Right to Privacy in the Context of the Privacy Paradox and Data Collection Patterns: Exploratory Study of Polish Facebook Users

Agnieszka Rychwalska
University of Warsaw
a.rychwalska@uw.edu.pl

Magdalena Roszczyńska-Kurasińska
University of Warsaw
m.roszczynska@uw.edu.pl

Anna Domaradzka-Widła
University of Warsaw
anna.domaradzka@uw.edu.pl

Abstract

Dark patterns in online data gathering infringe on citizens' right to privacy and create a profound imbalance of power between citizens of digitalizing societies and institutional actors. In effect, even when users declare their concern about privacy, this attitude is often not reflected in actions. In the study reported here we found that Facebook users indeed perceive an imbalance of control over privacy: they feel it is their responsibility to protect it but at the same time they feel that they are less capable to fulfil this task than institutional actors. We also found that privacy concerns were a good predictor of actual effects of privacy protective behaviors, while at the same time they did not correlate with declarations about privacy protection, which suggest a need for careful measurement of such constructs. Our results are a first step towards a comprehensive research agenda on individuals' attitudes towards institutional privacy.

1. Introduction

Digitalization of social interactions inevitably leads to production of data. By accessing digital services, individuals satisfy a plethora of needs but at the same time leave a trail of data that describe them and their relations. These data are then aggregated, cross-linked and analyzed by private and public sector actors to develop personalized products and services [1]. By shaping and creating individual needs, such products and services often incentivize individuals to share more data [2], thus creating an imbalance of control over data between citizens and institutional actors.

Individuals are the main suppliers of data and at the same time draw little value from the data they supply. Moreover, they might face deleterious consequences of extensive data sharing when the data are used to manipulate their opinions and attitudes, leading them to suboptimal choices or decisions [1]. Such decisions, when aggregated over many individuals, may also have society-wide consequences. Manipulation of public opinions can lead to polarization in societies [3] and if it is carried out by actors with an agenda it can determine the political course of a whole nation [4] or undermine public health [5].

Such negative consequences are an effect of violations of "institutional privacy" understood as privacy from institutional [6] or economic surveillance [7]. However, individuals are often more apprehensive about "social privacy", i.e. sharing personal information with other individuals [6]. Studies investigating such social sharing have shown a discrepancy between declared concerns for privacy and the actions that contradict these concerns, i.e. the privacy paradox [8]. For example, people declaring privacy concerns in a survey were nevertheless willing to answer sensitive questions when subsequently interviewed by an anthropomorphic shopping agent [9]. Social media users professing concerns over strangers finding out their sensitive information (such as sexual orientation or partners' names) did in fact reveal such facts on their social media profiles [10].

The privacy calculus model aims at explaining this paradox by positing that in each interaction with digital services, individuals assess the risks and opportunities involved in sharing their data [11] and value convenience over privacy. For example, future consequences of privacy breaching disclosures are discounted compared to immediate gratification which might be given for such behavior [12].

Moreover, whilst digital service users are mostly aware that private companies gather their personal data for economic purposes, many still do not acknowledge the scope of this process [2] or the potential power that abusing individuals' institutional privacy provides [13]. Often, they are unaware of the ways their data are aggregated and used by platform operators [14] or confuse social privacy protection (i.e. availability of privacy options within platforms which limit visibility of personal information to specific others) with institutional privacy protection [6, 15]. Out of those who realize the scope of surveillance, many display "resigned pragmatism": helplessness in face of their limited capacities to keep their data private [16]. We

URI: https://hdl.handle.net/10125/79671
978-0-9981331-5-7

posit that this lack of awareness, concern or capabilities to understand the scope and consequences of abuse of institutional privacy is precisely what increases the imbalances of power between individuals and institutions, limiting citizens' right to privacy.

