bait and Switch	Cannot be detected	There is too much variation in how this pattern is implemented.
	Friend Spam	Automated (partially)	Check if the site asks for email or social media permissions, and notify users.

Some patterns will have words or images that make them easy to identify ("opt in", "offer ends soon", "in demand", etc.) and therefore we can say that they are automatically detectable (either partially or fully). And, in contrast, some patterns are implemented in such a range of different ways depending on the particular interface (and the definitions of some patterns vary in different research literature), that they are impossible to consistently detect, so we classify these as "Cannot be detected". Other patterns require human judgement, such as determining if using pre-ticked checkboxes is being deceptive, or if the site is asking for security permissions, and so we classify these as being detectable manually (either partially or fully). To help recognise the patterns that can potentially be manually detected, the proposed system will allow the user to display an ancillary window that will help highlight some potential issues of concern on a given webpage or website. The new window can display things like:

- The percentage of the webpage that is visible in the browser window, to ensure the user is aware that there may be instructions or options that are not visible on the current page, but are elsewhere on the page.
- The total number of checkboxes on the page, and the number that are pre-ticked.
- The total number of radio buttons on the page, and the number that are pre-ticked.
- The shopping basket total, that will be zero if there are no items.
- A "fake review detection" tool that allows a user to select the text of a review, and to automatically search for that text elsewhere on the web.
- Highlight the number of links on the page, noting which are from text and which from images (to help detect potential Disguised Ads).
- Highlight which tick boxes or radio buttons are concerned with privacy issues, looking for words such as "privacy" or "GDPR".
- Indicate if the current webpage or website has already been reported as having a dark pattern.

Further, to help users locate suspected dark patterns on a webpage, the system will provide two modes of operation:

(1) where the system highlights all of the areas on that webpage to show suspected patterns on the page with suitable pointers, and
(2) if the user clicks on a particular type of issue on the auxiliary window, only those areas on the page will be highlighted, for example, if the user selects the "Radio Buttons" section of the panel, then all of the radio buttons on the webpage will be highlighted with pointers.

 ISBN: 978-1-61208-869-3

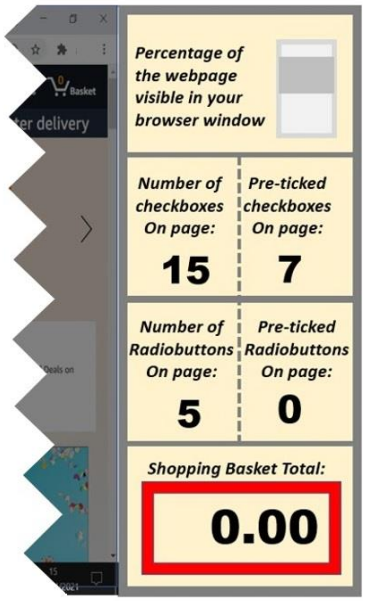

Figure 1. Appurtenant Window with Page Details

Two additional elements of the proposed system are the Reporting and Educational features:

- The *Reporting Feature* is designed to compensate for the fact that some patterns are difficult (or impossible) to detect, and it will allow users to record and report websites and webpages that they suspect have dark patterns. For example, if a user feels that they have been a victim of Forced Continuity, they can report the webpage or website, and indicate which pattern they feel is present.
- The *Educational Feature* which is designed to educate the users on each of the main dark patterns, as well as the variation among different researchers. This feature will help the users appreciate why they are being warned about a particular feature on a website as well as giving them sufficient information to allow them to accurately categorize patterns that they encounter if they wish to report them. It is envisioned that a central part of this feature will consist of a series of videoed micro-lessons.

V. IMPLEMENTATION AND LIMITATIONS

The goal of this research is to define a collection of dark patterns, and to explore whether or not it is possible to develop a framework to detect these dark patterns - in an automated way, a manual way, or not at all. The detection process not only categorizes whether each pattern is detectable, but it also describes to what extend it is detectable, and suggests some ways it might be detected. The development process of framework was as a result of the brainstorming sessions, and these crucially categorized the patterns into three groupings:

1. Automated Detection ("Disguised Ads", "Friend Spam", "Roach Motel" and "Trick Questions")
2. Manual Detection ("Hidden Costs", "Price Comparison Prevention", "Sneak into Basket")
3. Cannot be Detected ("Bait and Switch", "Confirmshaming", "Forced Continuity", "Misdirection", "Privacy Zuckering")

To help confirm the analysis process, an initial prototype system has been developed using the Python programming language which provides ample software libraries for web crawling and web scraping, specifically the HTMLparser and URLopen libraries were used in this case. Three patterns were selected to be implemented, “Trick Questions”, “Roach Motel”, and “Friend Spam” were chosen as they are the most straightforward to implement, since that have been classified as “Automated (partial)” and “Automated (fully)” in the above table. These three were implemented as separate methods in a single class. These methods were tested using over 60 of the dark patterns from the Mathur et al. [23] dataset, and the prototype was able to successfully detect all three of these patterns, each with significant variation. Early work has been undertaken on the development of methods to detect the other patterns classified as “Automated” as well as work on the “Manual” pattern types. Further implementation will explore additional dark patterns identified by other researchers, for example, UX Researcher, Reed Steiner [31] highlighted six more patterns, as follows:

- **Fake Activity:** On a commercial website, when the page says “three other people are viewing this item right now” this may not be a fully truthful claim.
- **Fake Reviews:** Research shows that several reviews and testimonials are fake, and exact matches with different customer names can be found on several sites.
- **Fake Countdown:** Some online purchases include countdown timers, in most cases countdown timers only add urgency to a sale.
- **Ambiguous Deadlines:** Some online purchases indicate that a product is only on sale for a limited amount of time, but don’t mention a specific deadline.
- **Low Stock Messages**: Sometimes sites claim that they are low on a particular item.
- **Deceptive High Demand:** Similar to low stock messages

These are commonly occurring patterns, and also worth exploring in more detail.

In terms of the limitations of this research, perhaps the most serious one is the fact that five of the patterns (“Misdirection”, “Confirmshaming”, “Forced Continuity”, “Privacy Zuckering”, and “Bait and Switch”) have been classified as “Cannot be detected”. If these cannot be detected, it significantly limits the efficacy of the final tool, therefore a thorough exploration of the Mathur et al. [23] dataset is planned to determine if there are any implicit characteristics associated with these five patterns that can be

 ISBN: 978-1-61208-869-3

used to detect them (either automatically or manually), as well as a number of further brainstorming sessions.

It is also worth noting that that the full implementation of this framework will result in some additional challenges, for example, some sites have a special file called Robots.txt that prohibits the use of web scraping, and it is also the case that some sites use technologies that make them more difficult to parse, for example, frames or webpages implemented in Javascript or CSS.

Finally, another consideration is that many shoppers use mobile applications instead of websites to purchase products and services, and the techniques outlined so far would be ineffective on these applications.

VI. CONCLUSIONS AND FUTRE WORK

This paper presented a framework for the detection of web-based dark patterns and an accompanying proposed software tool. It begins with a review of some of the key literature in this field, which highlights some of the reasons for the success of dark patterns, as well as their ubiquity. It follows this with an explanation of some of the key dark patterns, and a categorization of the patterns as being in one of the following three classifications:

1. A suspected pattern that can be detected in an automated way (partially or fully), in other words there is some characteristic either in the text, images or HTML of a webpage or website that indicates that it is a dark pattern.
2. A suspected pattern that can be detected in a manual way (partially or fully), in other words there is some characteristic either in the text, images or HTML of a webpage or website that indicates that there is potential for dark pattern on this page or site, but because it cannot be detected definitively, the potential pattern is highlighted to the user.
3. A suspected pattern that cannot be detected, in other words there is so much variation in either how the pattern is defined or in how the pattern is implemented, there is no direct way of detecting it just using web crawling and web scraping techniques.

This classification, in turn, leads to the design of a proposed software tool with the ability to detect patterns from category 1, and to highlight potential instances of patterns from category 2. For those patterns in category 3, even if there is no obvious way to identify them, nonetheless, it is important to deal with them in some way, therefore additional features are required for the system, a *Reporting feature* to address instances of patterns for category 3, as well as an *Educational feature* to create awareness about dark patterns in general.

Future work will focus on full implementation of the software tool and the inclusion of the Reporting and Education features. The Reporting features of the system are envisioned to work either in *stand-alone mode*, or *shared mode*. In stand-alone mode the reporting process is recorded locally on the user's own computer as a series of XML files, whereas in shared mode, the user can share their suspicions about potential dark patterns with other users also using the system, and they can also label and add a description to the suspected pattern.

The Educational features will consist of a series of micro-lessons describing the range of dark patterns. Also, a series of pop-up windows will be developed with simple explanations (and links to examples) of a specific pattern will be developed, to remind the users about the key characteristics of each specific pattern.

Finally, the framework provides a way forward to deal with dark patterns in a comprehensive and comprehensible manner. This has become more and more important as the number of services that have become available online continues to grow, and in many cases these services are available only exclusively online. It, therefore, becomes a matter of necessity that as many people as possible are aware of these deceitful patterns, and incumbent on IT practitioners to spread the word about these patterns.

ACKNOWLEDGMENT

The authors of this paper and the participants of the Ethics4EU project gratefully acknowledge the support of the Erasmus+ programme of the European Union. The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

REFERENCES

[1] A. Acquisti, "Nudging privacy: The behavioral economics of personal information". IEEE Security & Privacy, 7(6), pp. 82-85. 2009.

[2] A. Acquisti, I. Adjerid, R. Balebako, L. Brandimarte, L. F. Cranor, S. Komanduri, S. Wilson "Nudges for privacy and security: Understanding and assisting users' choices online", ACM Computing Surveys (CSUR), 50(3), pp. 1-41, 2017.

[3] A. Acquisti, L. Brandimarte, G. Loewenstein, "Privacy and human behavior in the age of information". *Science*, *347*(6221), pp. 509-514, 2015.

[4] A. Acquisti, J. Grossklags "Losses, gains, and hyperbolic discounting: An experimental approach to information security attitudes and behavior". In 2nd Annual Workshop on Economics and Information Security-WEIS (Vol. 3, pp. 1-27), 2003.

[5] I. Adjerid, A. Acquisti, L. Brandimarte, G. Loewenstein, "Sleights of privacy: Framing, disclosures, and the limits of transparency". In Proceedings of the ninth symposium on usable privacy and security, pp. 1-11. 2013.

[6] T. Alashoor, R. Baskerville, "The privacy paradox: The role of cognitive absorption in the social networking activity". In Thirty Sixth International Conference on Information Systems, Fort Worth, Texas, USA, pp. 1–20, 2015.

[7] H. Almuhimedi, F. Schaub, N. Sadeh, N.I. Adjerid, A. Acquisti, J. Gluck, Y. Agarwal, "Your location has been shared 5,398 times! A field study on mobile app privacy

nudging". In Proceedings of the 33rd annual ACM conference on human factors in computing systems, pp. 787-796, 2015.

[8] D. Ariely, G. Loewenstein, D. Prelec, "Coherent arbitrariness: Stable demand curves without stable preferences". The Quarterly journal of economics, 118(1), pp. 73-106, 2003.

[9] H. Brignull, "Dark patterns: Deception vs. honesty in UI design". Interaction Design, Usability, 338, 2011.

[10] C. Bösch, B. Erb, F. Kargl, H. Kopp, S. Pfattheicher, "Tales from the dark side: Privacy dark strategies and privacy dark patterns". Proceedings on Privacy Enhancing Technologies, 2016(4), pp. 237-254, 2016.

[11] A. Chernev, U. Böckenholt, J. Goodman, "Choice overload: A conceptual review and meta-analysis". Journal of Consumer Psychology, 25(2), pp. 333-358, 2015.

[12] R. Cialdini, "Influence. The Psychology of Persuasion". New York, NY: William Morrowe Company, 1984.

[13] R. B. Cialdini, "The science of persuasion". Scientific American, 284(2), pp. 76-81, 2001.

[14] J. D. D'Angelo, C. L. Toma, "There are plenty of fish in the sea: The effects of choice overload and reversibility on online daters' satisfaction with selected partners". Media Psychology, 20(1), pp. 1-27, 2017.

[15] GDPR, EU General Data Protection Regulation (GDPR): Regulation (EU) 2016/679 of the European Parliament, Article 4, 2016, https://gdpr-info.eu/art-4-gdpr/

[16] C. M. Gray, Y. Kou, Y., B. Battles, J. Hoggatt, A. L. Toombs,"The dark (patterns) side of UX design". In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, pp. 1-14, 2018.

[17] B. J. Grosz, D. G. Grant, K. Vredenburgh, J. Behrends, L. Hu, A. Simmons, J. Waldo, "Embedded EthiCS: Integrating Ethics across CS Education". Commun. ACM, 62(8), pp. 54–61, 2019. https://doi.org/10.1145/3330794

[18] S. Jilke, G. G. Van Ryzin, G. G., S. Van de Walle, "Responses to decline in marketized public services: An experimental evaluation of choice overload". J. of Public Administration Research & Theory, 26(3), pp. 421-432, 2016.

[19] D. Kahneman, "Thinking, Fast and Slow", Penguin Books, 2011.

