Proceedings of the 14th ILAIS Conference

June 29, 2020

The Israel Association for Information Systems (ILAIS) was founded in 2005 as the Israeli chapter of the Association for Information Systems (AIS). The goal of ILAIS is to promote the exchange of ideas, experiences and knowledge among IS scholars and professionals engaged in IS development, management and use.

Program Chairs:

Dr. Itzhak Shemer – Ramat-Gan Academic College

Dr. Idan Roth - The Academic College of Tel Aviv-Yaffo

Program Committee:

Dr. Dizza Beimel – Ruppin Academic Center

Dr. Ofir Ben-Assuli – Ono Academic College

Prof. Opher Etzion – Zefat Academic College

Dr. Adir Even – Ben-Gurion University

Prof. Lior Fink – Ben-Gurion University

Prof. Roy Gelbard – Bar-Ilan University

Dr. Adi Katz – SCE Ashdod College of Engineering

Dr. Dalit Levy – Zefat Academic College

Prof. Daphne Raban – University of Haifa

Dr. Idan Roth – The Academic College of Tel Aviv-Yaffo

Dr. Itzhak Shemer – Ramat-Gan Academic College

Dr. Lior Zalmanson – Tel-Aviv University

Preface

This volume contains the extended abstracts of the papers presented at ILAIS-2020, the 14th Israel Association for Information Systems Conference, held virtually on June 29th 2020 and hosted by Ramat-Gan Academic College.

This conference has been planned as a conventional conference, but the restrictions imposed by the health authorities due to the CoronaVirus pandemic forced the program committee to move out of the comfort zone and adopt a virtual format for the conference. One of the positive implications of this virtual format is the ability to broaden the audience and have participants (and speakers) from abroad, hence the conference is held is the afternoon (Israel time).

There were 19 submissions. Each abstract was reviewed by at least 3 program committee. Taking into account the virtual format of the conference – the program committee decided to accept 11 abstracts.

I would like to take this opportunity to whole-heartedly thank the people who made significant contributions to the conference:

- Our keynote speaker - Prof. Anindya Ghose from NYU.
- All the authors of the submitted abstracts, and the 11 authors who present their research in this conference.
- The panelists of the panel for doctoral students – led by Prof. Lior Fink from BGU.
- The panelists of the Zoon-in/Zoom-out panel – Led by Dr. Lior Zalmanson from TAU.
- Dr, Idan Roth – who shared his experience in integrating theory and practice in the CoronaVirus “trenches”.
- The members of the Program Committee (all listed in the cover page).
- ILAIS officers – David Schwartz, Dizza Beimel & Dalit Levy.

Itzhak Shemer

ILAIS-2020 Program

Pre-Conference Session (participation by invitation only**)**

11:30-13:00	Becoming a Researcher and a Faculty Member: An Interactive Panel for Doctoral Students *Panelists: Lior Fink (Moderator), Gal Oestreicher-Singer, David Schwartz, Iris Reinhartz-Berger, Ofir Ben-Assuli*

Opening Session

14:00-14:10	Welcome *David Schwartz – ILAIS Chair* *Itzhak Shemer – Conference Co-Chair*
14:10-14:25	An IS Faculty in the Eye of the CoronaVirus Storm (During Reserve Service) *Idan Roth*
14:25-15:10	Zoom in, Zoom Out: Perspectives on Remote Learning under COVID-19 and Beyond *Panelists: Lior Zalmanson (Moderator), Sheizaf Rafaeli, Carmel Vaisman, Shirley Bar Lev, Tal Levi, Dalit Levy*

Parallel Sessions

	Session 1A *Chair: Ofir Ben-Assuli*	**Session 1B** *Chair: Dizza Beimel*
15:20-15:40	Non-Alcoholic Fatty Liver Predictive Analytics *Orit Goldman, Ofir Ben-Assuli, Shira Zelber-Sagi, Shani Shenhar-Tsarfaty, Ori Rogowski, David Zeltser, Itzhak Shapira, Shlomo Berliner* **p.5**	The Effect of Consumer Engagement on Mobile Consumption *Iris Somech, Shachar Reichman* **p.11**
15:40-16:00	A Surprising Question on the Correct Way to Implement Association Rules – A Pharmacovigilance Case *Marina Goldman and David Bodoff* **p.7**	My Mom Was Getting this Popup: Understanding Motivations and Processes in Helping Older Relatives with Mobile Security and Privacy *Tamir Mendel, Eran Toch* **p.13**
16:00-16:20	Exploring Insights: Patient Interaction Using Personalized Innovative Technology *Kofi Amankwah Boamah, Iris Reychav, Roger McHaney, Esther Saiag, David Zeltser, Eran Lederman* **p.10**	The Guide to Content Moderation: Introducing Crowds to Mitigate the Challenges of the Human Moderator *Lior Zalmanson, Inbal Yahav, Dena Yadin* **p.15**

Parallel Sessions

	Session 2A *Chair: Opher Etzion*	**Session 2B** *Chair: Adir Even*
16:30-16:50	Third-Party Induced Cybersecurity Incidents: Who Pays the Price *Michel Benaroch* **P.18**	Learning the UML Class Diagram Using the Learning from Errors Approach *Ronit Shmallo, Tammar Shrot*
16:50-17:10	Does a Threshold Model Predict Individuals' Time of Adoption of Innovations? **P.22** *Mor Atlas, Arieh Gavious, Gilad Ravid*	The Motivation for Motivation Theories: A Systemic Literature Review on the Use of Motivation Theories in RE Research *Naomi Unkelos-Shpigel, Irit Hadar* **P.27**
17:10-17:30	Would you believe it if an AI told you that 2 + 2 is 5? Conformity to algorithmic recommendations *Yotam Liel, Lior Zalmanson* **P.24**	

Closing Session

17:45-18:30	**Keynote Talk: Trading Privacy for the Greater Social Good: How Did America React During COVID-19?** *Anindya Ghose* **P.30**
18:30-18:45	Best paper award
18:45-19:00	Open ILAIS Planning Meeting *Hosted by David Schwartz, Dizza Beimel & Dalit Levy.* All are welcome to join

NON-ALCOHOLIC FATTY LIVER PREDICTIVE NALYTICS

Research in Progress

Goldman Orit, Ben-Assuli Ofir, Faculty of Business Administration, Ono Academic College, Israel
Zelber-Sagi Shira, Faculty of Social Welfare and Health Sciences, University of Haifa, Israel
Shenhar-Tsarfaty Shani, Rogowski Ori, Zeltser David, Shapira Itzhak, Berliner Shlomo, Sackler Faculty of Medicine, Tel-Aviv university, Israel

Keywords: Machine Learning, Risk Prediction, Non-alcoholic fatty liver disease

1 Introduction

Non-alcoholic Fatty liver disease (NAFLD) is becoming a major global health burden in both developed and developing countries (Loomba & Sanyal, 2013). NAFLD is the most prevalent liver disease (30% in the general population) and may progress to liver cirrhosis, liver cancer and can increase the risk for cardiovascular disease and type-2 diabetes (Zelber-Sagi et al., 2006). So, it is crucial to identify NAFLD patients at risk, in order to implement therapeutic interventions that can help prevent/reverse the deleterious consequences of it. The Fatty Liver Index (FLI) is a simple validated calculating tool for the prediction of fatty liver in the general population (Bedogni et al., 2006, Zelber-Sagi et al., 2013).

2 Objectives

Few works have dealt with predicting FLI (Ruhl and Everhart 2015). New models focusing on predictive analytics for chronic patients based on data mining tools require large, broad datasets as we are using here. This study utilizes machine learning approach for risk stratification by identifying and predicting high risk patients for FLI and assessing the potential protective factors.

Our goal is to stratify patients' risks for NAFLD and advanced fibrosis after predicting their future development of their disease and suggest preventive medical decisions.

3 Research Methods

The data we use is a prospective cohort of apparently healthy volunteers from the Tel-Aviv medical center inflammation survey (TAMCIS), admitted for routine annual health check-up. The data analysis is performed according to the following steps :1. Assemble Visits' data from TAMCIS database. 2. Transforming into Patients' Level Data and create patient-level panel (FLI Prediction Panel containing 7,581 patients) 3. Data preparation - Computation of time series variables. 4. Machine learning- Classification models to predicting individual patient risk. 5. Analyzing the findings with performance measures. 5. Recommendations for preventable medicine.

4 Preliminary Results

Our preliminary results show after incorporating time covariates and other key variables that our technique slightly outperformed the predictive power of existing methods (AUC = 0.8486). In addition, we have identified new features that have been used as powerful factors in the predictive process.

The importance of the variables shown in Figure 1 is for the developed model on the whole data set. We have calculated the most important variables according to 5 models by their number of occurrences in the TOP10 of the 5 models. The most powerful factors for prediction in all the models were the last value of BMI, first value of weight, first value of GammaGT and last value of Triglycerides.

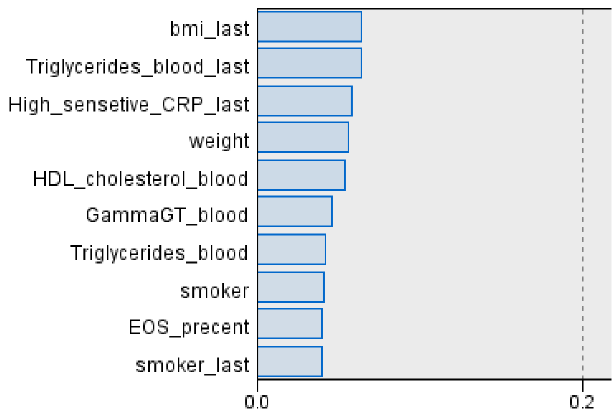


Figure 1. *FLI - CHAID variable importance*

5 Preliminary Conclusions

In this age of precision medicine, interventions to prevent the progression of chronic disease can reduce medical complications and disease burden substantially. Previous works have found AUC in predicting FLI of 0.82 (Zelber-Sagi et al., 2013), 0.78 (Ruhl & Everhart, 2015), 0.83 (Huang et al., 2015). Our results slightly outperformed the predictive power of existing works. The findings of this study (with extensions of more outcome variables) may help in risk stratification for disease progression and planning future preventive strategies based on lifestyle modifications and medical treatment implemented to reduce disease burden.