The status quo in institutional privacy protection is in line with particular interests of platform operators, who may actively increase the cost of privacy in the privacy calculus. For example, the prevalent solution to tackling privacy issues by service providers are privacy options and privacy policies that evolve to fulfil the minimal requirements imposed by local or global regulators (e.g. European GDPR policy). Yet, within the regulatory boundaries, platform operators can still make it hard for individuals to protect their privacy [15]. Often, by implementing privacy features they move the burden of protecting privacy from themselves to end users. Moreover, by creating an additional burden (e.g. the cost of finding privacy features or understanding privacy policies), such functionalities drain users' cognitive resources and reduce motivation, leading to cynicism or apathy [16]. For other users such practices can also create an illusion of agency and thus promote more data sharing [8], increasing the platform operators' advantage in data control.

The resultant imbalances in power over the main resource of the digital age – data – threatens to spur a self-reinforcing growth of societal inequalities: not only in economic terms but also in social and political rights [17, 18]. Thus, the wellbeing of citizens and the future of democratic societies depends on understanding how such imbalances can be mitigated. For this, a comprehensive research agenda that would investigate individuals' approach towards institutional privacy, as well as possible mitigating strategies, is needed. The study presented here is a step towards a better understanding on how individuals perceive the imbalance of power and agency, on how they can regain agency in privacy related behavior and, specifically, how they can be helped in shielding their privacy against institutional surveillance. Here, we focus on three aspects of privacy attitudes: a) the relation between privacy concerns and behaviors protecting institutional privacy; b) responsibilities and agency in institutional privacy protection of individuals and other actors; and c) risks and opportunities related to data gathering for various actors.

1.1. Research questions

Research on privacy paradox and privacy calculus suggest that users' declarations about privacy importance do not correspond to their actions, i.e. attitudes and behavior do not match, possibly due to lower importance of privacy than of convenience [11]. Although much research has been carried out to unpack the relation between privacy attitude and privacy behavior the results are inconsistent [19]. Many studies on the privacy paradox do not measure actual behaviors or their indicators, but declarations or intentions [19], or do not differentiate between social and institutional privacy [15]. As such, the relation between attitudes to privacy and behaviors protecting institutional privacy is largely unknown. Here we wanted to find out whether privacy concerns can predict a long term, institutional privacy protecting behavior across interactions with multiple business actors. Thus we posed the following research question:

RQ 1. Are higher concerns for privacy related to better data protection from institutional actors?

The discovery of the gap between privacy attitudes and behavior of digital media users has contributed to the rise of "privacy-by-design" trend in digital services: solutions that proactively integrate privacy protecting principles into system's design [20]. This approach draws from empirical results on privacy protecting behavior which show that, e.g. priming privacy related issues increases chances of privacy protecting behaviors [21] or that changing the privacy calculus leads to better privacy choices [22]. This approach attempts at shifting the burden of privacy protection from individuals to platform designers. However, it is unclear how users perceive the responsibilities related with privacy protection: as a burden or as personal agency. Thus we asked:

RQ 2a. According to digital media users, what is the responsibility of individuals vs. institutional actors in protecting privacy?

RQ 2b. According to digital media users, what is the perceived agency of individuals vs. institutional actors in protecting privacy?

Finally, as indicated by results of privacy-by-design solutions, raising individuals' awareness of the scope of data gathering and data use by institutional actors may increase their agency in privacy protection [21]. This might be especially true if individuals were previously unaware how much data are gathered and how that data are used [23]. However, even users that are aware of abuses of institutional privacy may lack in motivation due to the experience of resignation and helplessness [16]. Thus, the effects of raising awareness on privacy protecting behaviors may depend on what are their initial perceptions of risks and opportunities involved in data gathering and how accurate is their initial awareness of the scope and effects of data gathering.

RQ 3a. How do digital media users perceive the risks and opportunities of data gathering practices?

RQ 3b. Does increasing individuals' awareness of data gathering practices of institutional actors change their perception of data gathering practices?

To answer these research questions we designed a study investigating privacy related attitudes, behaviors and practices of Facebook users in Poland. The study procedure included an additional exercise to help users find and manage what data Facebook gathers about them and at the same time allowed us to test the effect of raising awareness of data gathering practices on individuals' attitudes towards privacy protection.