[20] S. Kokolakis, "Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon". Computers & security, 64, pp. 122-134, 2017.

[21] K. Kollnig, S. Datta, M. Van Kleek, "I Want My App That Way: Reclaiming Sovereignty Over Personal Devices", 2021, arXiv preprint arXiv:2102.11819.

[22] M. Maier, "Dark patterns: An end user perspective". Master's thesis. Umeå University, 2019.

[23] A. Mathur, G. Acar, M. J. Friedman, E. Lucherini, J. Mayer, J., M. Chetty, A. Narayanan, A. "Dark patterns at scale: Findings from a crawl of 11K shopping websites". Proceedings of the ACM on Human-Computer Interaction, 3(CSCW), pp. 1-32, 2019.

[24] R. Mehta, R. J. Zhu, "Blue or red? Exploring the effect of color on cognitive task performances", Science (New York, N.Y.), 323(5918), pp. 1226–1229, 2009.

[25] K. Nagar, P. Gandotra, "Exploring choice overload, internet shopping anxiety, variety seeking and online shopping adoption relationship: Evidence from online fashion stores". Global Business Review, 17(4), pp. 851-869, 2016.

[26] A. Narayanan, A. Mathur, M. Chetty, M. Kshirsagar, "Dark Patterns: Past, Present, and Future: The evolution of tricky user interfaces". *Queue*, *18*(2), pp. 67-92, 2020.

[27] P. A. Norberg, D. R. Horne, D. A. Horne, "The privacy paradox: Personal information disclosure intentions versus behaviors". Journal of consumer affairs, 41(1), pp. 100-126, 2007.

[28] D. O'Sullivan, D. Gordon, "Check Your Tech – Considering the Provenance of Data Used to Build Digital Products and Services: Case Studies and an Ethical CheckSheet", IFIP WG 9.4 European Conference on the Social Implications of Computers in Developing Countries, 10th–11th June 2020, Salford, UK, 2020.

[29] E. Peer, S. Egelman, M. Harbach, N. Malkin, A. Mathur, A. Frik, "Nudge me right: Personalizing online security nudges to people's decision-making styles". Computers in Human Behavior, 109, 106347, 2020.

[30] J. Puaschunder, "Towards a utility theory of privacy and information sharing and the introduction of hyper-hyperbolic discounting in the digital big data age". In Handbook of research on social and organizational dynamics in the digital era, pp. 157-200, IGI Global, 2020.

[31] R. Steiner, "Dark Patterns" . [Online]. Available from: https://www.fyresite.com/dark-patterns-a-new-scientific-look-at-ux-deception/, 2021.06.24

[32] M. Van Otterlo, "Automated experimentation in Walden 3.0.: The next step in profiling, predicting, control and surveillance". Surveillance & society, 12(2), pp. 255-272, 2014.

[33] G. A. Veltri, A. Ivchenko, "The impact of different forms of cognitive scarcity on online privacy disclosure". Computers in human behavior, 73, pp. 238-246, 2017.

[34] A. E. Waldman, "Cognitive biases, dark patterns, and the 'privacy paradox". Current opinion in psychology, 31, pp. 105-109, 2020.

[35] Y. Wang, G. Norcie, S. Komanduri, A. Acquisti, P. Leon, L. F. Cranor, "I regretted the minute I pressed share" a qualitative study of regrets on Facebook. In Proceedings of the seventh symposium on usable privacy and security, pp. 1-16, 2011.

[36] R. G. Weber., S. S. Condoor. "Conceptual design using a synergistically compatible morphological matrix." In FIE'98. 28th Annual Frontiers in Education Conference. Moving from'Teacher-Centered'to'Learner-Centered'Education. Conference Proceedings (Cat. No. 98CH36214), vol. 1, pp. 171-176. IEEE, 1998.

[37] S. Zuboff, "The Age of Surveillance Capitalism: The Fight for Human Future at the New Frontier of Power". London: Profile Books, ISBN 978-1-7881-6316-3, 2019.

Public Acceptance of Robots: Drivers and Barriers

Valentina Polyakova
Institute for Statistical Studies and Economics of Knowledge
National Research University Higher School of Economics
Moscow, Russia
e-mail: vpoliakova@hse.ru

Abstract — **Robots as companions, domestic assistants and nurses have great potential for the care of people who need assistance. At the same time in popular culture, technological progress is often represented as a threat to human distinctiveness. Studies demonstrate limited confidence of individuals in robotics and artificial intelligence. The author will consider how perceived risks and benefits, attitudes toward science and technology, social bonds and other individual socio-demographic characteristics influence on the public acceptance of robots.**

Keywords - public acceptance of technologies; robots; social attitudes; social bonds; trust; digitalization.

I. INTRODUCTION

Technology is one of the main factors of socio-economic development [7] that produces various effects: robotization and digitalization of production and organizational processes, increase diversity of digital goods and services, growth of demand for digital skills [12]. The spread of robotics in industry, as well as in everyday life is bringing us closer to the future described in science fiction. People are increasingly relying on artificial intelligence to process large amounts of data, make decisions about the solvency of borrowers, select staff and choose the best couple on dating sites.

Robots as companions, domestic assistants and nurses have great prospects for the care of people who need assistance (e.g., the elderly or sick people). At the same time studies demonstrate limited confidence in robotics and artificial intelligence [5][14]. In popular culture, technological progress is often represented as a threat to human distinctiveness [2][3]. The culturally ingrained fear of autonomous technologies has been called the "Frankenstein syndrome" [8]. A similar aspect is evident in the Uncanny valley effect of perception of technological objects, which was firstly described by Japanese robotics scientist and engineer Masahiro Mori [10][11]. The phenomenon implies that humans tend to dislike and detest robots and other objects that look or act roughly like humans (but not exactly like the real ones).

It is worth noting that this effect is not universal and has gender and age variations. Moreover some studies support its existence [9], while others do not [16]. Perhaps if the place of technology in our lives and attitudes to it changed, this effect would also disappear.

Thus, there is a certain paradox in public opinion. On one hand, the development of robotics and artificial intelligence is on the way to improving the capacity for self-learning and independent decision-making. On the other hand, people fear the autonomous technologies and are not ready to rely on them.

The remainder of this paper is organized as follows. Section 2 describes research approach and the applied methods, as well as presents key results of the analysis. Section 3 summarizes the main outputs of this study, encompass the limitations of the study and possible avenues for future research.

II. MAIN IDEA

The purpose of this research is to analyze social attitudes toward the use of artificial intelligence and robots for various tasks and the factors that have an influence on the public acceptance of autonomous technologies. The model of this study is based on the conception of public acceptance of technologies that is defined as the readiness to use a technology to solve tasks assigned to it [1]. Factors influencing technology acceptance are defined at micro- and macro-level. In this research, we analyze the effects of micro-level factors, which includes the indicators of the perception of robotics and technologies in general and social characteristics of the user. Studies of technology perception highlight such determinants as perceived risks and benefits, level of knowledge, attitudes toward science and technology, trust, values, and other individual socio-demographic differences [6]. Attention to the risks posed by the diffusion of new technologies (in reality or in imagination) contributes to the rejection of innovation, while understanding the benefits/benefits of use reduces the level of anxiety about it [4][13].

Data was collected in December 2018 – January 2019 during the 27th wave of the Russia longitudinal monitoring survey, which is a series of national representative surveys based on probabilistic stratified multistage territorial sample. The sample size is 7584 respondents aged 18-65. Method – face-to-face interviews.

The public acceptance of robots was assessed through respondents' estimations of 10 situations representing a robot in different roles. Situations were differentiated by the type of tasks (functional and social) and by the strength of influence on the user (assistance or dependence). Respondents were asked to rate the degree of perceived

 ISBN: 978-1-61208-869-3

comfort/discomfort in each of these situations on a four-point scale. The author used a binary logistic regression method to identify factors that influence robot acceptance in different situations.

The results demonstrate that there is a same dualism in public opinion in Russia, as in other countries. Positive attitudes toward robots as are a good thing for society are paired with an expectation of threat from it (54% vs. 63%) and a corresponding low level of acceptance of robots as autonomous actor. The use of robots in functional-assisting role as domestic helpers or delivery drones is rather acceptable (66% and 62% of respondents feel comfortable about these situations). At the same time, most respondents are unready for a high degree of robot autonomy and agency, and are not disposed to delegate responsibility for their lives or the lives of family members (driverless car, the use of robots in elderly parents care and in surgery seem comfortable for 21%, 19% and 14% of respondents).

According to the results of the regression, analysis general drivers of public acceptance of robots are confidence in one's own power and ability to influence the state of affairs, digital skills (as an indicator of digital adoption), engagement with science, positive attitudes to robots and belief in the robotization of human labor. General barriers to the public acceptance are science awareness and expectation of a threat from robots (lack of trust).

In addition, some factors enhance or reduce attitudes to robots only in particular situations. For example, gender have an effect (negative) only on acceptance of robots for assistance at home and delivery. Significant generational differences in attitudes are observed only in relation to driverless car.

Situation-specific predictors relate more to different aspects of well-being and personal characteristics. Health problems, selective trust in people and moderate loneliness contribute to a positive attitude toward the use of robots for assistance at home and delivery, while predisposition to trust most people and to value the spiritual aspects of life, exclusion from innovation consumption, living in low urbanized areas, an acknowledgement of the threat of human rights violations. Nevertheless, risks are not always negatively associated with the acceptance of robots. The effect depends on the type of risk and its specification. For example, understanding the risks of technology use by criminals has a positive effect on the attitude to robotic home assistants.

A significant negative effect of sensitivity to change due to technological advances on the acceptance of robots has been detected only in relation to their use for elderly care and surgery. Additional barriers to the adoption of robots for elderly care are traditional family ties and religion.

If the use of robots to care for the elderly is perceived in the context of social exclusion, robotic surgeon is too radical innovation in the eyes of the average person. Experience of innovation generation contributes to the acceptance of roboticization of surgery.

III. CONCLUSION

Nowadays, the society accepts only the idea of automating (human-controlled) certain processes with digital technology, but is not ready for fully autonomous digital technology. In this context, social attitudes towards autonomous technologies can be seen as an indicator that reveals the depth of changes in public consciousness in the digitalization of the society.

Technocentrism as characteristic of modern civilization is an important prerequisite in the formation of society's readiness for autonomous technologies. At the same time, the determining role is played not so much by the "pro-science" values of the population, as by the presence of an active interest in scientific and technological progress and involvement in the system of communications in this sphere.

However, awareness of science news has the opposite effect - the development of critical reflection on the consequences of the introduction of new technologies in the conditions of growing consciousness and resistance to progress.

Another important barrier to social integration of robots is the social bonds between people, trust, patterns and norms of relationships. Adherence to traditional family values of cohesion and mutual assistance, and having close relationships with other people are at odds with the idea of using robots to care for the elderly, as it is perceived as exclusion.

The resolution of the paradox lies in the sociocultural field. The results of the study indicate that the transformation of the digital environment from alien to natural may be one of the signaling factors that contribute to a change in the perception of autonomous digital technologies. According to the data, people today still predominantly consider the digital environment as an artificial phenomenon, depriving the life of the sense of life or acting as a poor substitute. Such a worldview prevents the acceptance of the changing social role of robotics. Therefore, future studies need to estimate the influence of cultural values on acceptance of robots in different situation. In this research, we did not have direct questions on human values and used only proxy indicators of traditional family values and trust.

More research is also needed to uncover the public perception of different types of social robots. In this survey, we used personalized situation (elderly parents care). Probably, impersonal examples of robots (e.g. elderly care in general or elderly care in nursing homes or in health care facilities) will provide other results.

ACKNOWLEDGMENT

The article was prepared in the framework of a research grant funded by the Ministry of Science and Higher Education of the Russian Federation (grant ID: 075-15-2020-928).

 ISBN: 978-1-61208-869-3

REFERENCES

[1] A. Dillon, User Acceptance of Information Technology. In W. Karwowski (Ed.) Encyclopedia of Human Factors and Ergonomics. London: Taylor and Francis, 2001.

[2] D. Dinello, Technophobia: science fiction visions of posthuman technology. University of Texas Press, 2005.

[3] F. Ferrari, M. P. Paladino, and J. Jetten, "Blurring human–machine distinctions: Anthropomorphic appearance in social robots as a threat to human distinctiveness," International Journal of Social Robotics, vol. 8, №2, pp. 287-302, 2016.

[4] G. Gaskell et al., "GM Foods and the Misperception of Risk Perception," Risk Analysis, vol. 24, №1. 2004, pp. 185-194.

[5] T. Gnambs and M. Appel, "Are robots becoming unpopular? Changes in attitudes towards autonomous robotic systems in Europe," Computers in Human Behavior, № 93, pp. 53-61, 2019.

[6] N. Gupta, A. R. H. Fischer, and L. J. Frewer, "Socio-psychological determinants of public acceptance of technologies: a review," Public Understanding of Science, vol. 21, №. 7, pp. 782-795, 2012.

[7] C. I. Jones and P. M. Romer, "The new Kaldor facts: ideas, institutions, population, and human capital," American Economic Journal: Macroeconomics, vol. 2, №. 1, pp. 224-245, 2010.