6 References

Bedogni G, Bellentani S, Miglioli L, et al. The Fatty Liver Index :a simple and accurate predictor of hepatic steatosis in the general population. BMC Gastroenterol 2006;6:33 doi: 1471-230X-6-33 [pii] 10.1186/1471-230X-6-33[published Online First: Epub Date].|

Huang, X., Xu, M., Chen, Y., Peng, K., Huang, Y., Wang, P., . . . Chen, Y. (2015). Validation of the fatty liver index for nonalcoholic fatty liver disease in middle-aged and elderly Chinese. *Medicine, 94*(40).

Loomba R, Sanyal AJ. The global NAFLD epidemic. Nat Rev Gastroenterol Hepatol 2013;10(11):686-90 doi: 10.1038/nrgastro.2013.171[published Online First: Epub Date]

Ruhl C, Everhart J. Fatty liver indices in the multiethnic United States National Health and Nutrition Examination Survey. Alimentary pharmacology & therapeutics 2015;41(1):65-76

Zelber-Sagi S, Nitzan-Kaluski D, Halpern Z, Oren R. Prevalence of primary non-alcoholic fatty liver disease in a population-based study and its association with biochemical and anthropometric measures. Liver Int 2006;26(7):856-63 doi: 10.1111/j.1478-3231.2006.01311.x[published Online First: Epub Date].|

Zelber-Sagi S, Webb M, Assy N, et al. Comparison of fatty liver index with noninvasive methods for steatosis detection and quantification. World journal of gastroenterol

A SURPRISING QUESTION ON THE CORRECT WAY TO IMPLEMENT ASSOCIATION RULES – A PHARMACOVIGILANCE CASE

Complete Research

Goldman Marina, Bodoff David, University of Haifa

Keywords: data mining, association rules, pharmacovigilance.

1 Introduction

Pharmacovigilance is the set of activities that manage the risk of harmful or dangerous drug side-effects, called Adverse Events (AEs). The process includes ongoing monitoring throughout a drug's life, as patients take the drug and report possible side-effects. These reports are collected in databases by the drug companies and by regulators such as the US FDA. Data mining techniques are then applied, looking for patterns of adverse events that exceed certain thresholds of probability. The regulator provides guidance when such patterns are discovered.

Association rules (AR) [1] is the primary data mining technique that is used for pharmacovigilance [2]. Originally developed to analyse consumer behaviour, its application to pharmacovigilance data seems straightforward. However, it came to our attention that different drug companies use different calculations in their implementations of AR. The two different calculations can affect whether a given pattern exceeds the thresholds and triggers a regulatory and medical response. The same two calculations apply to all applications of AR, yet the issue has not been raised in the literature. The purpose of this research is to lay out the problem. We demonstrate that the different calculations lead to different reports; we analyse the conditions under which the two calculations diverge; and we comment on some advantages of each method. In the particular case of pharmacovigilance, our work is a call to the industry and to regulatory bodies to address the issue, and adopt on a single standard approach.

2 Research Questions

The input for AR in pharmacovigilance has the form shown in Table 1a. Each row is a case report; a person was taking certain drugs A, B…, and experienced certain adverse effects I, II, etc. The goal of AR mining is to figure out if there is an association between any given drug and any given AE. From these case reports, contingency tables are derived for each drug-AE pair. Table 1b is an example, for the pair of Drug B and AE II. Finally from the contingency table, ratios are derived and compared with a threshold. For example P(II|B)/p(II| notB) =3/4 / 1/3 = 9/4. If this ratio is greater than a threshold (typically 2.0), then a red flag is raised for further investigation. This is the more common approach. An alternative approach, which is also correct and rests on more relaxed probabilistic assumptions, first breaks down each case report into every possible pair of drug and AE. Table 2a shows how the first 2 case reports from Table 1a would appear in this form. The resulting contingency table is shown in Table 2b. Now, P(II|B)/p(II| not-B) = 3/6 / 6/10 = 5/6, far below any red flag.

The research questions are: (1) Are both approaches correct? Under what probabilistic assumptions? (2) Under what circumstances do the two methods diverge? (3) What are the advantages of each approach?

Case ID	Items
1	B, I
2	A, B, C, I, II
3	B, I, II
4	A, B, C, II
5	A, I
6	C, I
7	A, C, II

	Drug B	Other drugs (Not drug B)	Total
Adverse event II	3	1	4
Other adverse events (Not adverse event II)	1	2	3
Totals	4	3	7

Tables 1a-b. Method #1. Table 1a shows the raw case reports; Table 1b shows the resulting contingency table for Drug B, possible AE II

Original Case No	A	B	C	I	II
1		B		I	
2	A			I	
2	A				II
2		B		I	
2		B			II
2			C	I	
2 **Etc**			C		II

	Drug B	Other drugs (Not drug B)	Total
Adverse event II	3	6	9
Other adverse events (not Adverse event II)	3	4	7
Totals	6	10	16

Tables 1a-b. Method #2. Table 2a shows part of the case reports, after they are broken down into Drug-AE pairs; Table 2b is the resulting contingency table for Drug B, possible AE II

3 Methods

We used the publicly available FDA database FAERS for 2018. Data cleaning was especially focused on removing duplicate case reports, a known problem with this data [3]. We implemented AR mining in Python for each of the two methods, and analyzed various data sub-samples. We first demonstrate empirically that results differ. Then we trace and analyze what drives the different results, which allows us to characterize when the two approaches will diverge. Finally, we use domain knowledge – what is important in pharmacovigilance – to shed light on the possible advantages of each approach.

4 Results

The two methods indeed produce different results. The methods diverge when, in the particular data subset being investigated, certain drugs occur in many cases, each with numerous AE's. An example is, all patients suffering from a given disease. We find that both methods are valid. Based on domain knowledge, the second approach better fulfills an important need for the pharmacovigilance task.

5 Discussion

This work calls attention to two methods of implementing AR mining, a supposedly mature method in data mining, and to the actual existing split within pharmacovigilance practice. Our results close an unaddressed open question that applies to all applications of AR mining. In the particular case of pharmacovigilance, there are obvious reasons why industry and regulators might wish to adopt a single standard approach.

6 References

[1] R. Agrawal , T. Imielinski and A. Swami, "Mining association rules between sets of items in large databases.," in *Proceedings of the ACM SIGMOD International*, 1993.

[2] R. Harpaz, H. Chase and C. Friedman, "Mining multi-item drug adverse effect associations in spontaneous reporting systems.," *BMC Bioinformatics, 11(S-9):S7,* vol. 11, no. Suppl 9, p. S7, 2010.

[3] C. Wong, S. Ho, B. Saini, D. Hibbs and R. Fois, "Standardisation of the FAERS database: a systematic approach to manually recoding drug name variants," *Pharmacoepidemiology and Drug Safety,* vol. 24, pp. 731-737, 2015.

EXPLORING INSIGHTS: PATIENT INTERACTION USING PERSONALIZED INNOVATIVE TECHNOLOGY

Complete Research

Boamah, Kofi Amankwah, Ariel University, Israel, kofiba@ariel.ac.il
Reychav, Iris, Ariel University, Israel, irisre@ariel.ac.il
McHaney, Roger, Kansas State University, USA, mchaney@ksu.edu
Saiag, Esther, Tel Aviv Medical Center, Israel, esthers@tlvmc.gov.il
Zeltser, David, Tel Aviv Medical Center, Israel, davidz@tlvmc.gov.il
Lederman, Eran, Tel Aviv Medical Center, Israel, eranled@tlvmc.govm.il

Keywords: Information Technology, Personalization, Human Computer Interaction

When used appropriately, information technology can positively impact healthcare with benefits for both patients and health service providers. Recent research into gender and age differences, and related influences on patterns of information technology usage suggests statistically significant differences exist along those dimensions (Gallagher et al., 2017; Goswami & Dutta, 2015; Shaouf, 2018). In contrast, other studies indicate these differences may not be clear (Shaouf, 2018). Specifically, general research that focuses on age differences often suggests younger people are more likely to use new technology than older people (Andone et al., 2016). However, in health care areas, this may not be true. For example, some studies show older adults have a higher need for the benefits offered by technology, and hence engage in more frequent usage (Kim et al., 2016; Zhao et al., 2018). The current study posits that age and gender differences should not prohibit targeted subsets of users from the potential benefits of new technology. The current study seeks to demonstrate how technology personalization can act as an equalizer in this area and provide benefits that cannot be overlooked.

Currently, four primary generations comprise the population: Baby-Boomers, Generation X, Generation Y, and Generation Z. Messarra (2014) indicated generational differences are likely to arise between individuals or groups due to variations in values, expectations, needs, workplace practices, and personalities. This could produce conflicting actions and preferences. Other factors including complexity, and inaccessibility may favor one group while leading to the detriment of another (Messarra, 2014). In the current study, we seek to ensure each group (broken down by age or gender) has an equal opportunity to access technology through personalization. To do this, Telesofia Medical has developed an innovative platform that generates personalized medical instructional videos to specifically meet discharged patients' needs and clarify written medical instructions. This tool specifically targets all discharged patients (Telesofia, 2020a). Each patient receives information related to their diagnosis, discharge instructions, recommendations and, if necessary, upcoming appointment schedule. Actions are demonstrated using models specific to each patient's demography (Figure 1).