2. Materials and methods

An online questionnaire was launched on the LimeSurvey platform. A post informing about the study with a link to the questionnaire was promoted on Facebook for nine days among adult Facebook users living in Poland and speaking Polish. The ad was visible only to users who logged in to Facebook on a computer. The study was not advertised to Facebook users logged in with their mobile devices because of the additional difficulties these users could face when switching between the Facebook app where users were to check their privacy settings and the questionnaire where they were to report their findings. Such differences in task difficulty could introduce uncontrolled bias in the sample.

To incentivize participation in the study we offered two gift vouchers to randomly chosen participants who would finish the main questionnaire and additional two vouchers to participants who would complete the additional task about their Facebook privacy settings. Informed consent was obtained from all participants. All procedures were approved by an Ethics Committee.

The questionnaire was composed of six sections measuring different aspects of privacy and management of online data: 1) privacy attitude and behavior, 2) attitude towards data acquisition by social media platforms, 3) General Data Protection Regulation, 4) privacy concerns, 5) social media habits, and 6) demographic questions. The participants who completed at least the third section of the questionnaire were invited to take part in an additional part of the questionnaire requiring inspection of Facebook privacy settings.

Privacy attitude was assessed by two questions: How important is it to you to protect privacy during offline activities? and How important is it to you to protect privacy in the online world? We assessed privacy behavior by two questions: How often do you speak about privacy with your friends? and How often do you adjust privacy settings on social platforms? Additionally, in this part of the questionnaire we asked to what extent different actors (e.g. users of social media, social media operators, governments, NGOs, and similar) should be obliged to take care of the privacy of digital media users; and to assess the chances of these actors to enforce protection of privacy for digital media users. Respondents were asked to provide answers to the above questions on the scale from 1 to 5 (1 indicating not at all/never/very low to 5 indicating very much/very often/very high).

Attitude towards data acquisition by social media operators was assessed by following questions: How beneficial is the acquisition of data by social media platforms to you/underage digital media users/senior media users/society as a whole/businesses/platform operators?; To what extent data acquisition by social media platforms has negative consequences for you/underage digital media users/senior media users/society as a whole/businesses/platform operators?; To what extent the accumulation of data about digital media users by social media platforms should be limited?; What do companies managing social media such as Facebook know about you?; How often do you happen to wonder before the publication of content on social media whether a post contains too much private content? Respondents were asked to mark their answers to the above questions on the scale from 1 to 5 (1 indicating not at all/nothing/never to 5 indicating very much/everything/very often). Additionally, we asked here to what extent the access to data regarding digital media users' activity should be changed for governments and private business owners. Answers were provided on a five-point scale (1 indicating "it should be reduced", 5 indicating "it should be significantly increased").

The attitude towards General Data Protection Regulation (GDPR) was measured by four questions: To what extent do you agree that GDRP has reduced the possibility of acquiring data on social media users by owners of social media platforms?; To what extent do you agree that you regained control over your data thanks to the GDRP?; To what extent do you agree that GDPR regulates the possibility of acquiring data on social media users effectively enough?, How irritated do you feel when having to make a decision regarding the privacy settings each time you enter a new webpage? The answers were given on a scale from 1 to 5 (1 indicating not at all and 5 indicating very much).

Privacy concerns were measured by the Internet Privacy Concerns scale (IPC) [24]. The IPC scale is composed of 26 questions. To minimize the effect of earlier questions on the later answers (by asking privacy related questions we might have increased privacy awareness) we rotated the order of the questions in the IPC questionnaire and the position of the IPC questionnaire in the whole survey.

To measure social media habits we asked respondents how often they use different social media such as Facebook, Instagram, TikTok, Reddit or Twitter and at what age they had set up their profile on Facebook.

The demographics section of the questionnaire consisted of questions about age, gender, education level and size of the place of residence.