[8] F. Kaplan, "Who is afraid of the humanoid? Investigating cultural differences in the acceptance of robots," International journal of humanoid robotics, vol. 1, № 03, pp. 465-480, 2004.

[9] W. J. Mitchell, Sr K. A. Szerszen, A. S. Lu, P. W. Schermerhorn, M. Scheutz and K. F. MacDorman, "A mismatch in the human realism of face and voice produces an uncanny valley," i-Perception, vol. 2, №1, pp. 10-12, 2011.

[10] M. Mori, "Bukimi no tani [the uncanny valley]," Energy, vol. 7, pp. 33-35, 1970.

[11] M. Mori, K. F. MacDorman, N. Kageki, "The uncanny valley [from the field]," Proc. IEEE Robotics & Automation Magazine, vol. 19, №2, 2012, pp. 98-100, DOI: 10.1109/MRA.2012.2192811

[12] OECD. *Getting Skills Right: Skills for Jobs Indicators*. Paris: OECD Publishing, 2017.

[13] L. Sjoberg, "Policy Implications of Risk Perception Research: A Case of the Emperor's New Clothes," Risk Management, vol. 4, № 2, pp. 11-20, 2002.

[14] Special Eurobarometer 460. *Attitudes towards the impact of digitization and automation on daily life*. Office for Official Publications of the European Communities, Luxembourg, 2017.

[15] M. Destephe, M. Brandao, T. Kishi, M. Zecca, K.Hashimoto, and A. Takanishi. "Walking in the uncanny valley: importance of the attractiveness on the acceptance of a robot as a working partner," Frontiers in Psychology, vol. 6, pp. 1-11, 2015.

[16] J. C. Thompson, J. G. Trafton, P. McKnight, "The perception of humanness from the movements of synthetic agents," Perception, vol. 40, №6, pp. 695-704, 2011.

 ISBN: 978-1-61208-869-3

Using Stylometric Features to Predict Author Personality Type in Modern Greek Essays

Gagiatsou Sofia
Department of Linguistics, School of Philosophy
National and Kapodistrian University of Athens
Athens, Greece
e-mail: sgagiats@phil.uoa.gr

Markopoulos Georgios
Department of Linguistics, School of Philosophy
National and Kapodistrian University of Athens
Athens, Greece
e-mail: gmarkop@phil.uoa.gr

Mikros George
College of Humanities and Social Sciences
Hamad Bin Khalifa University
Doha, Qatar
e-mail: gmikros@hbku.edu.qa

***Abstract*—We present a research focused on the prediction of the author's personality based on natural language processing techniques applied to essays written in Modern Greek by high-school students. Each writer has been profiled by filling in the Jung Typology Test. In addition, personality prediction is being discussed under the general research framework of author profiling by examining the effectiveness of several stylometric features to predict students' personality types. The feature set we employed was a combination of the word and sentence length, the most frequent part-of-speech tags, most frequent character/word bigrams and trigrams, most frequent words, as well as hapax/dis legomena. Since personality prediction represents a complex multidimensional research problem, we applied various machine learning algorithms to optimize our model's performance after extracting the stylometric features. We compared nine machine learning algorithms and ranked them according to their cross-validated accuracy. The best results were obtained by the Naive Bayes algorithm. According to the personality classification based on the Jung Typology Test, the author's personality prediction accuracy reached 80.7% on Extraversion, 79.9% on Intuition, 68.8% on Feeling, 75.7% on Judging, according to the personality classification. The reported results show a competitive approach to the personality prediction problem. Furthermore, our research revealed new combinations of stylometric features and corresponding computational techniques, giving interesting and satisfying solutions to the problem of the author's personality prediction for Modern Greek.**

Keywords-Author profiling; stylometry; Personality prediction; Jung Typology Test; corpus processing; computational stylistics; machine learning.

I. INTRODUCTION

Authorship identification represents one of the emerging text mining fields that stands at the intersection of Machine Learning, Information Retrieval, and Natural Language Processing. Under the stylometric framework, the author's identity is a multidimensional construct based mainly on writing patterns scattered across multiple linguistic levels and expressed quantitatively. The specific research domain splits into three subdomains: attributing a text to a particular author among a finite set of authors (Authorship Attribution), attributing a text to an author that does not belong to a closed set (Authorship Verification), and specifying the author's metadata such as demographic and psychological traits of the author (Authorship Profiling), including gender, age, personality, etc.

Language as a communication mechanism denotes the diversity of every individual. Therefore, the quantitative study of linguistic features can lead to predictions regarding the individual's character. The subject of Computational Personality Prediction (CPP) through natural language processing techniques constitutes a relatively new research field with many applications.

One critical application domain of this field is Forensic Linguistics. Criminals can be identified by the way they write. Moreover, conclusions can be drawn regarding their personalities and the way they are thinking. The example of identifying students' personality that carry guns and participate in school shootings is typical [1]. CPP can highlight their psychological traits, which afterward can be exploited in the successful identification of potential perpetrators.

Apart from the obvious contribution that provides in Behavioural Psychology by connecting personality traits to human behaviour, CPP can function in many other fields as well. For instance, in the marketing domain, personality analysis of users/consumers is utilized by companies to adopt effective recruitment techniques and customer service

 ISBN: 978-1-61208-869-3

techniques. Even in the field of human resources management, predicting the personality can affect or facilitate the selection and determine the eligibility of candidates for a particular job. Moreover, based on the user's personality, dialogic systems can be customized and brought closer to users' temperament making interaction more effective and satisfying.

Another vital analysis domain where automatic personality prediction is used is education. For example, by analyzing students' writings, talented students or students with difficulties could be recognized and thus receive adaptive teaching, addressing the appropriate cognitive level for each one or each group.

One of the most crucial issues in CPP research is developing appropriate linguistic resources enriched with the author's personality metadata. Unfortunately, these resources are challenging to create due to the increased level of manual interaction with the authors and the various privacy and ethical considerations linked with administering psychometric questionnaires to many individuals.

Another issue is that most Natural Language Processing (NLP) tools specialized in psychometric text profiling support only English. Therefore, research in other languages should be done by developing specialized dictionaries and other supporting linguistics resources from scratch (see, for example, the case of Linguistic Inquiry and Word Count-LIWC [2]).

To cover the above-mentioned research gaps, we performed the first CPP study in Modern Greek focused on high-school students. For this reason, we developed a model for predicting the personality of students based on Jung's taxonomy by analyzing their term-essays and applying various machine learning methods to rich document representation based on several stylometric features.

The rest of this paper is organized as follows. In Section II we provide an overview of previous work on personality prediction. Section III describes our researching methods. In Section IV we present the research results. We summarize our findings and discuss future work in Section V.

II. LITERATURE REVIEW

In this section we present the personality questionnaire used to profile the writers. Then, we review the findings of studies in the field of CPP from text.

A. *Carl Jung's and Isabel Briggs Myers' Personality Type Questionnaire*

Research in the field of personality prediction uses the Five-factor Model of Personality [3] or the Carl Jung's and Isabel Briggs Myers' personality type theory [4][5] to profile the participating authors. Therefore, the literature review presented in this section is referred to associated research, which involves the Jung Typology Test since our students have been profiled with the above-mentioned personality questionnaire.

According to Jung's theory of psychological types [4] people can be characterized by

- their preference of general attitude as Extraverted (E) or Introverted (I)
- their preference of one of the two functions of perception as Sensing (S) or Intuitive (N)
- their preference of one of the two functions of judging as Thinking (T) or Feeling (F)
- their orientation to the outer world as Judging (J) or Perceiving (P).

The Jung Typology Test classifies psychological differences of personality in four dichotomies which yield 16 different combinations or personality types. Each personality type can be assigned a 4-letter acronym of the corresponding combination of preferences: ESTJ, ISTJ, ENTJ, INTJ, ESTP, ISTP, ENTP, INTP, ESFJ, ISFJ, ENFJ, INFJ, ESFP, ISFP, ENFP, INFP.

B. *Personality Research from Text*

One of the first studies related to the author's personality prediction problem [6] defined the research problem as a text categorization task. They developed a corpus consisted of essays written in Dutch by 145 students (BA level). By selecting syntactic features and by training machine learning algorithms, the experiments in personality prediction suggested that the personality dimensions Introverted-Extraverted and iNtuitive-Sensing) can be predicted fairly accurately.

CPP studies have also expanded to social media texts with an emphasis on Twitter. A study for predicting Twitter users' personality type [7] showed that the classifier's performance on training data was quite good. Still, the classifier failed to achieve satisfying results for the test data. Another study [8] describes a logistic regression classifier's training process to predict each of the four dimensions of Jung Typology. Their results showed that linguistic features are the most predictive features. Although they succeeded in distinguishing between the personality dimensions Introverted-Extraverted and Feeling-Thinking, the other two dimensions were hard to predict.

In a study of a multilingual corpus of tweets [9], based on six languages (Dutch, German, French, Italian, Portuguese, and Spanish), the researchers extracted the most frequent word and character n-grams. Their results confirmed the findings of the previous work in that particular personality distinctions could be predicted from social media data with success. In another study focused on tweets [10], the researchers used a Naive Bayes classifier achieving 80% accuracy for Introverted-Extraverted and 60% for the other dimensions.

CPP has also being applied to languages with a different graphemic organization compared to Western languages. For example, in [11], researchers investigate the personality prediction of Twitter users in Japanese and conclude that the textual information of user behaviors is more useful than the users' cooccurrence behavior information such as the likes.

 ISBN: 978-1-61208-869-3

In this study, the problem of author personality prediction was treated as a set of binary classification tasks using Support Vector Machines.

III. CORPUS

To test our research hypothesis, that is, whether it is possible to detect personality traits of the authors of written Modern Greek texts, it is necessary to have a corpus of Modern Greek texts and at the same time to connect each author of these texts to a psychological profile. Due to the lack of such material, the first step was to collect primary textual data from native speakers of Modern Greek. In particular, the corpus that we developed consists of essays of 198 high school students and comprises 250.000 words in total. It is balanced both in size (number of words per student) and in students' demographics (gender and age).

The participating students of three different high schools were asked to write three essays to achieve our goal, which was to collect at least 1,000 words from each student. The task was voluntary, lasted three school years, and the writing was held in the classroom. The experiment was repeated three times at different periods. The authors had to write spontaneously and continuously for 60 minutes an essay. The topics, which were not given in advance, were related to the benefit of art, the role of school in raising environmental awareness, and fighting against child labor. Finally, since the provided texts were handwritten, we had to digitize them by manually typing all of them.

IV. METHODOLOGY

The following section describes the approach used to predict personality types of students.

A. Approach

In the literature, two approaches stand out for an automatic author's personality prediction. In a bottom-up approach, personality labels are predicted from linguistic features that are being extracted from the corpora used using standard NLP document representations (e.g., Bag-of-Words - BoW models, etc.) [12]-[14]. In a top-down approach, instead, specialized dictionaries with custom entries are used to check the potential correlation with personality traits [15]-[17]. Both approaches have advantages, as well as restrictions. Therefore, modern techniques are oriented towards hybrid methods that combine the use of a dictionary with extended document representations trained on machine learning algorithms to exploit the best from both approaches, i.e., speed and precision, respectively. In this study, we followed the bottom-up approach, which among other benefits explained above, is also language-independent.

B. Feature extraction

The features used in our research can be considered as part of a broader feature set that has been characterized as stylometric, i.e., models quantitatively the text's style. The linguistic features that have been used previously as stylometric indices are numerous. They increase continuously and belong to the whole range of linguistic levels. Stylometric features are compact, information-rich signaling linguistic devices. They are correlated with many different textual functions and carry multilevel information related to both the author's identity and his/her metadata. In CPP, stylometric features can unchain the hidden link between linguistic production and its correlation with specific personality types. This is because our personality traits are defining and be defined by our socio-cognitive and psychological conditions. In that sense, aspects of our linguistic behavior reflect these personality traits indirectly and amplify them using identity perceptions.

We processed the corpus with natural language processing tools during the pre-processing phase, i.e., tokenizer, lemmatizer, and POS tagger. The output of the preprocessing phase (matrix of stylometric features) was submitted to the data mining platform Rapidminer [18]. The text preprocessing pipeline was initially applied to the original texts of the students. However, we observed that various language errors scattered across all linguistic levels inserted significant bias in the modeling process and negatively affected the prediction results. Therefore, the essays were corrected manually without loss of information on the morphosyntactic level.

We designed and ran multiple experiments in order to extract and quantify many different subsets of stylometric features from the corpus. We extracted the most frequent character bigrams and trigrams, words bigrams, and trigrams, mean word and sentence length, the occurrence frequency of content and functional words, the most and less frequent words, the occurrence frequency of parts of speech, as well as hapax and dis legomena. These features have been proven effective in the field of authorship attribution [19] and gender identification [20], and we tested them for author personality prediction as well.

C. Classification Algorithms

In this project, the problem of predicting the personality type was treated as a binary classification task among the four dimensions of personality, **E**xtraversion-**I**ntroversion, **S**ensing-i**N**tuition, **T**hinking-**F**eeling and **J**udging-**P**erceiving. To have a valid prediction, the extracted stylometric features matched the texts whose authors clearly belonged to a positive or negative category.