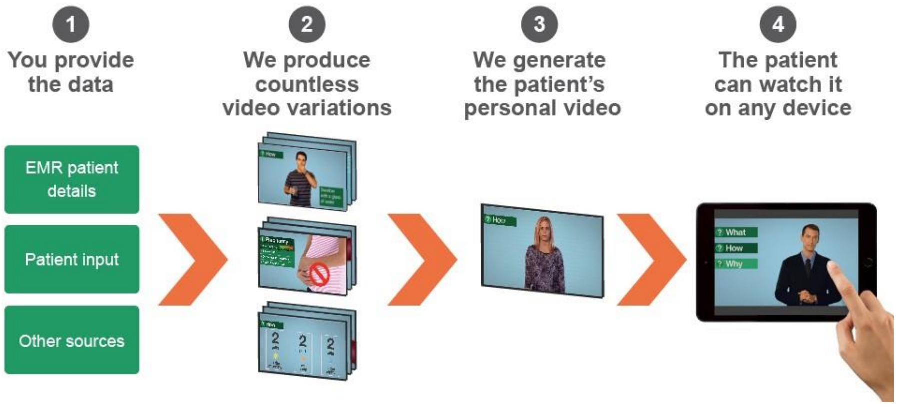


Figure 1. Telesofia Platform Process (Telesofia, 2020b).

THE EFFECT OF CONSUMER ENGAGEMENT ON MOBILE CONSUMPTION

Research in Progress

Somech, Iris, Tel Aviv University, Tel Aviv, Israel, iris@mail.tau.ac.il
Reichman, Shachar, Tel Aviv University, Tel Aviv, Israel, sr@tauex.tau.ac.il

Keywords: Mobile Commerce, Consumer Engagement, App Download, Positive and Negative Engagement

1 Introduction

Mobile platforms have become a mainstream venue for e-commerce, with mobile commerce (m-commerce) accounting for more than 40% of total e-commerce revenues in the U.S. in 2019. Revenues are projected to grow steadily in the next few years (Statista1). An important segment of m-commerce is the mobile application market, where consumers download applications from an app store. In 2019, 204 billion app downloads generated revenues of over $ 460 billion (Statista2, Statista3). In e-commerce, engagement, which is defined as a user's interaction with a website (Clifton 2012; Rojas and Alejandro 2014), has been shown to have a positive influence on sales and makes a significant contribution to sales prediction (Montgomery et al. 2004, Van den Poel and Buckinx 2005). As a result, retailers aim to maximize traffic and online engagement (Agarwal and Venkatesh 2002, Brynjolfsson et al. 2013). Since prior research on user engagement referred mainly to PC environments, and since other studies showed that users behave differently in mobile and PC environments (Bang et al. 2013, Ghose et al. 2013), we believe that understanding how mobile engagement influences consumption in general and mobile app sales in particular has significant academic and practical contribution.

2 Objective and Contribution

This study aimed to identify mobile engagement features that have a positive (vs. negative) contribution to mobile consumption. In addition, it aimed to evaluate the extent of this contribution and which user and/or device factors mediated this effect. The potential academic contribution is to provide insights into the causal effect of engagement features on mobile consumption. The potential practical contribution, is to increase conversion rate by designing more effective mobile interfaces. Designers are prompt to reconsider which engagement features to highlight and promote and which to selectively disable for certain types of consumers.

3 Methodology

To better understand the role of user engagement in conversion, we studied the causal effects of selected engagement features (gallery scroll, gallery open, read more description, video and reviews) on purchase probability. We evaluated the extent of each effect and whether it was positive or negative. We conducted a lab experiment to empirically measure the causal effect of engagement features on conversion in a controlled environment. For this purpose, we developed designated App Store/Google Play-like pages for two apps from two different categories (Games and Food & Drinks). Each of the application store pages had five variations (treatment groups), in which a single engagement feature was disabled at a time. In the control condition, all tested engagement features were enabled.

Amazon Mechanical Turk (MTurk) was used as an online recruitment platform to run the experiment, and only mobile users were allowed to participate. We first asked participants a few questions about their app download habits. Next, we instructed them to browse the page as they usually do and decide whether or not to install the app. Then, the participants landed on the store page. After clicking on the install button or exiting the app page, we asked participants to rate how similar their visit was to previous visits to other application stores. We also captured the activities users performed on the store page and user-related information such as operating system (OS) and browser.

We run logistic regression for different subsets of the data with and without user activities and attributes, to evaluate the effect of engagement features on app download probability,

4 Initial Results

2,934 participants took part in the experiment. We analysed several subsets of each treatment-control separately. When *disabling* the option to scroll the gallery, our results indicate a negative effect on downloads, for participants who used android devices and selected the Food & Drinks app. The relative decrease in the odds of downloading is 77%. For a female-only subset, the effect is even stronger. The opposite effect (positive) exists for participants who used iPhone devices.

When *excluding* the option to read the full description from the store page, the effect on downloads is positive. The relative increase in the odds of download is 26%. The effect is even stronger for users who were less interested in the app category they selected in the experiment.

When *excluding* the video from the store page, results indicate opposite effects by gender (for some of the subsets) – a negative effect for male and a positive effect for females.

Initial results show that the causal effect of engagement features on downloads may change with the user's device type, gender and level of interest. In this experiment setting, while gallery scroll is a positive engagement feature for android users, it is a negative engagement feature for iPhone users. Results also indicate that providing access to read the full description is a negative engagement feature, and that video effect on downloads may vary by gender.

5 References

Agarwal, R., and Venkatesh, V. 2002. "Assessing a Firm's Web Presence: A Heuristic Evaluation Procedure for the Measurement of Usability," *Information Systems Research* (13:2), pp. 168-186.

Brynjolfsson, E., Hu, Y. J., and Rahman, M. S. 2013. "Competing in the Age of Omnichannel Retailing," *MIT Sloan Management Review* (54:4), pp. 23-29.

Bang, Y., Han, K., Animesh, A., and Hwang, M. 2013. "From Online to Mobile: Linking Consumers’ Online Purchase Behaviors with Mobile Commerce Adoption," *in Proceedings of the Pacific Asia Conference on Information Systems*.

Clifton, B. 2012. "Advanced Web Metrics with Google Analytics", *John Wiley & Sons*.

Ghose, A., Goldfarb, A., and Han, S. P. 2013. "How is the Mobile Internet Different? Search Costs and Local Activities," *Information Systems Research* (24:3), pp. 613-631.

Montgomery, A. L., Li, S., Srinivasan, K., and Liechty, J. C. 2004. "Modeling Online Browsing and Path Analysis Using Clickstream Data," *Marketing Science* (23:4), pp. 579-595.

Rojas, C., and Alejandro, R. 2014. "Optimising User Engagement in the Public Sector: A Web and Social Media Analytics for the Lambeth Council," *University of Westminster. Westminster*. http://repositorio.educacionsuperior.gob.ec/handle/28000/1617.

Statista1. "U.S. mobile retail commerce sales as percentage of retail e-commerce sales from 2017 to 2021", from https://www.statista.com/statistics/249863/us-mobile-retail-commerce-sales-as-percentage-of-e-commerce-sales/

Statista2. "Number of mobile app downloads worldwide from 2016 to 2019", from https://www.statista.com/statistics/271644/worldwide-free-and-paid-mobile-app-store-downloads/

Statista3. "Worldwide mobile app revenues in 2014 to 2023", from https://www.statista.com/statistics/269025/worldwide-mobile-app-revenue-forecast/

Van den Poel, D., and Buckinx, W. 2005. "Predicting Online-Purchasing Behaviour," *European Journal of Operational Research* (16)

MY MOM WAS GETTING THIS POPUP: UNDERSTANDING MOTIVATIONS AND PROCESSES IN HELPING OLDER RELATIVES WITH MOBILE SECURITY AND PRIVACY

Complete Research

Mendel, Tamir, Tel Aviv University, Israel, tamirmendel@mail.tau.ac.il
Toch, Eran, Tel Aviv University, Israel, erant@tauex.tau.ac.il

Keywords: security and privacy, social computing, human computer interaction (HCI), older adults, mobile computing.

1 Introduction

Security and privacy pose a serious barrier to the use of mobile technology by older adults (Kuerbis et al. 2017). Older adults are more vulnerable to attacks based on mobile technology, such as phone-based fraud and phishing (Zhou et al. 2014). As older adults face more difficulties managing their security and privacy mechanisms and settings (Xie et al. 2012), without urgently addressing security and privacy challenges, their access to vital technology is seriously limited (Xie and Bugg 2009). Several studies and products are based on designing interfaces specifically for older adults, such as a social networking (Pensas et al. 2013), but these technologies do not cover all necessary applications and they do not cover the newest applications. Based on older adults' preferences for in-person support and being guided by friends and family (Damodaran et al. 2014), we are driven by curiosity regarding these types of family support networks.

2 Objectives and Research Questions

Our underlying goal is to improve older adults' security and privacy by understanding the process of helping older relatives with mobile security and privacy challenges. We focus on understanding and enhancing existing support networks, using social support to help older relatives overcome security and privacy problems. To design social support processes, we examined the following questions: What are the characteristics of help processes for older relatives? How do these processes differ from helping other social groups? What factors are related to the willingness to help older relatives?

3 Research Methods

To investigate our research questions, a user study was conducted based on a questionnaire. We recruited participants via Amazon Mechanical Turk. We have used qualitative and quantitative methods in analyzing support stories and questionnaire by 187 participants.