The additional part of the questionnaire - the Facebook privacy settings exercise - was composed of two tasks, each followed by a few questions. In the first task, each respondent was asked to log into her/his Facebook account and investigate how many companies deliver data to Facebook about the respondent's activity outside of Facebook To extract this information respondents had to, first, go to privacy settings, choose "Your information on Facebook", then "View or delete information about activity outside Facebook", and then click on "Activity outside Facebook". The list of companies provided there is password protected, therefore even logged in respondents had to enter their password and only then they could choose "Activity outside Facebook" again to see the list. This procedure was quite long and seemed to be complicated, therefore, we prepared detailed screenshots of what to do step by step. We asked respondents five questions related to this task: whether they managed to reach the information, what was the number of companies that transferred data on respondent's activities outside of Facebook to Facebook, whether they have been aware that other companies share such data with Facebook, how surprised they were by the amount of information that Facebook gathers about them, and to describe their reaction after seeing the list.

In the second task, we asked respondents to investigate the categories that Facebook assigned to them based on their activity. A detailed instruction with screenshots was presented. Respondents were asked to go to their privacy settings, choose "Privacy-shortcuts", go to "Advertising preferences", "Choose your advertising settings", "Categories used to reach you", and finally "Categories of interests". The description of the task was followed by eight questions. First we asked respondents whether they managed to reach the categories of interests that Facebook assigned to them, and if not - why. Those who managed to complete the task were asked to report the number of categories they found (a categorical variable, since Facebook does not provide the precise number), to assess how well the categories assigned by Facebook described them, to assess whether privacy management on Facebook was easy, whether they had known how to access privacy settings prior to the study and whether they were aware that they could reduce the accumulation of data about their activity by social media by these settings. Additionally, to check whether the conducted tasks influenced their attitude towards data acquisition by social media we repeated here two questions from the second section of the questionnaire: To what extent the accumulation of data about digital media users by social media platforms should be limited? What do companies managing social media such as Facebook know about you?

3. Results

3.1. Respondents

Online studies often suffer from a high attrition rate and our study was no exception: only 61% of respondents completed the whole survey. In order not to lose the answers of those who did not complete the whole survey (i.e. not to increase the self-selection bias) or those who did not provide answers to specific questions (some were not obligatory) we decided to use all the valid cases. In result, different analyses were run on different numbers of cases (yielding different Ns and dfs reported in the results).

Eighty nine out of 145 persons who agreed to take part in the study completed it in full. Four participants who declared to be under 18 years old were excluded from the analyses. Out of the respondents who completed the demographic questions 45 (53%) were women; the respondents' age varied from 18 to 87 years old (M = 41.82, SD = 18.48). The majority of participants lived in a big city (65%); 21% lived in a small city and the rest in a village. Half of the participants (54%) had completed higher education while 35% finished secondary school.

The respondents were frequent users of Facebook (M = 4.65, SD = .629 on a scale from 1 to 5) and occasional users of Instagram (M = 2.22, SD = 1.53). The least used social media in the studied sample were Twitter (M = 1.68, SD = 1.11), TikTok (M = 1.40, SD = 1), and Reddit (M = 1.26, SD = .74), which did not come as a surprise because the information about the study was promoted only on Facebook. The majority of respondents (67%) created an account on Facebook as adults, 27% did it as teenagers and only 6% as children.

3.2. Privacy attitude and behavior

First we analyzed whether our respondents displayed the privacy paradox. In the first analysis we determined if there was a difference in the need for privacy between offline and online activity. A Wilcoxon signed-rank test showed that of the 120

participants who answered both questions, 18 participants assessed privacy in offline activities as more important than on the internet, only 9 were of opposite opinion. The rest did not exhibit any differences in evaluation of the need for privacy protection in offline and online activities, $z = 111.5$, $p = .048$.

Next, we analyzed the distributions of privacy attitudes. These were generally skewed: the distributions of answers to both questions about privacy importance as well as of the scores on the IPC scale indicated that most participants valued their privacy highly. This might be an effect of the sample gathering method – possibly, only those concerned with privacy agreed to fill in the survey. However, this result might also be due to the privacy paradox: the attitude of high concern for privacy might be considered socially desirable but might not translate into behaviors.