Since personality detection presents a complex classification task, we decided to use several different machine learning algorithms to find the best approach in terms of model performance. We compared nine machine learning methods, i.e., Naive Bayes, Generalized Linear Model, Logistic Regression, Fast Large Margin, Deep Learning, Decision Trees, Random Forest, Gradient Boosted Trees, Support Vector Machines and we ranked them according to their cross-validated accuracy (10-fold). We evaluated the machine learning algorithms in terms of their predictive ability using as training data the students' essays.

 ISBN: 978-1-61208-869-3

Their personality type had been defined before using the appropriate psychometric questionnaire.

V. RESULTS

In this section, we present the results of the procedure that we followed to automatically classify the students' essays based on the personality type defined by the personality questionnaire they filled in. From the nine algorithms that were trained in the textual data, we present the evaluation metrics of the most effective algorithm (Table I) along with the corresponding weights that positively affected the prediction of the personality type.

Regarding the prediction of all personality types of Jung's typology, the algorithm with the best results was Naive Bayes. The accuracy rate revealed a range from 68.8% to 80.7%, with an average of 76.5%. Extraversion type was predicted with 80.7%, the Intuition type with 79.9%, the Feeling with 68.8%, and the Judging type with 75.7% [21]. A more detailed list of evaluation metrics (accuracy, precision, and recall) is reported in Table I.

TABLE I. NAIVE BAYES MODEL PERFORMANCE

Personality Type	**Naive Bayes Classifier**		
	Accuracy	***Precision***	***Recall***
Extraversion	80.7%	80.5%	100%
Intuition	79.9%	81.3%	92.6%
Feeling	68.8%	67.7%	96.7%
Judging	75.7%	76.2%	95.2%

The remaining algorithms that were trained in the corpus produced the following results in terms of classification accuracy: Regarding the Extraversion type, the Generalized Linear Model, Logistic Regression, Fast Large Margin, Deep Learning, Decision Trees, Random Forest and Gradient Boosted Trees algorithms have the same percentage of accuracy being 80.0% and the Support Vector Machine algorithm has 79.0%. The Intuition type was predicted with 75.0% by Gradient Boosted Trees algorithm, with 71.9% by Deep Learning and with 71.7% by Generalized Linear Model and Logistic Regression. For the Feeling type, the Decision Tree algorithm exhibits the second best performance with 63.2%, Random Forest being in third position with 63.1% and the next best result was 63% using Gradient Boosted Trees. The algorithms with the best performance for the Judging type were Support Vector Machine, Fast Large Margin and Deep Learning with calculated accuracies of 71.1%, 71.0% and 70.3% respectively.

The study aimed to classify the essays of the students in personality types by using stylometric indices. Therefore, we had to check whether and which of these features are the most useful and contribute to the prediction accuracy of the algorithm. For this reason, we extracted the weights from the Naive Bayes model that measure the importance of each stylometric feature to the classification decisions of the algorithm for each personality type separately.

For Extraversion (Figure 1), the use of verb types in active voice had a significant impact. In addition, the mean length of the sentence in words of all sentences, the words that occur only twice in one text, the most frequent content words, and finally the personal pronouns complete the list with the five most important stylometric features.

Figure 1. Weights for Extraversion.

Figure 2 depicts the prediction ability of the stylometric features for Intuition used by the algorithm. The word's mean length in characters had the most significant impact. The features that follow are the most frequent trigrams of characters, the hapax legomena, the personal pronouns, the content words, the most frequent word bigrams, the rarest words, the most frequent word trigrams, and all content words.

Figure 2. Weights for Intuition.

The stylometric features that affected the result of the classification of the essays in terms of Feeling are the verbs, the adjectives, the most frequent content words, the personal and the possessive pronouns, the nouns, and the adverbs (Figure 3).

Finally, in Figure 4, the eight stylometric features that contributed to the prediction of the Judging type were in descending order: The most common word trigrams, the most common word bigrams, the mean length of the sentence in words, the most common character bigrams and the most common character trigrams with the same percentage, the personal and possessive pronouns, the articles, and the mean length of the word in characters.

 ISBN: 978-1-61208-869-3

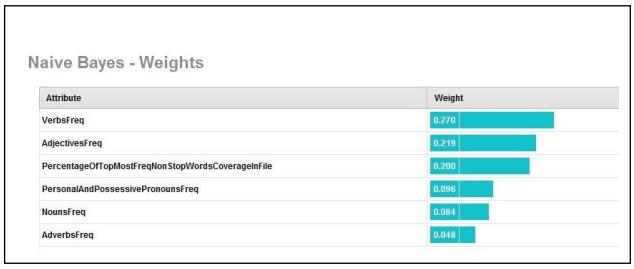

Figure 3. Weights for Feeling.

Figure 4. Weights for Judging.

The most important features extracted from the model vary considerably for each personality type. Therefore, we can infer that each type is based on different combination of linguistic features and these subsets are different between the different personality types.

It also becomes clear that the predictive accuracy of the proposed classification model is high compared to the existing literature on the field of personality prediction. Regarding Jung's Typology Test, we got an average accuracy of 76.5%, compared to the 68.62% reported for Dutch [6]. The other studies mentioned [7]-[11] implemented machine learning techniques in textual data that were retrieved from social media. Therefore, their results involve research with textual data from adults written under different circumstances and in a different language.

VI. Conclusion and Future Work

To summarize, in this paper, we presented the results of our research in the field of personality prediction. We applied CPP for the first time in texts written by high-school students, making our dataset unique. Our results confirmed our initial research hypothesis that stylometric features could be used as reliable prediction indices for the author's psychological profile. Our findings further support the latent link of personality traits with a wide array of linguistic behaviour aspects. Different personality types correlate with different stylometric features that belong to different linguistic levels. Therefore, the personality prediction through text demands a highly dynamic feature set to capture the widest possible spectrum of linguistic structures.

To this direction, future research will employ experimentation with more linguistic features. Furthermore, we plan to localize in Modern Greek well-known psychometric lexicons (e.g., LIWC) and use them complementing our feature sets.

References

[1] Y. Neuman, D. Assaf, Y. Cohen, and L. J. Knoll, "Profiling school shooters: Automatic text-based analysis," Front. Psychiatry, vol.6, p.86, 2015.

[2] J.W. Pennebaker, R.L. Boyd, K. Jordan, and K. Blackburn, "The Development and Psychometric Properties of LIWC2015". Austin, TX: University of Texas at Austin. (www.LIWC.net), 2015. [retrieved: June, 2021]

[3] Jr. P. T. Costa and R. R. McCrae, "NEO-PI-R: Professional Manual," Odessa, Fla.: Psychological Assessment Resources, 1993.

[4] C. G. Jung, "Psychological Types," Princeton, New Jersey: Princeton University Press, 1971.

[5] I. Briggs Myers and P. B. Myers, "Gifts Differing: Understanding Personality Type," Mountain View, CA: Davies-Black Publishing, 1980.

[6] K. Luyckx and W. Daelemans, "Personae: A corpus for Author and Personality Prediction from Text" The Sixth International Language Resources and Evaluation Conference (LREC 2008), 28-30 May 2008, pp. 2981-2987.

[7] D. Brinks and H. White, "Detection of Myers - Briggs Type Indicator via Text based Computer-mediated Communication," CS 229 Machine Learning Projects, Stanford, 2012.

[8] B. Plank and D. Hovy, "Personality Traits on Twitter-or-how to get 1,500 Personality Tests in a Week" The Sixth Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, Association for Computational Linguistics, 2015, pp. 92-98.

[9] B.Verhoeven, W. Daelemans, and B. Plank, "Creating TwiSty: Corpus Development and Statistics," Computational Linguistics and Psycholinguistics Research Center CLiPS Technical Report Series, University of Antwerp, Belgium, CTRS-006, 2016.

[10] L. C. Lukito, A. Erwin, J. Purnama, and W. Danoekoesoemo, "Social Media User Personality Classification using Computational Linguistic" The Eighth International Conference on Information Technology and Electrical Engineering, Oct. 2016, pp. 1-6.

[11] K. Yamada, R. Sasano, and K. Takeda, "Incorporating Textual Information on User Behavior for Personality Prediction" The 57th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop, Jul.-Aug. 2019, pp. 177-182.

[12] J. Oberlander and S. Nowson, "Whose Thumb is it anyway? Classifying Author Personality from Weblog Text" The 44th Annual Meeting of the Association for Computational Linguistics ACL, Jul. 2006, pp. 627-634.

[13] F. Iacobelli, A. J. Gill, S. Nowson, and J. Oberlander, "Large Scale Personality Classification of Bloggers" The Fourth International Conference on Affective Computing and Intelligent Interaction, 2011, Heidelberg: Springer-Verlag, pp. 568-577.

[14] Y. Bachrach, M. Kosinski, T. Graepel, P. Kohli, and D. Stillwell, "Personality and Patterns of Facebook Usage" The Fourth Annual ACM Web Science Conference, 2012, pp. 36-45.

 ISBN: 978-1-61208-869-3

[15] S. Argamon, S. Dhawle, M. Koppel, and J. W. Pennebaker, "Lexical Predictors of Personality Type" The Joint Annual Meeting of the Interface and the Classification Society of North America, 2005, pp. 1-16.

[16] F. Mairesse, M. A. Walker, M. R. Mehl, and R. K. Moore, "Using Linguistic Cues for the Automatic Recognition of Personality in Conversation and Text," Journal of Artificial Intelligence Research, vol. 30, pp. 457-500, 2007.

[17] J. Golbeck, C. Robles, and K. Turner, "Predicting Personality with Social Media" The 2011 Annual Conference Extended Abstracts on Human Factors in Computing Systems, pp. 253-262.

[18] I. Mierswa and R. Klinkenberg. Rapidminer Studio (9.1) [Data science, machine learning, predictive analytics]. (https://rapidminer.com). [retrieved: June, 2021].

[19] G. K. Mikros and G. Markopoulos, "Using Multiword Sequences as Features in Authorship Attribution: Experiments based on Greek Blog Texts," In A. Christofidou (Ed.), Aspects of Corpus Linguistics: Principles, applications and challenges Vol. 14, pp. 56-67, 2017. Athens: Academy of Athens: Research Center for Scientific Terms and Neologisms.

[20] G. K. Mikros, "Authorship Attribution and Gender Identification in Greek Blogs," In I. Obradović, E. Kelih & R. Köhler (Eds.), Selected papers of the VIIIth International Conference on Quantitative Linguistics (QUALICO), Apr. 2013, Belgrade: Academic Mind, pp. 21-32.

[21] S. Gagiatsou, "Automatic author profiling based on natural language processing techniques", Ph.D. Thesis, National and Kapodistrian University of Athens: Greece, 2021.

 ISBN: 978-1-61208-869-3

Blockchain for Smart Grid Flexibility

Handling Settlements between the Aggregator and Prosumers

Lasse Berntzen
School of Business
University of South-Eastern Norway
Horten, Norway
email: lasse.berntzen@usn.no

Qian Meng
School of Business
University of South-Eastern Norway
Horten, Norway
email: qian.meng@usn.no

Boban Vesin
School of Business
University of South-Eastern Norway
Horten, Norway
email: boban.vesin@usn.no

Marius Rohde Johannessen
School of Business
University of South-Eastern Norway
Horten, Norway
email: marius.johannessen@usn.no

Thomas Brekke
School of Business
University of South-Eastern Norway
Horten, Norway
email: thomas.brekke@usn.no

Inessa Laur
School of Business
University of South-Eastern Norway
Horten, Norway
email: inessa.laur@usn.no

***Abstract*—This paper shows how the Ethereum blockchain can register settlements between an aggregator and prosumers in a smart grid. By providing flexible use of electricity to the aggregator, customers get rewarded. The flexibility is valuable for the aggregator since the power infrastructure may be used more efficiently. Blockchain is an exciting technology for handling settlements which, however, also has some clear limitations. For example, the cost per transaction on the public Ethereum blockchain is too high compared to the value of the actual transactions. A private blockchain is an alternative but removes some of the original benefits of using the public blockchain. The paper concludes that blockchain is a promising technology, and a private blockchain is more suitable for transactions containing minimal amounts.**

Keywords-smart grid; blockchain; Ethereum; smart contract; aggregator; flexibility; Smart-MLA.

I. INTRODUCTION

Smart grids [1] are electric power grids supported by electronics that keep track of power consumption and production. Typically, smart meters [2] keep track of energy flow between producers, consumers, and prosumers. Smart meters normally connect to a service maintained by the Distribution System Operator (DSO).

Prosumers refers to the combination of producers and consumers, with production coming from energy sources like solar panels. On sunny days, the energy production of a household or a building may exceed the consumption. Then the prosumer may sell excess energy to the grid. At other times, the household gets its electricity from the grid, as shown in Figure 1.

The price of energy varies throughout the day. The energy price is calculated based on demand and supply forecasts, and the pricing intervals typically differ from one hour down to ten minutes.