4 Results

We found a variety of triggers and assistance modes in security and privacy support processes. Triggers are defined as an event that starts the support process. We found that in 41% of the support reports participants have fixed the problem for their older relatives, giving advice was 17%, guiding was 19%, and demonstrating was 23%. The helper actual assistance mode is significantly different between the older relative and other social groups (Chi-square test: $\chi 2= 10.178$; df = 3; p-value= 0.017), with guiding is more expected for older relatives (18/86) than to other social groups (5/101).

A linear regression model was created to predict the willingness to assist older relatives (Adjusted $R^2 = 0.35$). The dependencies variables are emergency, time, familiarity, exposure, altruism, assisting frequency and attitude. The familiarity with preferences (β=0.29, p-value<0.001) and exposure approval (β=0.26, p-value<0.05) have a significant positive effect on the willingness to assist older relatives.

5 Discussion and Conclusions

When the helper perceives that older relatives expected guidance, then in-depth assistance provides to older relatives especially. Currently, existing support technologies mostly focus on allowing others to fix the problem on behalf of the person, but effective support technologies should support most of assistance modes. Furthermore, we show that familiarity with an older relative's preferences is essential in providing meaningful support and protecting the older relative's information is important to the helper in providing assistance. Our findings in the context of mobile security and privacy show important design insight to develop collaborative technologies for social support.

6 References

Damodaran, L., Olphert, C.W., and Sandhu, J. (2014) 'Falling off the Bandwagon? Exploring the Challenges to Sustained Digital Engagement by Older People'. *Gerontology* 60 (2), 163–173

Kuerbis, A., Mulliken, A., Muench, F., A. Moore, A., and Gardner, D. (2017) 'Older Adults and Mobile Technology: Factors That Enhance and Inhibit Utilization in the Context of Behavioral Health'. *Mental Health and Addiction Research* [online] 2 (2).

Pensas, H.., Kivimäki, T.., Vainio, A.-M.., Konakas, S.., Costicoglou, S.., Kölndorfer, P.., Summanen, K.., Moisio, H.., and Vanhala, J.. (2013) 'Building a Client-Server Social Network Application for Elders and Safety Net'. *Proceedings of the 17th International Academic MindTrek Conference: 'Making Sense of Converging Media', MindTrek 2013* 310–312

Xie, B. and Bugg, J.M. (2009) 'Public Library Computer Training for Older Adults to Access High-Quality Internet Health Information'. *Library and Information Science Research* [online] 31 (3), 155–162.

Xie, B., Watkins, I., Golbeck, J., and Huang, M. (2012) 'Understanding and Changing Older Adults' Perceptions and Learning of Social Media'. *Educational Gerontology* 38 (4), 282–296

Zhou, J., Rau, P.L.P., and Salvendy, G. (2014) 'Age-Related Difference in the Use of Mobile Phones'. *Universal Access in the Information Society* 13 (4), 401–413

The Guide to Content Moderation: Introducing Crowds to Mitigate the Challenges of the Human Moderator

Lior Zalmanson, Inbal Yahav, Dena Yadin , Tel-Aviv University

1 Introduction

The increased reliance on User Generated Content (UGC) in the business models of online environments, brings with it a challenge of content moderation (Gillespie 2017). Namely, making sure the content submitted to the website by its users is not uncivil, illegal, harmful, deceitful, or offensive to the websites' community (Langvardt 2017). Website owners are required to employ, in some cases, thousands of human content moderators, to address this issue, and this number increases with the growth in website popularity (Roberts 2019). Recently, discussions in both academia and the media have addressed the managerial and personal challenges of human content moderators who have to make many moderation decisions under pressure as well as endure being exposed to many hours of harmful content (Gillespie 2017, Langvardt 2017, Link et al. 2016, Roberts 2019)[1].

At present, content moderation requires the human-in-the-loop approach due to multiple statistical and algorithmic challenges. The first challenge is the subjective, dynamic and culture-specific nature of UGC (Link et al. 2016). The second is that methodologically, the problem requires "understanding" a large amount of terminology and is sensitive to sentence structure, grammar and type of speech that in themselves vary in meaning in the many different domains and contexts (Durate et al. 2018).

In current best practices, the firm provides the human moderators with a list of guidelines. To ensure the relevance of the guidelines, the rules need to be constantly assessed, examined and updated by the firm, which adds to the pressure and the confusion of moderators. In our pretest, we collaborated with "Haaretz" - one of the largest news sites in Israel and found that while 43% of comments posted on the website were rejected in practice, only 18.8% were, in fact, inappropriate. Further investigation revealed a highly inconsistent and controversial moderating process.

Our work seeks to offer a different approach to content moderation, that avoids the perils of human moderation and lowers the complexity of algorithmic moderation. Here, we show how crowdsourcing can be used in symbiosis with NLP algorithms to compile free-text information on perceptions of inappropriate content that can be used inclusively to the automatic generating of content-moderation guidelines.

2 Preliminary study

Data for this study consisted of 140 Facebook comments on articles on the topics of Bitcoin and gun control, out of which 70 comments were qualitatively labeled as uncivil, and 70 comments were civil. Our objective was to identify the different types of online inappropriateness and, in specific, online incivility. For that, we used a hybrid of crowd freetext annotation of the comments' content and advanced NLP models. Specifically, 255 nonexpert human crowd members were asked to re-label each comment as "civil" or "uncivil" and explain the reasoning for their labeling decision. Each comment was labeled and meta-labeled by 10 crowd members. The crowd annotation served as "content meta-data," which held a higher abstraction level of information of the raw comments. Examples for crowd annotations were:

(1) *"Its a little rude, and provides nothing to further the conversation."* (for an uncivil comment)

(2) *"It's not conventional to use all capital letters. This is a little impoilte since it implies screaming. "* (for an uncivil comment)

(3) *"There's really no name calling or provacation."* (for a civil comment)

(4) *"It is polite"* (for a civil comment)

The content meta-data was then algorithmically analyzed and categorized to establish highlevel concepts of inappropriate content, following the Formal Concept Analysis (FCA, see Priss, 2006) framework.

As a benchmark, we used definitions of types of incivility based on the best practices depicted in the literature on incivility (Coe et al. 2014, Kenski et al. 2017), as well as the definition of the types based on professional qualitative analysis of our observational data (the comments, as labeled by a qualitative research expert). The

[1] https://edition.cnn.com/2019/02/28/tech/facebook-google-content-moderators/index.html

types identified in the literature, our qualitative analysis, and the crowd-based algorithm are presented in Table 1. Figure 1 shows the word embedding graphs of civil and uncivil content meta-data, that was collected from the crowd. An attempt to run unsupervised topic analysis on the raw uncivil comments did not yield insightful results with regards to the types of incivility but rather captured the topics of the articles themselves, which were discussed by the commenters.

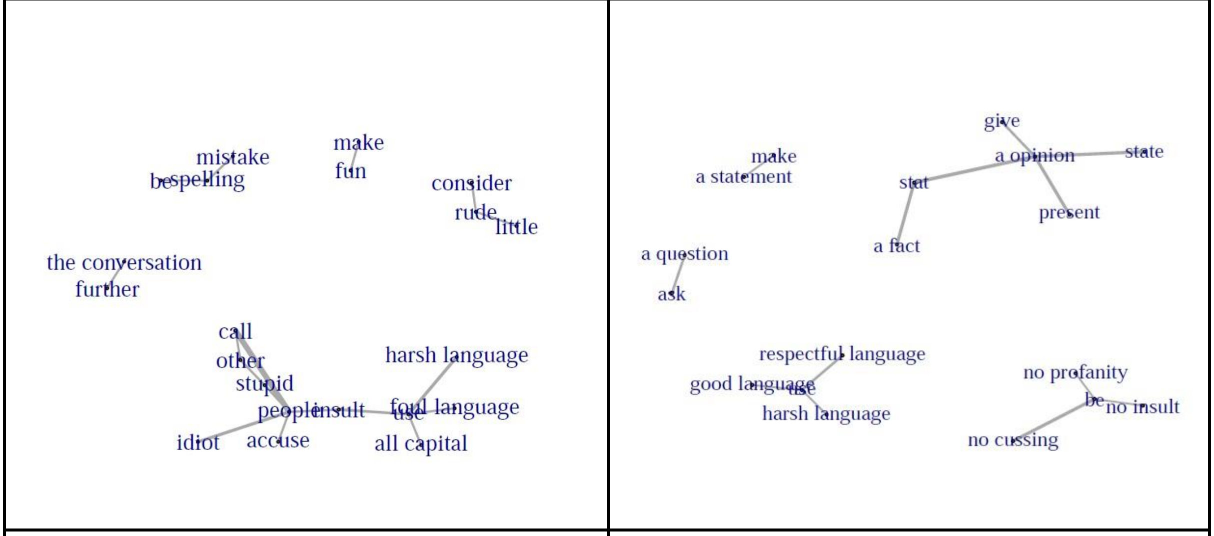
The analysis reveals that (1) there's a misalignment between our qualitative analysis, conducted in the context of online comments and current 'literature' that defined incivility at large. This finding implies detecting incivility in practice is highly content and context-dependent; (2) the crowd data has a high level of agreement with the professional qualitative analysis, indicating that our approach can identify and capture the main types of incivility.

Overall, we hope that implementing such collaborative crowdsourcing and NLP approach will lead to clearer, constantly up-to-date guidelines that take the different contexts and domains into consideration and can help lower the human toll in the stressful work of content moderation.

Table 1: types of incivility

Literature:	Name-calling, Aspersion, Vulgarity, Lying, Pejorative for speech.
Qualitative Analysis:	Name-calling, Aspersion: Insulting/ Cursing, Shouting, No reasoned argument, Condescending, Bad grammar.
Our approach (crowd + NLP):	Aspersion: Cursing / Rudeness / Insult, Comment written in all capital letters ("feels like shouting"), Comment that does not add information to the discussion, Condescending (make fun), Bad grammar.