To verify this we tested for correlations between attitudes and behaviors. There was a statistically significant positive correlation between the two measures of privacy attitudes: importance of privacy protection online and in offline activities, $\tau_b = .71$, $p < .0005$ and between the importance of offline, and online privacy and the IPC scale score ($\tau_b(94) = .24$, $p = .004$, $\tau_b(94) = .22$, $p = .009$). However, none but one of these measures correlated significantly with the propensity to talk about privacy or to change privacy settings on Facebook, $p > .05$, confirming the privacy paradox. The propensity to talk about privacy correlated with attitude toward online privacy ($\tau_b(120) = .18$, $p = .02$). These two behavior related questions yield significantly correlated answers ($\tau_b(120) = .22$, $p = .004$).

As explained in the introduction, such paradox might not preclude a positive relation between privacy attitude and privacy protecting behaviors. To answer RQ1 – whether individuals' attitudes towards privacy impact their behavior online and its effects in terms of data traces left – we checked whether privacy attitudes could predict how much data about activities outside Facebook our respondents allowed the social media company to gather. The respondents reported this after being guided to access it in the Facebook privacy settings: they were asked to report how many companies sent data on the respondents' to Facebook.

Amongst others, Facebook can gather users' activities on other websites when they allow it to leave a third-party cookie, or use the Facebook login, share or like buttons within their services. To limit this data gathering users have to meticulously decline cookies, clean their cookie cache regularly and, if advanced, use dedicated websites that limit data flows between data collectors (e.g. youradchoices.com). Finally, they have the option to block Facebook from using such data for ad personalization by accessing their privacy options on the platform. All these actions require substantial effort on part of the users. Thus, we concluded that the number of companies that sent data to Facebook about a particular user might be an estimate of how much the user is engaged in privacy protecting behavior online.

For our predictor variable we chose the IPC score, as the other measures of privacy attitude were on scales with smaller range and thus had lower variance. The response variable was approximately log-linearly distributed; therefore we used its log transform in the model. The model was statistically significant and explained 31% of the variance in the number of companies sharing data on the user ($F(1,40) = 17.6$, $p < .001$): the higher the IPC score the less companies were reported as sending data about the user to Facebook (Fig. 1).

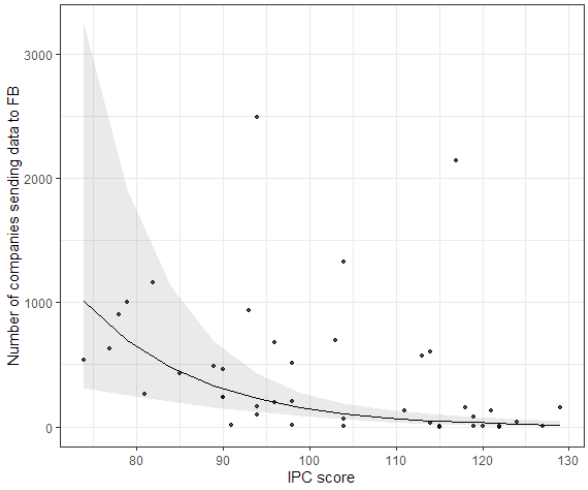

Fig. 1. Relationship between IPC score and the number of companies providing Facebook with data on users; line corresponds to predicted values, and the greyed area to 95% CI.

3.3. Privacy protection: responsibility and agency

To answer RQ2 - who should be responsible for protecting the privacy of online users and who has the highest chances of changing existing standards that enable privacy violation - we run a series of Friedman tests. First, we examined which actors the respondents felt should be most obliged to take care of the privacy of digital media users. Pairwise comparisons were performed with a Bonferroni correction for multiple comparisons. Perception of obligation was statistically significantly different between different actors, $\chi 2(3) =$

43.17, p < .001. Post hoc analysis revealed statistically significant differences in perception of responsibility between NGOs (Mdn = 2.15) and users (Mdn = 2.70) (p = .006); NGOs and social media platforms (Mdn = 2.83) (p < .001); government (Mdn = 2.33) and social media platforms (p = .016); but not between NGOs and government, government and users, nor between users and social media platforms.