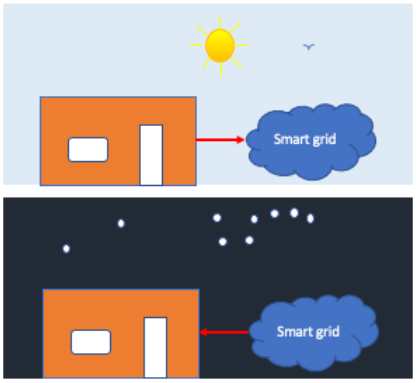

Figure 1. Prosumers

Flexibility occurs when a household or building can delay the use of electricity to a timeslot when the price is lower. A typical example is the charging of electric vehicles. Charging requires a rather large consumption over a relatively short period. If the customer chooses the optimal time, charging will be less expensive.

The next step is to transfer the flexibility to an aggregator. In this case, the customer defines some constraints, e.g., the electric car should be fully charged at 7 am. The aggregator then selects the optimal time slots when the actual charging takes place.

The flexibility is essential for grid management since it can reduce the chances of overloading the grid and delay investments in upgraded electric power infrastructure.

The customer is rewarded for giving up control, either by favorable pricing or a discount on the electricity bill.

Blockchain technology has the potential to have a significant impact on the energy sector. Numerous use cases have been proposed, including wholesale and retail energy trading [3][4].

 ISBN: 978-1-61208-869-3

The purpose of this paper is to present the Smart Multi-Layer Aggregator (Smart-MLA) project and how blockchain technology can handle settlements between the aggregator and its customers (prosumers).

Section II discusses the flexibility project, followed by a section explaining Blockchain technology. Section IV presents some related work, followed by a section showing our approach for the project. Finally, Section VI concludes and provides some input on future work.

II. THE SMART MULTI-LAYER AGGREGATOR

The Smart-MLA project [5] is an ERA-Net Smartgrid Plus research project with academic and industrial partners from Denmark, Norway, Romania, Sweden, and Turkey. The aim is to develop and demonstrate a cloud-based multi-layer aggregator solution to facilitate optimum demand response and grid flexibility for energy systems to utilize up to 100% renewable energy. The project implements three layers of flexibility aggregation, as shown in Figure 2.

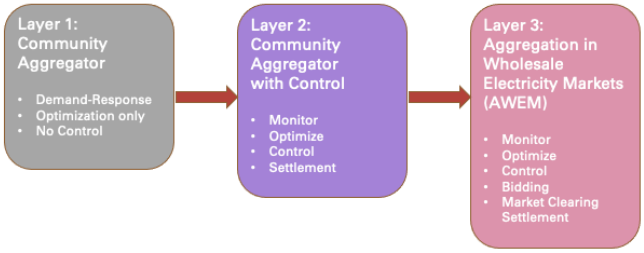

Figure 2. The Smart-MLA layers

On the lowest layer (shown in Figure 3), flexibility occurs within a household or a building. The main objective of this layer is to improve the awareness of the prosumers on the merits of flexibility at the prosumer level. Here, the customer sets constraints and the preferred scheduling for flexible appliances. Then, the proposed algorithm [6] optimizes the operation of the flexible appliances and comes up with the difference between the cost of using user-preferred and optimal schedules. The solution considers both production capacity (solar panels) and storage capacity (batteries) to obtain the best possible result.

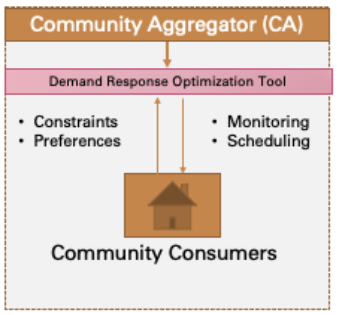

Figure 3. Smart-MLA layer 1

On the second layer (shown in Figure 4), the customer transfers the control of charging electric vehicles, heat pump, and other appliances to the aggregator. When the aggregator controls many households/buildings connected to the grid, the aggregator may optimize the power consumption for the whole smart grid.

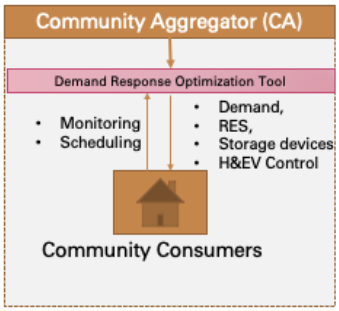

Figure 4. Smart-MLA layer 2

This layer includes settlements between the aggregator and the prosumers. In this layer, the grid is used as an on-demand delivery service that may also buy back excess energy and reward the prosumer for transferring flexibility. The settlements are between the prosumers and the aggregator, as shown in Figure 5. The settlements are registered on the blockchain.

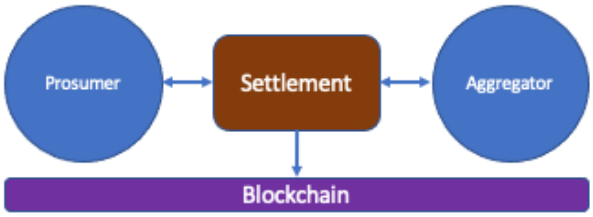

Figure 5. Settlements on layer 2

The flexibility aggregated from the prosumers will be traded in the local flexibility market in coordination with the Distribution System Operator (DSO) on the upper layer. The aggregator uses an iterative process based on schedules, prices, and bids for flexibility, as shown in Figure 6.

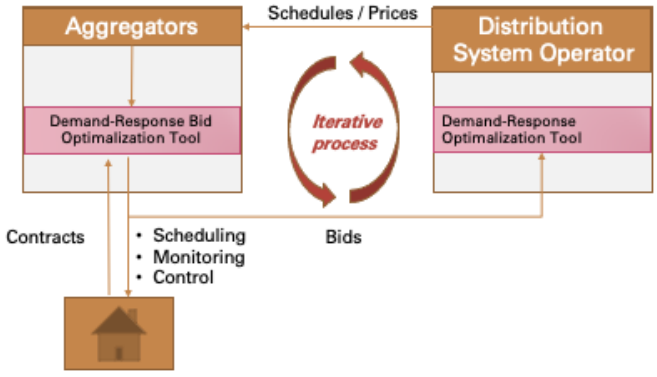

Figure 6. Smart-MLA layer 3

Further details on the three-layer aggregator flexibility can be found in a separate paper [7].

 ISBN: 978-1-61208-869-3

The settlements are made between the aggregator and the prosumers, and the aggregator and the DSO, as illustrated in Figure 7.

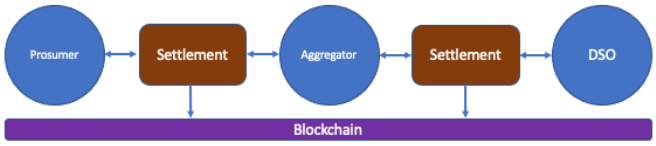

Figure 7. Settlements on layer 3

Both types of settlements are registered on the blockchain. The settlements include the amount of energy, time, period, and the two entities involved in the transaction.

III. BLOCKCHAIN BASICS

Blockchain technology was initially developed to handle cryptocurrency safely and transparently [8]. Bitcoin was the first cryptocurrency using blockchain as a platform. The blockchain is a decentralized, immutable ledger. The ledger is implemented as blocks of data chained together [9], as shown in Figure 8. Cryptographic techniques ensure that it is impossible to change the block's content when put on the blockchain [9]. The first block is called the genesis block.

Figure 8. The blockchain

Each block contains a link to the previous block and also a hash value of the previous block. Each new block is encrypted (including the hash value of the previous block). A block cannot be altered after it is added to the blockchain. In that case, the whole blockchain becomes invalid [9]. Figure 9 shows the layout of the blocks and the block header.

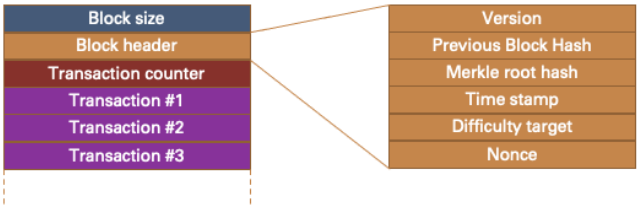

Figure 9. The blockchain header

A "Merkle" tree [9] secures the integrity of the transactions stored in the block. First, a hash value is calculated for each transaction. Hash values are then calculated for each pair of transactions, and the process is repeated until the root of the tree appears. Finally, the Merkle root hash is stored in the block header. The construction of the Merkle tree is illustrated in Figure 10.

The block is added after validation. The validation is done by so-called miners who solve a mathematical problem and get rewarded (with cryptocurrency) for solving the problem [9].

A new block is distributed to many nodes in the network. A majority consensus protocol secures the integrity of the blockchain. Therefore, an attack needs to succeed with a majority of the nodes in the network, which is nearly impossible.

If one or more nodes fail, the data is still obtainable from the other nodes. Thus, if there should be an inconsistency, the majority will win.

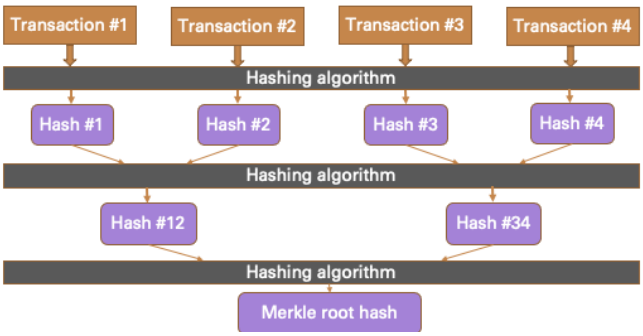

Figure 10. The Merkle tree

A blockchain may be public or private. Everyone has access to the public blockchain and can examine the transactions that have been made. The blockchain is semi-anonymous; the user is identified with a binary address. A private blockchain may require permission to examine the transactions.

IV. BLOCKCHAIN IN SMART GRIDS

In recent years, the centralized power grids have experienced significant changes and challenges as follows [10]:

- The production is unpredictable and variable due to an increase in renewable energy sources.
- The power transmission and distribution become more controllable and fault-tolerant due to network digitization.
- Prosumers (entities that both produce and consume energy) have the active dual role of both producer and consumer.
- Loads become more interactive and dynamic.

Blockchain as a decentralized and distributed technology secures suitable applications in the transition to smart grid infrastructure. The utilization of the blockchain in smart grids could offer various advantages to the electrical power system with increased security, improved data privacy, data transparency and immutability, removal of third-party control and trust, ubiquitous solution, and greater data accessibility. A lot of research has been done about blockchain technology applied to the power sector [11]. The integration of blockchain and renewable energy sources, energy storage devices, and electric vehicles into the electrical grid has been a broad research area [12]. Blockchain technology brought agreeable advantages and

 ISBN: 978-1-61208-869-3

created much interest in applying this technology in smart grids [13]. Blockchain applications in the smart grid could be divided into three parts of the smart grids as follows [14]:

- Power generation: Blockchains and smart contracts can allow a generating unit to directly trade with a consumer or a retail energy supplier via autonomous trading agents, cutting out the middleman [15]. Blockchains could also provide innovative trading platforms for integrating small-scale renewables, distributed generation, and flexibility services [16].
- Power transmission and distribution: Blockchains enable independent power grid nodes, without prior trust in each other, to reach an agreement on the optimal power flow solution through simulation experiments on the 39-bus New England transmission system [17]. Blockchains also guarantee the origin of the energy from the voltage distribution across the network [18].
- Power consumption: Emerging blockchain technologies ensure the seamless and secure implementation of a decentralized demand-side management approach following optimized demand profiles, including battery devices, electric vehicles' charging [19].

Among the examined research, a focused topic is about electric vehicles' connection to the smart grids. The random charging of these vehicles may give a heavy burden to the power grid. To solve this problem, blockchain technology is introduced. An electric vehicle integration scheme is proposed based on blockchain technology for reducing fluctuations in the grid [20]. It is suggested that a platform that integrates electric vehicles and charging stations enables a negotiation between them over blockchain [21]. It explains using blockchain technology to make electric cars find a close-by charging station [22]. To improve voltage stability of the grid while minimizing charging costs for electric vehicle users, an adaptive blockchain-based electric vehicle participation scheme is proposed to schedule electric vehicles charging/discharging demand to flatten the impact of consuming/injecting excess amount of energy at the level of the transformer substation [23]. The preservation of vehicle owners' privacy is also considered by a smart contract that allows vehicles to signal their demand and receive offers by charging stations using predefined regions without revealing their exact location [24]. The integration of electric vehicles with blockchain technology may secure the best location and price for electric vehicle users and ensure security and privacy [25]. It proposes a secure and efficient vehicle to grid energy trading framework based on blockchain and edge computing [26].

The following section will show how blockchain is used in the Smart-MLA project to handle settlements between the aggregator and the prosumers.

V. BLOCKCHAIN IN THE SMART-MLA PROJECT

One of the aims of the Smart-MLA project is to demonstrate the use of blockchain technology for settlements. The reason for including blockchain was an increasing interest in using blockchain for applications other than the pure cryptocurrency transfer. In Smart-MLA, the communication is between the aggregator and prosumers connected to the smart grid and between the aggregator and the DSO.