Figure 1: Word-embedding on the crowd's content meta-data



Incivility	Civility

3 References

Coe, K., Kenski, K., & Rains, S. A. (2014). Online and uncivil? Patterns and determinants of incivility in newspaper website comments. Journal of Communication, 64(4), 658-679.

Duarte, N., Llanso, E. and Loup, A. (2018). Mixed Messages? The Limits of Automated Social Media Content Analysis. In FAT (p. 106).

Kenski, K., Coe, K., & Rains, S. A. (2017). Perceptions of uncivil discourse online: An examination of types and predictors. Communication Research, 0093650217699933.

Langvardt, K. (2017). Regulating online content moderation. Geo. LJ, 106, 1353.

Link, D., Hellingrath, B., & Ling, J. (2016). A Human-is-the-Loop Approach for SemiAutomated Content Moderation. In ISCRAM.

Priss, U. (2006). Formal concept analysis in information science. Annual review of information science and technology, 40(1), 521-543.

Roberts, S. T. (2019). Behind the screen: Content moderation in the shadows of social media. Yale University Press.

THIRD-PARTY INDUCED CYBERSECURITY INCIDENTS: WHO PAYS THE PRICE

Research in Progress

Michel Benaroch, Syracuse University

Keywords: Cyber incident, Third-party service provider, IT outsourcing, Event study, Market Reaction, Economics of cybersecurity.

1 Introduction

Cybersecurity incidents originating with IT vendors and third-party service providers arrive regularly. For example, in July 2019 data on over 100 million of CapitalOne's customers hosted on Amazon's cloud was hacked by a former Amazon Cloud Services employee, and in May 2019 Salesforce had a multi-hour cloud meltdown due to a database blunder that granted users access to all data. Many other examples involve smaller firms. One study suggests that 56% of surveyed firms experienced a data breach caused by one of their vendors (Ponemon 2018). Other studies report numbers closer to 30% in financial services and healthcare (Benaroch and Chernobai 2015, Vasishta et al. 2018).

IT service providers contribute to cybersecurity risk because of their complex IT solutions and lack of adequate security controls. They add to and exacerbate their client firms' cyber risk exposure. Client firms may assume that their IT service providers are responsible for cyber risk, but in effect they cannot outsource their cybersecurity liability and regulatory responsibility. So, more client firms are demanding assurance over the security of their IT providers' services and more IT providers are obtaining information security certifications (e.g., ISO27002, SOCI/II, NIST800-53, and CSA). Such certifications are costly and renewed yearly, but their effect on cyber risk exposure remains unknown.

2 Objectives

Our research objective is threefold: (1) examine the prevalence of cyber incidents induced by IT service providers, compared to other cyber incidents? (2) assess whether these cyber incidents harm firms more or less than other cyber incidents? And, if so, (3) evaluate if the answer is different if IT service providers have had information security certification prior to experiencing incidents?

3 Theoretical Underpinning

The economic harm cyber incidents inflict on firms is easier to measure for publicly traded firms. The harm includes damage to firm reputation and customer trust, monetary losses, recovery costs, and so on. As these factors can lower firms' future profitability, public firms that suffer cyber incidents see their stock price drop. On this basis, tenths of studies use the event study methodology to measure the economic impact of cyber incidents. Third-party induced incidents involve a client firm and an IT service provider, raising the question: who suffers more of the economic harm? Applying the event study methodology on different subsamples of cyber incidents can help answer this question.

Questions concerning the effect of information security certification, in addition, require reliance on the notion of *market-based trust* (Benaroch 2020) Many argue that client firms must simply develop *trust* in their partners. Market-based trust is anchored in market mechanisms for establishing the reputation of IT service providers based on an independent certification of their information and security controls. Reputation, or the fear of its loss, constrains opportunistic behaviour and exemplifies how markets self-regulate. Service providers can hire trusted third parties to evaluate and certify their quality (e.g., Dun & Bradstreet offers dependable credit information on businesses of all sizes). Evaluation standards are often set by regulators, especially when market-based reputation mechanisms and standards are slow to develop. Reputation mechanisms reward IT service providers who obtain information security certifications from independent, trusted agencies. Indeed, firms that obtained certifications, such as ISO27002, see their stock prices rise (e.g., Dean et al. 2019; Malliouris and Simpson 2019). Just as the market reacts positively and creates value for certified IT providers, so would IT providers that contribute to cybersecurity incidents indicating a weakness of their information security controls suffer punitive market reactions that destroy firm value. It is this dual market-based mechanism that should hold IT service providers accountable for cybersecurity risk. This is the basis for the way we address the question: does the stock market react differently when IT service providers do, or do not, obtain information security certification?

4 Methods and Results

We answer our research questions using a unique sample of about 1,600 cybersecurity incidents that occurred in public firms between 2000 and 2018 (Table 1). We constructed the sample using: Verizon's VERIS data breaches database, Privacy Rights Clearinghouse Repository, IBM's AlgoFirst database of operational risk events, and Wikipedia entry on data breaches. Excluded from the sample are incidents: with no exact "announcement" date, in firms without stock data in the CRSP database, and from the six months following the August 2008 financial market meltdown. Data collection on firms with information security certified is ongoing.

	All incidents			3^{rd} party incidents				
	No. of Incidents	In public firms & w/ clear incident date	W/ stock data	No. of Incidents	In public firms & w/ clear incident date	W/ stock data	& 1^{st} party identified	& 3^{rd} party Identified
AlgoFirst	838	374		181	92			
RPC	9,362	763		197	52			
Wikipedia	521	118		154	31	**217**	**151**	**67**
Veris	8,009	319		266	20			
total	**18,730**	**1,574**	**1373**	**798**	**195**			

Table 1. Construction of Sample of Cyber Incidents

We use the event study methodology to estimate the economic impact cyber incidents on public firms. We use *Eventus®* and WRDS' Event Study tool to conduct a univariate analysis of the average cumulative abnormal returns (CAR) on stocks of several subsamples of firms. We also use multi-variate regression of CARs against various determinants, including a 1^{st}- and 3^{rd}-party indicator.

We have several preliminary results so far (see Appendix A). First, third-party induced cyber incidents account, year to year, for less than 15% of all incidents. Second, for all non- 3^{rd}-party induced pairs, average CARs are statistically significant and around -0.5% in the first 10 days, -1.2% after 80 days, and -1.9% after 200 days. Third, firms involved in 3^{rd}-party induced incidents, client firms and 3^{rd}-party firms, have significant average CARs around -0.25% in the first 10 days, -1.4% after 30 days, and above 0% after 100 days. Lastly, while average CARs for 1^{st}-party firms are not negative, for 3^{rd}-party firms they consistently negative for 250 days and drop to -4.5% within 70 days.

5 Conclusion

This research seeks to make two contributions. It offers an empirical, fact-based assessment of the scope of third-party induced cyber incident and their impact on firms. It also evaluates the economic value of information security certification to client firms and to IT service providers. Both contributions are important because reliance on IT service providers is likely to continue and grow. The latter contribution is vital given known obstacles to reliance on market-based trust, including the slow adoption by IT service providers, the explosion of cybersecurity regulations, and the concurrent proliferation of cybersecurity certifications designed to meet different information security criteria.

6 References

Benaroch M. 2020. "Cyber Security and IT Outsourcing – Challenges and Emerging Realities," in R. Hirschheim et al. (Ed.), *IS Outsourcing: The Era of Digital Transformation* (5^{th} Edition), Springer.

Benaroch M. and Chernobai A. 2015 "Linking Operational IT Failures to IT Control Weaknesses," *Proceedings of AMCIS'2015*, Puerto Rico.

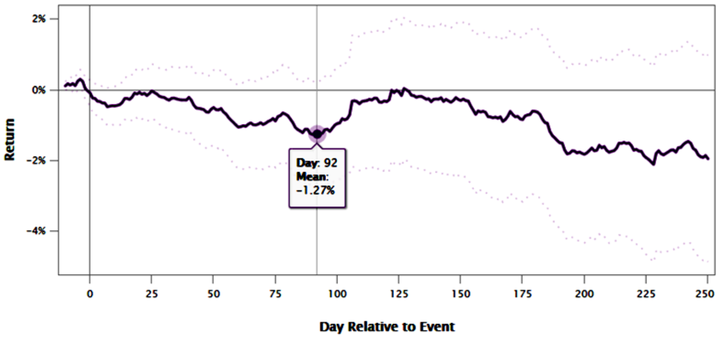
Deane K., Goldberg M., Rakes R. & Rees P. 2019. "The effect of information security certification announcements on the market value of the firm," *Information Technology and Management*, published online.

Malliouris D.D. & Simpson A.C. 2019. "The stock market impact of information security investments: The case of security standards," *Workshop on Economics of Information Security*, Boston, MA.

Ponemon, 2018. "Data Risk in the Third-Party Ecosystem," Ponemon Institute.

Vasishta V., Gupta M., Misra K., Mulgund P. & Sharman R. 2018 "Optimizing Cybersecurity Program - Evidence from Data Breaches in Healthcare," *13th Annual Symposium on Info. Assurance* (ASIA'18).

Appendix A: Some Preliminary Results

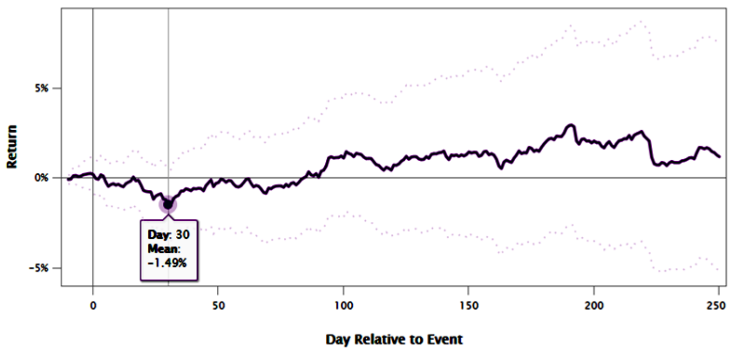


Market Model Abnormal Returns, Standard & Poor's 500 Composite Index

Days	N	Mean Cumulative Abnormal Return	Precision Weighted CAAR	Positive: Negative	Uncorrected Patell Z	Portfolio Time-Series (CDA) t
(-1,+10)	1156	-0.39%	-0.42%	543:613	-2.994**	-1.676*
(-1,+20)	1156	-0.04%	-0.34%	559:597	-1.796*	-0.116
(-1,+30)	1156	-0.16%	-0.53%	565:591	-2.332**	-0.422
(-1,+100)	1156	-0.73%	-1.49%	524:632<<	-3.640***	-1.065
(-1,+150)	1156	-0.51%	-1.58%	532:624<	-3.144***	-0.607
(-1,+250)	1156	-1.82%	-2.64%	531:625<	-4.080***	-1.696*

***p<0.001 **p<0.01, *p<0.05, $p<0.1

Figure A.1: mean CAs for non-3rd party incidents

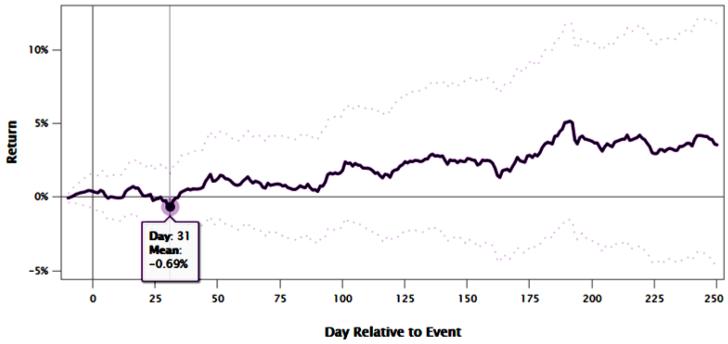


Market Model Abnormal Returns, Standard & Poor's 500 Composite Index

Days	N	Mean Cumulative Abnormal Return	Precision Weighted CAAR	Positive: Negative	Uncorrected Patell Z	Portfolio Time-Series (CDA) t
(-1,+10)	217	-0.05%	-0.53%	101:116	-1.838*	-0.114
(-1,+20)	217	-0.25%	-0.91%	96:121(	-2.315*	-0.427
(-1,+50)	217	-0.72%	-1.69%	98:119	-2.762**	-0.786
(-1,+100)	217	-0.39%	-0.75%	111:106	-0.797	-0.309
(-1,+150)	217	-0.32%	-1.01%	110:107	-0.893	-0.205
(-1,+250)	217	2.41%	0.91%	118:99)	0.790	1.200

***p<0.001 **p<0.01, *p<0.05, $p<0.1

Figure A.2: mean CARs for firms in 3rd-party incidents

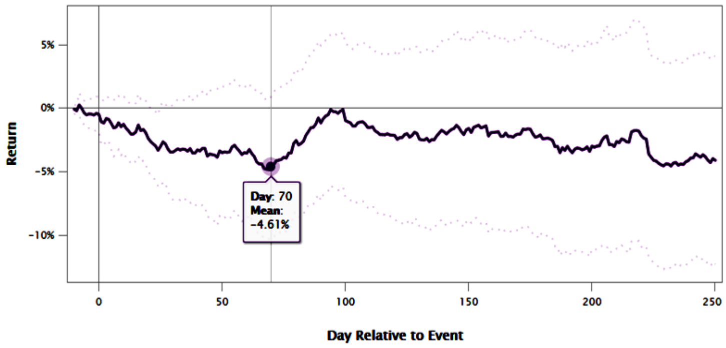


Market Model Abnormal Returns, Standard & Poor's 500 Composite Index

Days	N	Mean Cumulative Abnormal Return	Precision Weighted CAAR	Positive: Negative	Uncorrected Patell Z	Portfolio Time-Series (CDA) t
(-1,+10)	151	-0.06%	-0.49%	67:84	-1.392$	-0.123
(-1,+20)	151	0.07%	-0.64%	71:80	-1.336$	0.099
(-1,+30)	151	-0.67%	-1.81%	65:86(	-3.134***	-0.799
(-1,+100)	151	-0.90%	-1.08%	75:76	-1.026	-0.606
(-1,+150)	151	-0.90%	-1.52%	75:76	-1.207	-0.496
(-1,+250)	151	2.15%	0.84%	77:74	0.491	0.919

***p<0.001 **p<0.01, *p<0.05, $p<0.1

Figure A.3: mean CARs for 1st-parties in 3rd-party incidents



Market Model Abnormal Returns, Standard & Poor's 500 Composite Index

Days	N	Mean Cumulative Abnormal Return	Precision Weighted CAAR	Positive: Negative	Uncorrected Patell Z	Portfolio Time-Series (CDA) t
(-1,+10)	50	-0.68%	-1.28%	23:27	-2.101*	-0.916
(-1,+20)	50	-2.02%	-2.56%	19:31<	-3.101***	-2.016*
(-1,+50)	50	-2.41%	-2.82%	21:29	-2.184*	-1.563$
(-1,+100)	50	0.33%	-0.16%	26:24	0.002	0.154
(-1,+150)	50	-0.19%	-1.08%	23:27	-0.391	-0.071
(-1,+250)	50	0.53%	-0.90%	29:21	-0.154	0.157

***p<0.001 **p<0.01, *p<0.05, $p<0.1

Figure A.4: man CARs for 3rd parties in 3rd-oarty Incidents

DOES A THRESHOLD MODEL PREDICT INDIVIDUALS' TIME OF ADOPTION OF INNOVATIONS?

Complete Research

Mor Atlas, Arieh Gavious, Gilad Ravid, Gen Gurion Universiry

Keywords: Social network, Diffusion, Contagion, Adoption, Threshold models, Prediction accuracy.

1 Introduction

The diffusion-of-innovation models capture social influence that leads to adoption. Diffusion models generally fall into two categories: 1) threshold models that are deterministic in which people adopt an innovation once they reach their threshold (Burt, 1987, Granovetter, 1978, Rogers, 2010, Valente, 1996), and 2) probabilistic models that predict the probability of adopting an innovation based on personal and structural attributes. (Strang, Tuma, 1993, Kleinberg, 2007).

In this paper, we assess three commonly used threshold models for prediction at the micro-level to examine whether these threshold models represent the underlining process in adoption diffusion. We show that, surprisingly, these models fail to predict an individual time of adoption. Thus, the statistical agreement reported in the literature does not lead to the ability to predict individuals' time of adoption at the very least better than the naïve uniform model. Finally, we suggest that adding some elements based on human behavior in a social network such as an incubation period to the existing models improves their predictive abilities.

2 Research Methods

2.1 Models simulation and data

Three well established threshold models where chosen for examination: 1) Cohesion (Granovetter, 1978) where the threshold is compared to the neighbours adoption percentage. 2) Structural Equivalence (Burt, 1987) - where the threshold is compared to the structural equivalence adoption percentage. 3) Diffusion of innovations (Rogers, 2010) – a normal distribution for the time of adoption. We utilized three datasets that were important in establishing the basic threshold models and have various attributes: (a) medical innovations (Coleman, Katz & Menzel, 1966, Burt, 1987, Valente, 1996) (b) hybrid seed corn by Brazilian farmers (Rogers, 2010, Valente, 1996) and (c) Korean family planning (Granovetter, 1978, Valente, 1996). Simulation was performed for time of adoption prediction for each node in each dataset using each model's algorithm. Comparing these times with the real times of adoption allowed for evaluation of the different models. The uniform model is used as a baseline and should be a basic model to use when introducing a model for diffusion.

2.2 Model enhancements

we suggest three enhancements that are in line with human behavior and influences. First is the cohesion and random incubation period to account for the time need to cultivate prior to adoption. In this model we add to the cohesion threshold model incubation period drawn from an exponential distribution. Second is the linear combination of cohesion and structural equivalence models to account for both influences and third, cohesion with regression for a threshold that explains the differentiation between the thresholds using node attributes.

2.3 Evaluation method

We assert that a model's ability to predict the time of adoption would describe the adoption process sufficiently and contain the factors influencing the adoption decision. To examine accuracy of prediction against real data, the assessment should include both goodness-of-fit and relative error measures (Legates, McCabe, 1999). The four evaluation measures we used follow the work of Bellocchi et al. and Legates et al.

(Bellocchi et al., 2002, Legates, McCabe, 1999): (i) Mean Absolute Error (MAE) (ii) Root Mean Square Error (RMSE), (iii) Nash-Sutcliffe Efficiency and (iv) Pearson's Correlation Coefficient.

3 Results

All simulated models produced a negative Nash-Sutcliffe Efficiency (EF). Surprisingly, the uniform baseline model yielded the best results, however, it did not reflect the macro behavior of the whole network. It is important to note that even while implementing fitting using the actual real calculated threshold (RCT), which is the percentage of the nodes' direct connection who have already adopted at the node time of adoption assigned as the nodes' threshold, the EF was also negative. These results eliminated any problem of overfitting that might occur using the same data, as we did in the simulation. All our enhancements improved the prediction results in all the measures across all datasets examined.

4 Discussion and Conclusions

We demonstrate a standard analytical approach of measuring a model ability to explain the diffusion process by assessing the micro-level time of adoption predictors in a social network. The results show that the examined known threshold models poorly performed as the sole predictors inferior to an average value. Additionally, the enhancement suggestions offer new insights into social influence and directions to the optimization of the accuracy in using threshold models for prediction.