Ethereum [27] is a further development of blockchain technology introducing smart contracts. Smart contracts [28] are self-executing programs stored on the Ethereum blockchain that run when some predetermined conditions are met. It provides the possibility to do more than transfering amounts between actors. For example, smart contracts can be used for bidding processes between prosumers and the aggregator and between the aggregator and the DSO.

Smart contracts are written in the Solidity programming language [29], an integrated part of the Ethereum blockchain technology.

The demonstration of blockchain technology for settlements has been done as a set of activities:

- Setting up a local blockchain using Ganache and Remix
- Writing to and reading from the blockchain using web3
- Setting up a test blockchain using the Ropsten test network
- Using Nethereum to build applications using the .NET framework
- Making a test API for settlements

A. Local Blockchain

The local blockchain was implemented by using Ganache. Ganache provides a free Ethereum test blockchain network and is part of the Truffle blockchain development framework [30]. Figure 11 shows the Ganache graphical user interface (GUI) with ten accounts set up on the local blockchain.

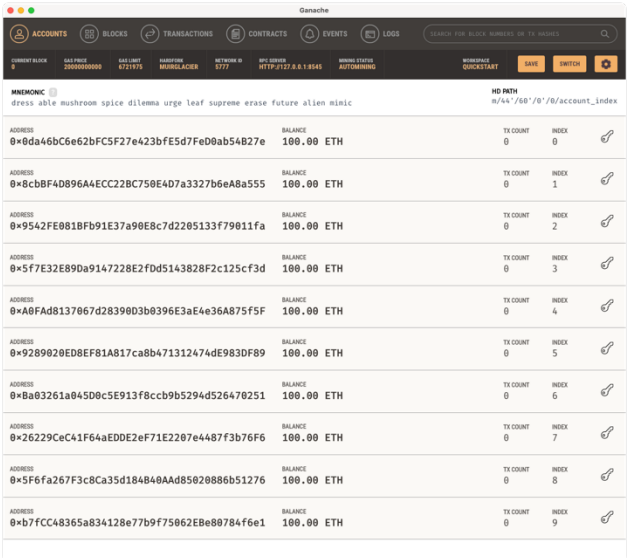

Figure 11. Ganache GUI

Ganache can be accessed from the Remix integrated development environment (IDE) [31]. Remix facilitates the writing, compilation, and deployment of smart contracts written in the Solidity language. The combination of Remix and Ganache was used to test the smart contracts. The smart contract validates senders and receivers, uses tariffs to compute correct amounts to be transferred, and checks if the sender has the necessary amount available.

B. Using web3

The web3 [32] library provides access to the Ethereum blockchain from programs. The library may be used with Ganache, the Ethereum test networks (e.g., Ropsten), or the public Ethereum network.

C. Using the Ropsten test network

Ropsten is one of four Ethereum test networks listed on the Ethereum website [33]. Like the public Ethereum network, Ropsten requires payments for making transactions. However, coins for payment are free and requested by sending a Twitter message to the network operator.

D. Using Nethereum

Nethereum [34] is a .NET library that makes it possible for .NET programs to write to and read from the Ethereum blockchain. User interfaces, as well as application program interfaces, may then be written in .NET code.

E. Test API

Finally, a test API was made to demonstrate how settlements may be done in practice. The smart meter or an agent talking to the smart meter writes a record to the blockchain about a settlement. The blockchain can also be accessed from the billing service and other services doing statistics and monitoring. Figure 12 shows the architecture of the settlement service.

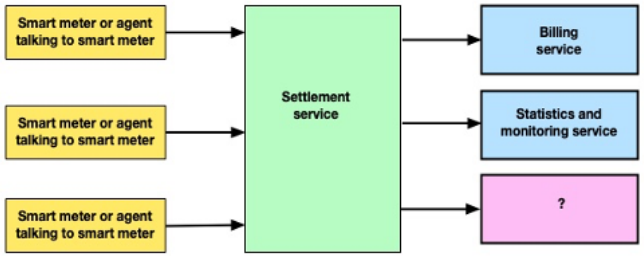

Figure 12. The settlement service

As shown earlier, the settlement records contain the amount of energy, time, period, and the two entities involved in the transaction. The smart contract checks the validity of the sender and receiver; and if the sender has the necessary amount available. The smart contract may also create a competitive environment based on bidding between the aggregator and prosumers and between the aggregator and the DSO.

VI. Conclusions and Future Work

The Smart-MLA project has demonstrated the use of blockchain for a specific smart grid application – registering settlements between an aggregator and the prosumers connected to the smart grid. The demonstration showed some disadvantages of using blockchain for this purpose.

To register a transaction on the real Ethereum blockchain requires a transaction fee. The current transaction fee on the Ethereum network is around USD 4.00 (June 2021). Therefore, if blockchain is used to register car owner transfers, the transaction fee would be negligible compared to the transaction itself.

However, the settlements between the aggregator and prosumers will be tiny amounts. Even if aggregated for each hour or even each day, the transaction fee would create a considerable overhead for the transactions.

Therefore, the public Ethereum blockchain should not be used for transactions involving minimal payments. Also, aggregation does not solve the problem since the whole idea of using blockchain was to achieve full transparency of all transactions.

The alternative is to set up a private blockchain. In that case, the fees for registering will not have a real value. But the infrastructure itself, mainly servers, will have a price tag. Blockchain is supposed to have a large number of copies on a distributed network. For a private blockchain, the number of nodes will be limited, and the blockchain will be more vulnerable to security attacks.

A traditional system of record (e.g., a relational database) could handle settlements in a much more efficient way. However, the problem with a database is the centralization of physical control combined with the ability to delete or modify records.

The blockchain excels in being immutable and transparent.

The Smart-MLA project strives to demonstrate how aggregators may benefit from customer flexibility. The smart grid may be enhanced by letting prosumers trade with each other through a stock-market exchange. Prosumers with storage capacity may compete to buy energy from other prosumers for future sales. This will create an internal market within the smart grid. The aggregator will then have to compete for excess energy produced by prosumers, optimizing the revenues for the prosumers. Today, the aggregator decides the buying price in a monopolistic way. An internal market within the smart grid would remove the monopoly of the aggregator and make a sounder competitive environment.

Acknowledgment

This work was supported under the ERA-Net Smart Grids Plus scheme Grant number 89029 and funded through the Research Council of Norway grant number 295750 with the project title "Multi-layer aggregator solutions to facilitate optimum demand response and grid flexibility."

 ISBN: 978-1-61208-869-3

REFERENCES

[1] J. C. Stephens, E. J. Wilson, and T. R. Peterson, "Smart Grid (R)evolution – Electric Power Struggles," Cambridge University Press, 2015.

[2] C. Meinecke, "Potentiale und Grenzen von Smart Metering," Springer VS, 2016.

[3] M. Gough, R. Castro, S.F. Santos, M. Shafie-khah, and J. P. S. Catalão, "A panorama of applications of blockchain technology to energy," in Blockchain-based Smart Grids, M. Shafie-khah, Ed. Elsevier/Academic Press, pp. 5-41, 2020.

[4] C. Liu, X. Zhang, K.K. Chai, J. Loo, and Y. Chen, "A survey on blockchain-enabled smart grids: Advances, applications, and challenges," IET Smart Cities no. 3, pp. 56-78. 2021.

[5] Smart-MLA Consortium. Multi-layer aggregator solutions to facilitate optimum demand response and grid flexibility (Smart-MLA). Project web site. [Online] Available from https://smart-mla.stimasoft.com 2021.06.01

[6] S. Teimourzadeh et al., "Enlightening customers on merits of demand-side load control: A simple-but-efficient-platform," IEEE Access, vol. 8, pp. 193238-193247, 2020.

[7] X. Jin, S. Teimourzadeh, O. B. Tor, and Q. Wu, "Three-Layer Aggregator Solutions to Facilitate Distribution System Flexibility," in Flexibility in Electric Power Distribution Networks, H. H. Alhelou, E. Heydarian-Forushani, and P. Siano, Eds. CRC Press, ch. 8, 2021.

[8] S. Nakamoto, "Bitcoin: A Peer-to-Peer Electronic Cash System," 2008. [Online] https://bitcoin.org/bitcoin.pdf 2021-05-01

[9] D. Hellwig, G. Karlic, and A. Huchzermeier, "Build Your Own Blockchain - A Practical Guide to Distributed Ledger Technology," Springer Nature, 2020.

[10] M. Foti and M. Vavalis, "What blockchain can do for power grids? " Blockchain: Research and Applications, vol. 2, pp. 1-14, 2021.

[11] A. S. Musleh, G. Yao, and S. M. Muyeen, "Blockchain Applications in Smart Grid – Review and Frameworks," in IEEE Access, vol. 7, pp. 86746-86757, 2019.

[12] H. Farhangi, "The path of the smart grid," IEEE Power Energy Magazine., vol. 8, no. 1, pp. 18-28, 2010.

[13] Z. Dong, F. Luo, and G. Liang, "Blockchain: A secure decentralized trusted cyber infrastructure solution for future energy systems," Journal of Modern Power Systems and Clean Energy, vol. 6, no. 5, pp. 958-967, 2018.

[14] W. Su and A. Q. Huang, "The Energy Internet," Sawston, U.K.:Woodhead Publishing, 2018.

[15] V. Grewal-Carr and S. Marshall, "Blockchain enigma paradox opportunity," [Online] Available from https://www2.deloitte.com/content/dam/Deloitte/uk/Documents/Innovation/deloitte-uk-blockchain-full-report.pdf 2021.05.01

[16] M. Andoni et al., "Blockchain technology in the energy sector: A systematic review of challenges and opportunities," Renewable and Sustainable Energy Reviews, vol. 100, pp. 143-174, 2019.

[17] M. Foti, C. Mavromatis, and M. Vavalis, "Decentralized blockchain-based consensus for Optimal Power Flow solutions," Applied Energy, vol. 283, 2021.

[18] M. L. Di Silvestre et al., "Transparency in transactive energy at distribution level," 2017 AEIT International Annual Conference, pp. 1-5, 2017.

[19] S. Noor, W. Yang, M. Guo, K. H. van Dam, and X. Wang, "Energy Demand Side Management within micro-grid networks enhanced by blockchain," Applied Energy, vol. 228, pp. 1385-1398, 2018.

[20] Z. Su et al., "A Secure Charging Scheme for Electric Vehicles with Smart Communities in Energy Blockchain," IEEE Internet of Things Journal, vol. 6, no. 3, pp. 4601-4613, 2019.

[21] M. Pustišek, A. Kos, and U. Sedlar, "Blockchain based autonomous selection of electric vehicle charging stations," Proc. 2016 International Conference on Identification, Information and Knowledge in the Internet of Things (IIKI); pp. 217-222, 2016.

[22] J. Kang et al., "Enabling localized peer-to-peer electricity trading among plug-in hybrid electric vehicles using consortium blockchains," IEEE Transactions on Industrial Informatics., vol. 13, no. 6, pp. 3154-3164, 2017.

[23] C. Liu, K. K. Chai, X. Zhang, E. T. Lau, and Y. Chen, "Adaptive blockchain-based electric vehicle participation scheme in smart grid platform," IEEE Access, vol. 6, pp. 25657-25665, 2018.

[24] F. Knirsch, A. Unterweger, and D. Engel, "Privacy-preserving blockchain-based electric vehicle charging with dynamic tariff decisions," Computer Science – Research and Development. 33, pp. 71-79, 2018.

[25] Y. Hou et al., "A resolution of sharing private charging piles based on smart contract," Proc. 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), pp. 3004-3008, 2017.

[26] Z.Y. Zhou, L. Tan, and G. Xu, "Blockchain and edge computing based vehicle-to-grid energy trading in energy internet," Proc. 2nd IEEE Conference on Energy Internet and Energy System Integration (EI2), pp. 1-5. 2018.

[27] ethereum.org [Online] Available from: https://ethereum.org 2021.06.01

[28] D. Rossbach, "Smart Contracts in Blockchain," Amazon, 2018.

[29] M. Mukhopadhyay, "Ethereum Smart Contract Development," Packt Publishing, 2018.

[30] N. Bhaskar, "Truffle Quick Start Guide," Packt Publishing, 2018.

[31] ethereum.org [Online] Available from https://remix.ethereum.org 2021.06.01

[32] Web3js [Online] Available from https://web3js.readthedocs.io/en/v1.3.4/ 2021.06.01

[33] ethereum.org [Online] Available from https://ethereum.org/en/developers/docs/networks/ 2021.06.01

[34] nethereum.com [Online] Available from https://nethereum.com 2021.06.01

 ISBN: 978-1-61208-869-3

Vehicle to Grid and Crisis Management

Potential of V2G for smart city power grids in Norway

Marius Rohde Johannessen
School of Business
University of South-Eastern Norway
Horten, Norway
email: marius.johannessen@usn.no

Lasse Berntzen
School of Business
University of South-Eastern Norway
Horten, Norway
email: lasse.berntzen@usn.no

Boban Vesin
School of Business
University of South-Eastern Norway
Horten, Norway
email: boban.vesin@usn.no

Qian Meng
School of Business
University of South-Eastern Norway
Horten, Norway
email: qian.meng@usn.no

Thomas Brekke
School of Business
University of South-Eastern Norway
Horten, Norway
email: thomas.brekke@usn.no

Inessa Laur
School of Business
University of South-Eastern Norway
Horten, Norway
email: inessa.laur@usn.no

Abstract— **Researchers and practitioners alike have discussed Vehicle-to-Grid (V2G) technology for quite some time, and we have extensive coverage of the technology needed, as well as opportunities, challenges, algorithms, and business models. However, few studies have examined V2G from a crisis management perspective. This paper presents a review of the current V2G literature. It gives an overview of the possibilities for using electric cars as a backup power supply in case of emergencies, such as citywide power outages. We calculate the potential power available in a typical Norwegian mid-sized city and examine to what extent this can be part of crisis management in case of a massive power outage. We then review existing business model literature to analyze possible incentives for people to make their cars available for V2G. Finally, we conclude by pointing out several technical and social issues that need to be addressed for V2G to become a viable option and suggest vehicle-to-home (V2H) as the most likely scenario in the short to mid-term.**

Keywords-vehicle-to-grid; V2G; smart city; smart grid; electric vehicles.