Researchers and practitioners should use caution when utilizing threshold models. These models are indeed straightforward, however, as the results indicate, they do not express the real diffusion process. We showed that the correlations presented in the threshold models did not shed light on the underlying mechanisms, and the threshold models did not accord with the time of adoption sufficiently.

5 Reference

Bellocchi, G., Acutis, M., Fila, G. & Donatelli, M. 2002, "An indicator of solar radiation model performance based on a fuzzy expert system", Agronomy Journal, vol. 94, no. 6, pp. 1222-1233.

Burt, R.S. 1987, "Social contagion and innovation: Cohesion versus structural equivalence", American journal of Sociology, , pp. 1287-1335.

Coleman, J.S., Katz, E. & Menzel, H. 1966, Medical innovation: A diffusion study, Bobbs-Merrill Indianapolis.

Granovetter, M. 1978, "Threshold Models of Collective Behavior", American Journal of Sociology, vol. 83, no. 6, pp. 1420-1443.

Kleinberg, J. 2007, "Cascading behavior in networks: Algorithmic and economic issues", Algorithmic game theory, vol. 24, pp. 613-632.

Legates, D.R. & McCabe, G.J. 1999, "Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation", Water Resources Research, vol. 35, no. 1, pp. 233-241.

Rogers, E.M. 2010, Diffusion of innovations, Simon and Schuster.

Strang, D. & Tuma, N.B. 1993, "Spatial and Temporal Heterogeneity in Diffusion", American Journal of Sociology, vol. 99, no. 3, pp. 614-639.

Valente, T.W. 1996, "Social network thresholds in the diffusion of innovations", Social networks, vol. 18, no. 1, pp. 69-89.

Would you believe it if an AI told you that 2 + 2 is 5? Conformity to algorithmic recommendations

Yotam Liel and Lior Zalmanson – Tel Aviv University

In a world where humans and AI agents' collaborative work is becoming increasingly common, an essential factor for the success of such collaboration lies in humans' ability to call out algorithms biases and errors. A recent line of research highlights the existence of systematic biases hidden behind machine learning models (d'Alessandro, O'Neil, & LaGatta, 2017). Biased and erroneous algorithmic recommendations have substantial impacts on the subjects involved and, in some scenarios, even critical effects on the lives of the subjects of such recommendations (can cause a denial of access to loans, rejection of job application, profiling suspects in a crime, etc.). Until this day, past research in IS has mostly focused on the advantages of algorithmic recommendations. There have been many studies that showed that forecasts made by algorithms are superior to human judgment in terms of accuracy (Dawes, Faust, & Meehl, 1989; Yeomans, Shah, Mullainathan, & Kleinberg, 2017) and that in recent years, humans have begun to exhibit an increased willingness to accept and adopt algorithmic recommendations (Dietvorst, Simmons, & Massey, 2018; Gunaratne, Zalmanson, & Nov, 2018; Logg, Minson, & Moore, 2019). Thus, in the face of algorithmic recommendation, studying the issue of exercising caution, defined as the ability to balance reliance on algorithmic recommendations and critical judgment towards them holds immense importance and potential social gain.

As the first step in this direction, this research seeks to ask a broad question: Will workers relying on recommendations be alert and ready to detect errors in the algorithmic outputs?

The statistical challenge we faced was in the design of an environment that will allow us to separate and compare the volition of workers' cognition (in terms of confident choices) to the persuasion power of the algorithmic recommendations. To do so in an observational study will pose many identification challenges, and thus, we have conducted a series of controlled lab experiments. A field study was tested as a possible alternative, but due to the deceptive aspect of the study (supplying erroneous recommendations), it was disqualified.

Instead, we have focused on gig-economy platform workers (MTurk) and found in a series of online lab experiments, that they tend to accept algorithmic advice without proper judgment, even when it's evidently erroneous. The design of the tests was greatly influenced by the classic works on conformity (Asch, 1956; Crutchfield, 1955). Drawing from these works, participants were asked to perform a short series of simple perceptual judgments tasks. The pressure from peers was replaced with AI recommendations, in the form of "algorithmic" aid that was provided to the treatment group – recommendations presented as the results of an image recognition algorithm analysis (see Fig. 1). In all of the experiments administered, the algorithm provided what are clearly wrong answers (as was evident by the almost unanimous answer given by workers when no algorithmic recommendation was given). The participants were recruited from MTurk, and the tasks were presented as standard image classification tasks. We choose this setup for two main reasons: (1) Image classification is a common task in MTurk with which many of the platform's users are familiar. By using this type of assignments we were able to simulate a typical work scenario for the participants; and (2) it allowed us to design tasks that very much resembled Asch's classic lines tasks and maintained their important traits – simple and easy perceptual judgment tasks with a clear difference between the correct and incorrect answers.

Our results indicate that algorithmic recommendations, even when erroneous, hold strong persuasive power. In our first study, participants followed the err algorithmic recommendations on average in 30% of the tasks, and 53% of the participants followed the err recommendations at least once. Moreover, participants in the control group choose these options in only 3.4% of the tasks, thus indicating it was easy not to mistake them for the correct answer. In our subsequent studies, we added mechanisms to ensure that participants spend sufficient time on each trial or are incentivized for top-quality work. In doing so, we were able to reduce the levels of conformity to an average of 14% (when ensured that they examined each problem for thirty seconds at least) and 19% (when presented with monetary incentives if they get all answers correctly). However, even with those mechanisms in place, we still observed substantial and significant levels of conformity.

We also observed that users who have exhibited conformity have reported higher confidence in the algorithm, hinting that they were consciously persuaded by the algorithm.

Our experiments are the first to show the persuasion of AI algorithms using Asch-like tasks, in which the algorithm's mistakes are supposedly blunt and easily detectable. This notion differentiates our work from prior works that tended to study acceptance and aversion of algorithmic recommendations toward questions that involved complicated cognitive predictions and uncertainty (Dietvorst, Simmons, & Massey, 2015; Gunaratne, Zalmanson, & Nov, 2018; Logg, Minson, & Moore, 2019) or subjective preferences (Adomavicius, Bockstedt, Curley, & Zhang, 2013).

This research contributes to our understanding of the persuasive power of algorithmic recommendations, and specifically to the (lack of) judgment towards suspicious and irregular suggestions produced by algorithms. By shedding light on these phenomena, we hope to increase the awareness of the risks of conforming and over-trusting recommendations provided by algorithms.

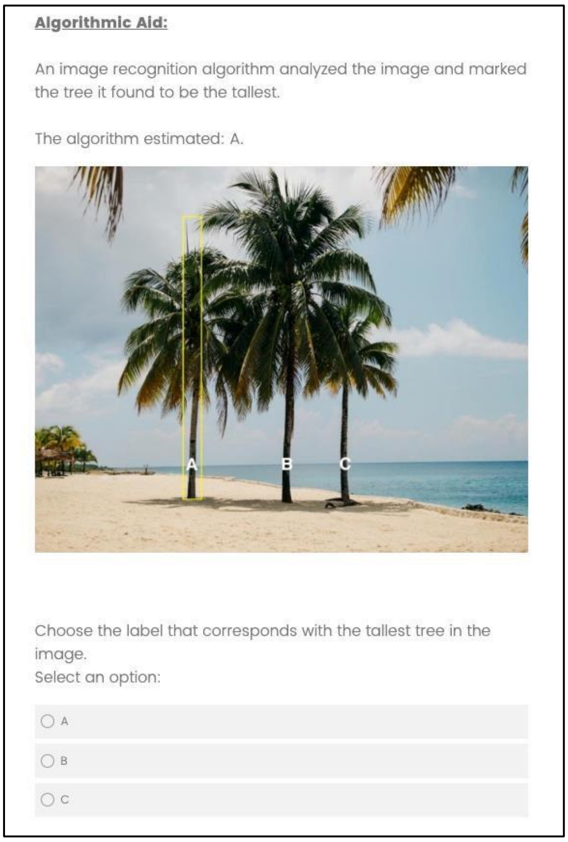


Fig. 1. One of the tasks as shown to participants in the treatment condition.

Bibliography

Adomavicius, G., Bockstedt, J. C., Curley, S. P., & Zhang, J. (2013). Do recommender systems manipulate consumer preferences? A study of anchoring effects. *Information Systems Research*, 956-975.

Asch, S. E. (1955). Opinions and social pressure . *Scientific American*, 31-35.

Asch, S. E. (1956). Studies of independence and conformity: I. A minority of one against a unanimous majority. *Psychological monographs: General and applied*, 1.

Crutchfield, R. S. (1955). Conformity and character. *American Psychologist*, 191.

d'Alessandro, B., O'Neil, C., & LaGatta, T. (2017). Conscientious classification: A data scientist's guide to discrimination-aware classification. *Big data*, 120-134.

Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science*, 16681674.

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm Aversion: People Erroneously Avoid Algorithms After Seeing Them Err. *Journal of Experimental Psychology: General*, 114.

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2018). Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them. *Management Science*, 11551170.

Gunaratne, J., Zalmanson, L., & Nov, O. (2018). The Persuasive Power of Algorithmic and Crowdsourced Advice. *Journal of Management Information Systems*, 1092-1120.

Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes*, 90-103.

Yeomans, M., Shah, A., Mullainathan, S., & Kleinberg, J. (2017). Making sense of recommendations. *Journal of Behavioral Decision Making*.

THE MOTIVATION FOR MOTIVATION THEORIES: A SYSTEMATIC LITERATURE REVIEW ON THE USE OF MOTIVATION THEORIES IN RE RESEARCH

Research in Progress

Naomi Unkelos-Shpigel, Ort Braude College of Engineering
Irit Hadar, University of Haifa

Keywords: Requirements engineering, motivation, systematic literature review

1 Research motivation and objective

The requirements engineering (RE) process requires knowledge sharing and collaboration among different stakeholders, who may vary in their backgrounds and technical knowledge. However, stakeholder collaboration is not easy to achieve, and is highly dependent on stakeholder motivation. The objective of this ongoing research is to understand how motivation has been researched in RE thus far, focusing on the role motivation theories have played in this context. To this end, we are conducting a systematic literature review (SLR). This paper presents results obtained so far, based on preliminary analyses of the full set of papers retrieved in the SLR.

2 Research method and findings

The searched items included peer-reviewed journal, conference, and workshop papers. The publication period of the searched papers started at the start date specified in each digital library, and ended on December 2019. Five major digital libraries were searched: Association for Computing Machinery (ACM), Institute of Electrical and Electronics Engineers (IEEE), ScienceDirect, Springer Link, and Association for Information Systems (AIS)

In order to include all relevant papers, we defined the search query to be as general as possible, including the terms of interest, namely, *motivation* and *requirements engineering*. We did not include the search word *empirical* in the queries, despite our focus on empirical studies, since we found, while iterating on candidate queries, that papers often describe empirical research studies without using this exact term.

We applied inclusion/exclusion criteria to each of the papers as follows. The criteria for paper inclusion were that the paper: (a) reports an empirical study (or studies), (b) is in the context of specific RE task(s); (c) examines association(s) between motivation and any performed RE task; (d) describes research that involved humans (rather than automated procedures). We excluded papers reporting on motivation in educational contexts, since our research questions target motivation in the workplace, and motivational aspects in educational and work settings differ significantly. In some cases, where the tasks were designed in the context of an organizational task related to RE, we did include studies with university students' participation meant to simulate RE practitioners. The papers screening process is depicted in Figure 1. The final number of included papers was 327.

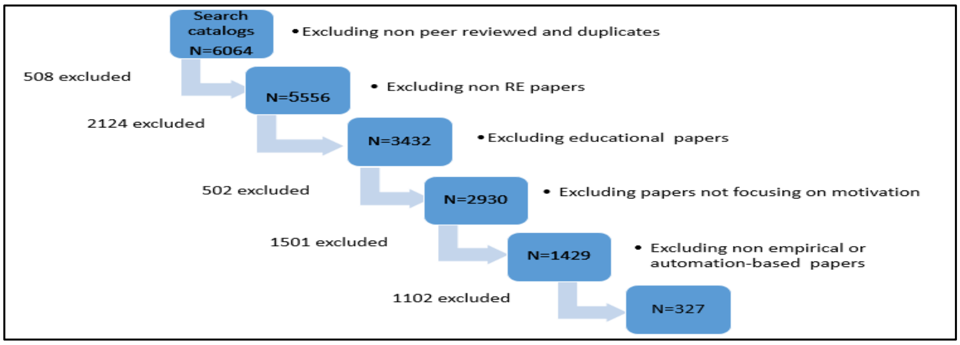


Figure 2. Papers screening process

In order to understand the nature of the research studies described in the included papers, we distinguished between those employing qualitative, quantitative and mixed research approach, and further documented the specific research methods used. Qualitative research was the most frequently used approach, with case studies as the leading research method. This is a surprising outcome, given the fact that motivation theories typically rely on quantitative measures. In quantitative studies, experiments were mostly used. Figure 2 presents these data.

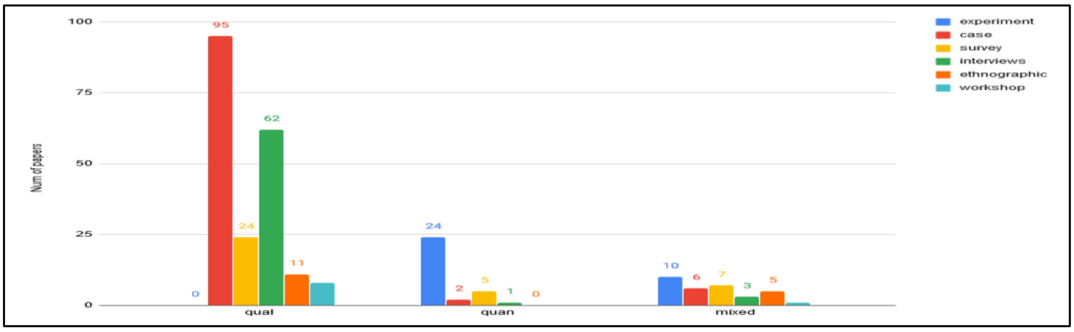


Figure 2. Papers' research approach and methods

An essential part of an SLR is quality assessment of the papers included [9]. We used a set of quality criteria for paper quality assessment, composed of the set of criteria proposed by Ambreen et al. (2018), and three new criteria, derived from our research objective, focusing on the use of theory in the analyzed papers. Due to space limitation, we will focus here on the theory use criteria only.

Of the analyzed papers, 78% do not rely on any motivational or cognitive theory. This is a concerning outcome, as these theories may provide reliable metrics for measuring motivation of participants for performing their tasks, possible explanations for cognitive, motivational and behavioral phenomena identified in the research, and validated strategies for improving motivation. Of the papers that did rely on theories, we looked into the theories used, many of which found to appear only once in our SLR. The most frequently used theories were (no. of papers in brackets): Activity Theory (7), Self-Determination Theory (8), Distributed Cognition (4), and Flow Theory (2). Next, we checked for each paper its research approach: whether it took a decriptive appraoch, describing and analyzing a given situation, or a design science approach, identifying a problem and proposing a solution. Of the 193 papers classified as descriptive, 41 papers referred to theories; of the 134 papers classified as design sceince, 32 leveraged on a theory. Figure 3 presents these findings.

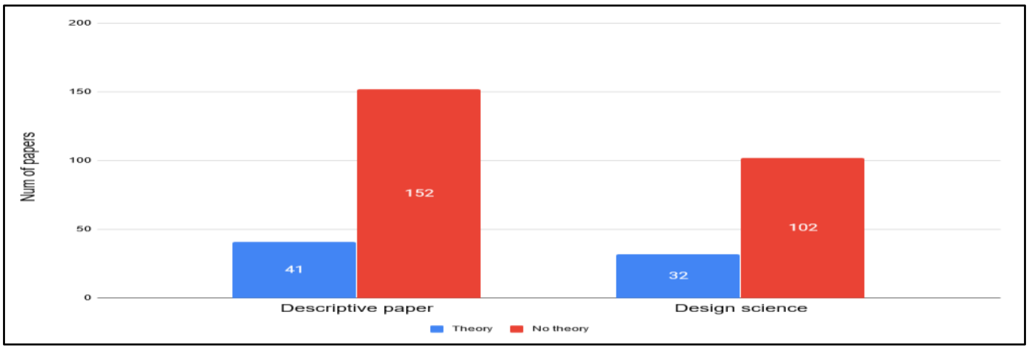


Figure 3. Papers types and theories

Our working assumptions is that using theories has the potential of increasing research rigor and validity. As this research progresses, we further plan to test whether correlations can be found between the use of motivation theories and other research quality indicators, to further substantiate and identify what benefits, if any, stem from the use of motivation theories in RE research. However, our preliminary findings already uncover the need for a more rigorous research on motivation in RE, taking advantage of the vast body of knowledge of motivation research.

3 References

Ambreen T., Ikram N., Usman M., Niazi M. 2018. "Empirical research in requirements engineering: trends and opportunities". *Requirements Engineering*, (23:1), pp.63-89.

Csikszentmihalyi M., 1997. *Flow and the Psychology of Discovery and Invention*. Harper Perennial, New York.

Deci E. L., and Ryan R.M. 2000. "The" what" and" why" of goal pursuits: Human needs and the self-determination of behaviour". *Psychological inquiry*, (11 :4), pp. 227-268.

Engeström Y.1999. "Activity theory and individual and social transformation". *Perspectives on Activity Theory*,(19:38).

Kitchenham, B., 2015. *Procedures for performing systematic reviews.* Department of Computer Science, Keele University and National ICT, Australia Ltd.

Keynote Talk:

Trading Privacy for the Greater Social Good: How Did America React During COVID-19?

Anindya Ghose, NYU

Digital contact tracing and analyses of social distancing from smartphone location data are two prime examples of non-therapeutic interventions used in many countries to potentially mitigate the impact of the COVID-19 pandemic. While many understand the importance of trading personal privacy for the public good, others have been alarmed at the potential for surveillance via measures enabled through location tracking on smartphones. In this research, we analyze a massive yet atomic individual level location data with over 11 billion records from 10 'Blue' (Democrat) and 10 'Red' (Republican) cities in the U.S. to present some of the first evidence of how Americans responded to the increasing privacy concern that government authorities, private sector, and public health experts may have to use individual-level location data to track the COVID-19 spread. First, we find that there is a significant decreasing trend of opt-out of location sharing with mobile apps in the U.S. While areas with more Democrats are more privacy-concerned than areas with more Republicans before the advent of the COVID-19 pandemic, there is a significant decrease in the overall opt-out rates after COVID-19 and this effect is more salient amongst the Democrat than Republican cities. Second, people who practice social distancing (i.e., those who travel less and interact with fewer close contacts during the pandemic) are also less likely to opt-out, whereas the converse is true for people who practice less social-distancing. This relationship is also more salient amongst the Democrat than Republican cities. Third, the demographic analysis reveals that high-income population and males are more privacy-conscious and more likely to opt-out from location tracking, compared to low-income population and females. Overall, this research demonstrates that people in both Blue and Red cities generally formed a unified front in sacrificing personal privacy for the societal good during COVID-19, while simultaneously exhibiting a divergence in the extent of such a sacrifice along the lines of political divide, social distancing compliance, and demographics.