I. INTRODUCTION

Vehicle-to-Grid (V2G) technology is the ability to reverse the flow of energy between Electric Vehicles (EVs) and the home/power grid so that EV batteries function as a backup power source for peak hours or power outages. V2G can also help balance renewable energy sources by feeding electricity to the grid when winds are calm or the sun is not shining. The charging interface initiative (Charin) estimates that V2G technologies are widely available around 2025 [1].

Many technical papers are describing V2G algorithms for optimizing grid and charging efficiency [2], optimal placement and planning of charging and charging locations taking supply challenges and demand into consideration [3], or cost-optimization for end-users [4]. There are even examples of modeling V2G capacity [5]. However, few papers combine this with a crisis management perspective. Thus, this paper focuses on the potential of vehicle-to-grid technology in the context of crisis management. As Norway has been the leading country in adopting electric vehicles, we use Norway as our example. We examine the potential of V2G to act as a backup power source for households in case of emergencies and present a scenario for potentially available energy in 2025 and 2030, based on the current EV fleet and predictions of future uptake. We have calculated the potential for the country as a whole and for a typical mid-sized Norwegian city of 80.000 inhabitants. However, few papers combine this with a crisis management perspective. Crisis management refers to handling a crisis, which in our case is a potential failure in the power grid and the use of energy stored in EV batteries as part of the solution.

The reason for only examining households is that critical functions such as hospitals already have backup systems such as generators in place. For these functions, the varying energy available from EV batteries would probably not meet the strict criteria for reliable backup (but could perhaps be of interest as a last resort if everything else fails).

The rest of the paper is structured as follows: Section II presents a literature review of V2G, including technical requirements, available algorithms, and the challenges,

 ISBN: 978-1-61208-869-3

opportunities, and willingness of users to adopt the technology. Section III presents our research approach. Section IV presents our findings; Available energy on the national and mid-sized city level as a best-case scenario for summer and winter, and a more realistic scenario where assume less available energy. Finally, in Section V, we present our conclusions and possibilities for future research.

II. LITERATURE REVIEW

This section presents related research on the vehicle to grid definition and background, business case, challenges, opportunities, and crisis management.

A. Vehicle to grid definition and background

The power grid has little storage capability and needs to be carefully managed to handle fluctuating customer demand. Storage capacity is costly, except perhaps for hydropower, where energy is stored in dams. According to Kempton and Tomic, electric vehicles are already there, and the number is increasing. On average, cars are stationary 96 % of the time, making their stored energy potentially available for the power grid. The basic definition of V2G is simply the ability of vehicles to provide power to the grid when stationary [6].

There are different ways of connecting vehicles and the power grid. This paper is concerned with bi-directional V2G, where power can flow both ways, and EV batteries can aid in grid balancing and backup. This requires aggregator services, communication between DSO/TSO and homes, as well as smart meters and equipment for two-way transfer of energy between home and grid [7]. Other definitions include unidirectional V2G, where charging is determined based on variables such as available energy and the current prices of electricity. Unidirectional V2G requires less infrastructure and has considerable potential for overall energy-saving and load-balancing of the grid but does not allow for using batteries as a backup energy source [2]. Literature has also examined hybrid solutions, such as Vehicle-to-Home (V2H), where the EV battery is used as a backup power source for individual homes, and Vehicle-to-Vehicle (V2V), where cars can charge each others' batteries [8].

B. Business case, challenges, opportunities

The overall business case for all variants of V2G is the potential for significant cost savings due to energy saving and load balancing, which in turn requires less investment in peak capacity [8]. In Norway, for example, the challenge is not (yet) the availability of electricity but the fact that peak demand sometimes exceeds the grid's capacity to deliver.

As electricity moves to renewable sources such as wind and solar, the need for backup and balancing increases. Winds fluctuate, and the sun doesn't always shine when the need for electricity is greatest. So one scenario for V2G is to act as a backup to ensure the grid is balanced and operational, replacing polluting alternatives such as coal- or gas-driven power plants. Calculations indicate that V2G could stabilize half the US electricity grid with a low base capacity in the form of fixed batteries and 8-38% of the EV fleet providing reserves [6]. Høj, Juul, and Lindegard [9] claim that V2G technology offers the potential for new and profitable business models if they manage to balance and integrate intermittent renewable energy into the grid, reduction of peak load, charging optimization, and regulation of participating capacity. They, therefore, call for more research into the creation of such business models.

Almenning, Bjarghol, and Farahmand [10] have responded and examined the impact of grid tariffs on peak power demand. They found an 11% decrease in peak power use in a test neighborhood where peer-to-peer energy trading in local markets allowed for local energy exchange from solar and batteries. Economic incentives are also important for the large-scale acceptance of peer-to-peer "prosumerism", including willingness to participate in a V2G scheme [11].

In recent years, several studies have modeled the potential economic costs and benefits of V2G. Gough et al. applied a monte Carlo-based data-driven analysis of V2G in UK commercial buildings, where the car park was equipped with V2G capabilities. They found that the most significant potential was wholesale market trading involving peak and off-peak tariffs, with a potential income of around £8400 per vehicle over a 10-year investment period. Vehicle-to-home peak reduction generated less value. Short-term operating reserve (providing extra power at times of high demand) was not viable because battery degradation cost exceeds the savings. This factor is likely to change as battery prices drop [12]. Similarly, Ahmadian et al. [13] found battery degradation costs to exceed the potential income unless set up in a system for balancing wind power. Unidirectional V2G (smart charging) was found to be economical and with no impact on battery degradation.

Li and colleagues [14] found that battery costs might already have dropped enough for a short-term operating reserve to be profitable. They conducted a case study in Shanghai and found the total net profit of V2G to be positive, but with higher profit for power plants than EV users and negative profit for grid companies. Finally, Berglund et al. [15] found that effective use of battery storage could shave peak energy costs by as much as 13,9%, enough to cover the costs of battery storage facilities. This makes it even more likely that V2G could become profitable even when considering increased battery degradation.

There are also some challenges to V2G: A survey conducted in the Nordic countries regarding EV adoption and willingness to be included in V2G capacities showed that only two of the four countries (Norway and Finland) were interested in V2G. The study concludes that the public knows little of the potential of V2G and calls for greater education and awareness-building regarding V2G to increase EV drivers' willingness to participate. However, monetary incentives seem to be popular if people are aware of the potential. Sæle [16] conducted a survey to examine the willingness of Norwegians to change their electricity user habits and found that up to 64% was willing to either allow remote control of appliances, water heating, etc. or to

 ISBN: 978-1-61208-869-3

contribute with a manual response if they save 200 Euros or more in a year.

Another challenge, or rather a criticism, is that many studies of V2G and electricity efficiency take the idea of the rational human for granted and that cost-benefit analyses fail to include human factors (human beings are not necessarily rational) or social inequality. Smart equipment can be costly and is not necessarily something everyone can afford. There is also the question of what is the underlying philosophy of this constant "surveillance" of users [17]. While the ethics of technology is beyond the scope of this paper, the topic should definitely be on the agenda of researchers.

C. Crisis management

Crisis management is a sub-field of organizational studies, which draws on various fields such as psychology, management, technology, and politics. Coombs defines crisis management as "a set of factors designed to combat crises and to lessen the actual damage inflicted by a crisis" [18] Pearson and Clair [19] define crisis management as a process, where some kind of event (blackout, natural disaster, terrorism, etc.) triggers a crisis, which is met by individual and collective reactions and handled in both planned and ad hoc ways. Successful handling of a crisis depends on a range of contextual and environmental factors, of which preparedness and adoption of a crisis management mindset are among the most important ones [19].

Crises come in many shapes and sizes, ranging from the local pub running out of beer via industrial-scale crises (metal quality reports being falsified, the VW diesel scandal, BPs oil spill in the Gulf, or unexpected side effects of drugs) [20] to natural disasters such as earthquakes, floods and extreme weather [21]

The power industry is not exempt from crises. Literature shows issues ranging from natural disasters via short-term power failure to more deep-seated issues such as lacking infrastructure or too low capacity for energy generation. Examples include Pakistan, which struggles with massive energy shortages and power outages ranging between 8 hours a day in cities and 18 hours a day in rural areas [22]. In Brazil, seasonal changes in weather patterns have proven to be a challenge for their hydropower dams, with the latest major supply challenge in 2015 [23], and Nepal has struggled with both earthquakes and blockades [24].

Western countries also face electricity-related crises. In the UK, strikes in the 1980s threatened the entire power grid [25]. The US state of California had supply issues in the early '00s, forcing energy companies to introduce incentives for reducing consumption [26]. More recently, we have seen electricity prices soar in Texas due to extreme weather conditions. The current transition towards renewable energy sources also brings challenges and potential crises that need to be handled [27]. Zyadin and colleagues [28] point out lack of storage and variations in production as major challenges. The wind is not always blowing, and the sun doesn't always shine, so we need backup sources or energy storage. Traditionally, coal-fired or nuclear power plants have provided this, but environmental concerns and social pressure mean this is no longer an option. This is where V2G comes in as a possible (partial) solution to the problem, as discussed by Zdrallek et al. [29].

III. RESEARCH APPROACH

We used data from statistics Norway for household energy use, which has been fairly constant over the past years. Thus, we assume that household energy needs will remain at the same level for the next decade. To calculate the number of EVs in 2025 and 2030, we examined data from the Norwegian Information Council for the Road Traffic [32] and from the Norwegian Public Roads Administration's open data platform [33]. In addition, we conducted a document analysis of white papers on V2G and smart grid from Distribution/Transmission System Operators (DSO/TSO) and the EU. Finally, we have conducted interviews with regional DSO and TSO to discuss their views on V2G as a viable source for backup electricity.

Using the available statistical data and input from white papers and interviews, we have made a calculation of the potentially available energy from EVs, using the following approach (calculations and details are presented in detail in section IV):

1. Calculate daily energy needs nationally and for a mid-sized city
2. Calculate potential number of EV's in 2025 and 2030
3. Estimate available battery capacity in 2025 and 2030
4. Estimate available power available, taking use into account

IV. FINDINGS – SCENARIOS FOR BACKUP POWER

In this section, we present the potential of EVs as a power source in a crisis, using Norway as our case and looking at both the national level and a mid-sized city of 80.000 inhabitants.

Household energy demand

We started by examining data from statistics Norway on energy use for households (Table I). We have kept this number constant, as increased demand from EV charging so far has been offset by more efficient appliances and heating. Demand is higher in winter (about double) due to the cold climate. Nationally, households consume 70,5 GWh/day in summer and 140 GWh/day in winter. For our mid-sized city, demand is 1 GWh/day in summer and 2 GWh/day in winter.

TABLE I. DAILY ENERGY DEMAND, HOUSEHOLDS

	National demand	City demand
Summer	70,5 GWh	1 GWh
Winter	140 GWh	2 GWh

Estimation of EV numbers

The Norwegian government has said all new cars should be electric from 2025, and annual EV sales growth has been 10% in the last four years. Based on total annual sales of 150.000 cars (average new car sales 2017-19) and 10% annual sales growth, this means about 900.000 EVs

 ISBN: 978-1-61208-869-3

nationwide, and 11.000 in a mid-sized city by 2025, and 1,7 million/18.000 in 2030 (Table II). This number is close to the calculations by Saele [30], who estimates around 1,5 million EVs in Norway by 2030

TABLE II. ESTIMATED NUMBER OF EV'S

	Nationally	Mid-sized city
2025	900.000	11.000
2030	1.700.000	18.000

Estimation of battery capacity in EV's

This is, of course, nothing more than an informed guess, based on current battery sizes and predicted battery size for coming models. Currently, city cars and small cars have battery packs of 30-50 kWh, compact cars average 50 kWh, and medium-large cars have battery packs ranging from 75-100 kWh [34]. As costs of batteries go down, battery pack size is likely to increase. However, the current information on coming models shows the increase might not be as big as expected, with most models announced for the coming years have battery packs between 50 and 80 kWh.

Thus, a conservative estimate is that the average EV in 2025 has a 40 kWh battery pack, increasing to 60 kWh in 2030, as the oldest models with 18-24 kWh battery packs are gradually decommissioned.

Estimation of available energy in EV batteries

Battery capacity is one thing, state of charge something else. In order to estimate how much energy is available, we need to consider the average state of charge on parked cars at different times of the day. We have calculated this for morning and evening. As smart meters allow dynamic pricing of electricity, we assume most people charge their cars at night when prices are low. Thus, there should be more electricity available in the morning than in the evening when cars have been driven back and forth to work and activities. As with gasoline-powered vehicles, there is no reason to assume that everyone keeps a "full tank" of electricity every morning, but also that few people let their "tank" remain empty for long since both a high and a low state of charge can damage the battery. Thus, we estimate 30 kWh/car in the morning in 2025 and 40 in 2030 (Table III).

The average private car drives 11.883 km in a year, or 33 km a day. Based on the current US Environmental Protection Agency numbers from fueleconomy.com [35], if the average car mileage is 200 Wh/km, this means 6,6 kWh spent for driving each day. As we drive more some days than others, we round this up to 10 kWh, which leaves a conservative estimate of 20 kWh (2025) and 30 kWh (2030) available energy from each EV in the evening (Table III). This is, of course, again an informed guess, based on what we know of EV users' current charging habits – According to the Norwegian EV survey, most people charge at home, during the night, every 2-3 days [31]

TABLE III. AVAILABLE ENERGY IN THE MORNING AND EVENING, AVG/CAR

	Morning	Evening
2025	30 kWh	20 kWh
2030	40 kWh	30 kWh

EV's potential as a backup power source

Given the above assumptions and estimates, we can calculate the backup power potential of EVs is as follows: Multiply the number of EVs (nation and city) with available energy (morning and evening) as a percentage of daily energy demand. Tables IV and V below show how much of daily energy demand can potentially be covered by EV batteries:

TABLE IV. POTENTIALLY AVAILABLE ENERGY IN EVS, 2025. SUMMER AND WINTER, AS A PERCENTAGE OF AVERAGE HOUSEHOLD DEMAND

2025	Nation, morning	Nation, evening	City, morning	City, evening
Available energy	28 GWh	18.7 GWh	0.3 GWh	0.21 GWh
Summer, percentage of demand	40%	27%	30%	21%
Winter, percentage of demand	20%	13%	15%	10%

TABLE V. POTENTIALLY AVAILABLE ENERGY IN EVS, 2030. SUMMER AND WINTER, AS A PERCENTAGE OF AVERAGE HOUSEHOLD DEMAND

2030	Nation, morning	Nation, evening	City, morning	City, evening
Available energy	67 GWh	42 GWh	0,72 GWh	0,45 GWh
Summer, percentage of demand	95%	60%	72%	45%
Winter, percentage of demand	48%	30%	36%	23%

These numbers are not meant as absolutes, and much can change between now and 2030 in terms of household demand, EV battery sizes, and charging habits. However, they do provide us with an indication of the potential of EV batteries as part of a crisis management plan to address power outages. This is probably an optimistic estimate, as it assumes every EV sold has V2G capabilities and that households have equipment in place for feeding electricity back to the grid – both assumptions that are not true as of today. Our intention with this paper is, however, simply to point out the potential, which is not insignificant, at least for shorter-term power outages. In winter, heating is critical in the northern countries, and even a few hours can lead to people freezing in their homes. At least in 2030, the potential is there to use

 ISBN: 978-1-61208-869-3

the energy stored in EV batteries in case of power outages. The challenge then is to be ready both on a technological level and with proper crisis management plans for distribution and priorities.

In the next section, we outline some of the future research challenges that need to be addressed for this to be feasible in the near future.

V. CONCLUSION AND FUTURE RESEARCH

This paper has examined the potential of using energy stored in EV batteries, so-called V2G technology, as part of a crisis management strategy in case of power outages. Using numbers from Norway, we have shown that in 2030, the energy stored in EV batteries can potentially cover almost the entire household energy demand in the summertime. Our estimates are based on current energy demand, a careful increase in battery capacity, and Norway reaching its objective of all new cars being sold from 2025 are electric.

Of course, many factors can push this estimate both higher and lower. The percentage of new cars might decline, new battery technologies might significantly increase the potentially available energy, to name but a few.

To realize the potential, there are several challenges and factors in need of further research, both on the technical and social level:

Technical issues: As EVs cannot necessarily cover all of the demand in an emergency, grid balancing and demand needs to be carefully managed, most likely using automated algorithms and requiring smart hubs installed in homes with the ability to control household energy use. One possible research approach could be to examine "neighborhood grids", attempting to balance supply and demand within, for example, the grid served by one transformer.

Currently, few homes have the technology to reverse the current and draw power from EV batteries. Mostly we find this in households with solar panels or other forms of battery storage, such as Tesla's powerwall. Thus, equipment supporting energy transfer between home, EV, and grid needs to be installed in homes. These challenges are technically possible to overcome, but so far, households have few incentives to install expensive equipment. The most likely scenario is perhaps not V2G, but V2H – using the car's battery as a backup for individual homes. This brings us over to the **social and organizational issues:**

As the literature review showed, monetary incentives are important for households to invest in smart energy products, so research on business models that facilitate this is important. Other areas that need examination include transaction handling, where studies have appeared on, for example, the use of blockchain to handle micro-transactions between households. From a crisis management perspective, plans need to be made and implemented on both technical and social aspects. One important issue is the creation of guidelines for prioritizing demand. Even if the equipment is there to switch households appliances on and off, there is still a need for a plan on how to prioritize. Should families with small children be allowed to use more than other households? What about essential medical equipment for the elderly living at home, with various home care appliances? Heating in winter and cooling in summer demands a lot of energy, and here too, there is a need to plan and prioritize.

In summary, EVs have the potential to become a significant part of energy crisis management, but in order to make this feasible, numerous challenges need to be overcome in the coming years. As we are still in the early days of smart grids/meters/hubs/homes, we have the opportunity to plan for these things now in order to make the energy system more robust in the future.

ACKNOWLEDGMENT

This work was supported under the ERA-Net Smart Grids Plus scheme Grant number 89029 and funded through the Research Council of Norway Grant number 295750 with the project title "Multi-layer aggregator solutions to facilitate optimum demand response and grid flexibility."

REFERENCES

[1] Charging Interface Initiative, "Grid Integration Levels 2020-06-26 V5.2" [Online] http://combined-charging-system.org/fileadmin/Downloads/Papers_and_Regulations/CharIN_Levels__Grid_Integration_v5.2.pdf 2021-06-01.

[2] E. Sortomme and M.A. El-Sharkawi, "Optimal charging strategies for unidirectional vehicle-to-grid" IEEE Transactions on Smart Grid, 2(1), pp. 131–138, 2011.

[3] M.A. Quddus, M. Kabli, and M. Marufuzzaman, "Modeling electric vehicle charging station expansion with an integration of renewable energy and Vehicle-to-Grid sources" Transportation Research Part E: Logistics and Transportation Review. vol. 128, pp.251-279, 2019.

[4] H. Turker and S. Bacha, "Optimal Minimization of Plug-In Electric Vehicle Charging Cost with Vehicle-to-Home and Vehicle-to-Grid Concepts" IEEE Transaction on Vehicular Technology. vol. 67(11), pp.10281-10292, 2018.

[5] H. Zhang, Z. Hu, Z. Xu, and Y. Song, "Evaluation of Achievable Vehicle-to-Grid Capacity Using Aggregate PEV Model" IEEE Transactions on Power Systems. 32(1), pp. 784–794, 2017.

[6] W. Kempton and J. Tomić, "Vehicle-to-grid power fundamentals: Calculating capacity and net revenue" Journal of Power Sources, 144(1), pp. 268–279, 2005.

[7] F. Mwasilu, J.J. Justo, E.K. Kim, T.D. Do, and J.W. Jung, "Electric vehicles and smart grid interaction: A review on vehicle to grid and renewable energy sources integration "Renewable and Sustainable Energy Reviews 34 pp. 501-516, 2014.

[8] C. Liu, K.T. Chau, D. Wu, and S. Gao, "Opportunities and challenges of vehicle-to-home, vehicle-to-vehicle, and vehicle-to-grid technologies" Proceedings of the IEEE, 101(11), pp. 2409–2427, 2013.

[9] J. Høj, L. Juhl, and S. Lindegaard, "V2G—An Economic Gamechanger in E-Mobility?" World Electric. Vehicle Journal, 9(3) pp. 1-7, 2018.

[10] O.M. Almenning, S. Bjarghov, and H. Farahmand, "Reducing Neighborhood Peak Loads with implicit Peer-to-Peer energy trading under Subscribed Capacity tariffs" In Proc. 2nd International Conference on Smart Energy Systems and Technologies (SEST), 2019.

[11] T.H. Jackson Inderberg, H. Sæle, H. Westskog, and T. Winther, "The dynamics of solar prosuming: Exploring interconnections between actor groups in Norway" Energy Research and Social Science 70, 2020.

[12] R. Gough, C. Dickerson, P. Rowley, and C. Walsh, "Vehicle-to-grid feasibility: A techno-economic analysis of EV-based energy storage" Applied Energy, 192, pp. 12–23, 2017.

[13] A. Ahmadian et al., "Cost-Benefit Analysis of V2G Implementation in Distribution Networks Considering PEVs Battery Degradation" IEEE Transactions on Sustainable Energy, 9(2), pp. 961–970, 2018.

 ISBN: 978-1-61208-869-3

[14] X. Li et al., "A cost-benefit analysis of V2G electric vehicles supporting peak shaving in Shanghai" Electric Power Systems Research, 179, 2020.

[15] F. Berglund, S. Zaferanlouei, M. Korpås, and K. Uhlen, "Optimal Operation of Battery Storage for a Subscribed Capacity-Based Power Tariff Prosumer—A Norwegian Case Study" Energies. 12(23), pp. 1-24, 2019.

[16] H. Sale, "Flexibility potential at Norwegian households - Customer evaluations and system benefits" In Proc. International Conference on the European Energy Market, EEM. IEEE Computer Society, 2020.

[17] T.M. Skjølsvold, I.F. Fjellså, and M. Ryghaug, "Det fleksible mennesket 2.0: Om sosiale relasjoner i fremtidens digitale elektrisitetsystem" Norsk Sosiolgisk Tidsskrift. 3(3), pp. 191–208, 2019.

[18] W. Coombs, "Ongoing crisis communication: Planning, managing, and responding 5th edition," Sage Publishing, 2019.

[19] C.M. Pearson and J.A. Clair, "Reframing Crisis Management," Academic Management Review, 23, pp.59-76, 1998.

[20] W.T.Coombs and D. Laufer, "Global Crisis Management – Current Research and Future Directions," Journal of International Management, 24 pp. 199–203, 2018.

[21] BK. Sovacool, M. Tan-Mullins, W. Abrahamse, "Bloated bodies and broken bricks: Power, ecology, and inequality in the political economy of natural disaster recovery," World Dev. 110, 243–255 (2018

[22] G. Valasai et al." Overcoming electricity crisis in Pakistan: A review of sustainable electricity options," Renewable Sustainable Energy Review, 72 pp. 734–745, 2017.

[23] J.D. Hunt, D. Stilpen, M.A.V. de Freitas, "A review of the causes, impacts and solutions for electricity supply crises in Brazil," Renewable Sustainable Energy Review 88, pp. 208–222, 2018.

[24] G. Underwood, D. Hill, and S. Lamichhane, "Earthquakes, blockades and energy crises: A conceptual framework for energy systems resilience applied to Nepal," Energy Research and Social Science. 69, 2020.

[25] F. Ledger and H. Sallis, "Power struggles: crisis management in the UK electricity industry," IEE Review, 41(1), pp. 21–25,1995.

[26] C.A. Goldman, G.L. Barbose, and J.H. Eto, "California customer load reductions during the electricity crisis: Did they help to keep the lights on?", Journal of Industry, Competition and Trade, 2, pp. 113-142, 2002.

[27] F.W. Geels et al., "The enactment of socio-technical transition pathways: A reformulated typology and a comparative multi-level analysis of the German and UK low-carbon electricity transitions (1990–2014)," Research Policy, 45(4), pp. 896–913, 2016.

[28] A. Zyadin, P. Halder, T. Kähkönen, and A. Puhakka, "Challenges to renewable energy: A bulletin of perceptions from international academic arena," Renewable Energy, 69, pp. 82–88, 2014.

[29] M. Zdrallek, I. Hübl, and C. Böse, "Special Report - Session 3: Operation, Control and Protection," 23rd International Conference on Electricity Distribution, 2015.

[30] H. Saele and I. Petersen, "Electric vehicles in Norway and the potential for demand response," In Proc. 53rd International Universities Power Engineering Conference, UPEC 2018. IEEE, 2018.

[31] Elbilforeningen: Elbilisten 2020 - a survey of Norwegian EV users.

[32] Information Council for Road Traffic [Online] https://ofv.no, 2021.06.01

[33] Norwegian Public Roads Administration Open Data Portal [Online] https://dataut.vegvesen.no, 2021.06.01

[34] EV Norway [Online] https://elbil.no/om-elbil/elbiler-idag/ 2021.06.01

[35] US Department of Energy [Online] http://fueleconomy.org 2021.06.01

 ISBN: 978-1-61208-869-3