There are two interesting takeaways from this analysis: users assign themselves the same level of responsibility for protecting privacy as to social media platforms, and the government is not perceived as responsible for ensuring the users' privacy to the same degree as social media platforms are. Although it is the government that has the legislative power to implement regulations, users seem not to expect the government to act on it. This attitude towards regulators and their effectiveness in curbing the imbalances in control of data might stem from the fact that when asked about the effectiveness of broad data regulatory acts (here: the GDPR) our respondents assessed them rather negatively. On average, they did not agree with a statements that GDPR limited the practice of accumulating data by social media platforms (M = 2.19, SD = 1.21), that GDPR regulates the practice of collecting data by social media platforms effectively enough (M = 1.96, SD = 1.01), nor did they feel that they regained control over their data (M = 1.99, SD = 1.12). Moreover, on average, respondents felt slightly irritated by the need to control privacy settings each time they entered a new site (M = 3.31, SD = 1.4). The assessment of the GDPR was available only to the respondents who were familiar with the regulation, but only one respondent declared that they did not know what GDPR was.

Second, a Friedman test was run to determine if there were differences in perception of agency in protecting privacy between different actors (RQ2b). Pairwise comparisons were performed with a Bonferroni correction for multiple comparisons. Perception of agency was statistically significantly different between actors, $\chi 2(3) = 44.04$, $p < .001$. Post hoc analysis revealed statistically significant differences in perception of agency between social media platforms (Mdn = 3.01) and government (Mdn = 2.56) (p = .039), social media platforms and users themselves (Mdn = 2.15) ($p < .001$), and social media platforms and NGOs (Mdn = 2.28) ($p < .001$), but not between users and government, users and NGOs, nor between NGOs and government. Social media platforms are perceived as the most capable actor in changing the existing standards in data collection practice. The users themselves, governments and NGOs are assessed as less powerful actors in this system.

3.4. Risks and opportunities of massive data collection

To answer RQ3a on individuals' perceptions of risks and opportunities of data gathering we analyzed answers to questions about negative and positive consequences of data gathering for various actors: the users themselves, underage digital media users, senior media users, society as a whole as well as businesses and platform operators. By performing a principal component analysis (PCA) on the question about benefits we found that media users divide the beneficiaries of data gathering into two groups: the first consists of business and platform operators (i.e. the private sector) and the second of all other actors. The overall Kaiser–Meyer–Olkin (KMO) measure was 0.74, with individual KMO measures not lower than 0.5. Bartlett's Test of Sphericity was statistically significant ($p < .0005$), indicating that the data was factorizable. The two revealed components had eigenvalues greater than one and explained 47.33%, 25.35% of the total variance, respectively. A Varimax orthogonal rotation was employed.

We obtained the same division of actors: private business (business and social media platforms) and society (e.g., elderly, minors, the respondent herself, society as whole) after running a PCA on assessment of risks related to gathering of data by social media platforms. The overall Kaiser–Meyer–Olkin (KMO) measure was 0.78, with individual KMO measures not lower than 0.5. Bartlett's Test of Sphericity was significant ($p < .0005$). The two components had eigenvalues greater than one and explained 56.96%, 21.32% of the total variance, respectively. A Varimax orthogonal rotation was employed.

We then compared these two groups of actors as determined by the PCA with regard to how their benefits and risks from data gathering were perceived by respondents. We found that businesses accrue higher benefits and positive consequences than negative consequences and risks: on a scale from 1 ("there are no benefits"/"there are no risks") to 5 ("there are great benefits"/"there are great risks") their benefits were assessed at 4.16 on average (SD = .09) while their negative consequences at 2.41 on average (SD = .18), $F(1,108) = 101.96$, $p < .0001$. On the other hand the non-business users of social media get significantly less benefits from data gathering (M = 1.99, SD = .09) than negative consequences (M = 3.9, SD = .1), $F(1,108) = 171.82$, $p < .0001$.

To better understand the consequences of these perceptions of data gathering we also analyzed questions that tackled the effects of this process: whether the respondents thought that data gatherers (in our case Facebook) knew them well (from 1 - "they

don't know anything about me" to 5 - "they know everything about me") and whether they thought that data gathering should be limited (from 1 - "it shouldn't be limited at all" to 5 - "it should be strongly limited"). We found that answers to both questions reflected the perception of negative consequences of data gathering: the respondents assessed that platform owners knew them well (M = 3.79, SD = 1.13) and that data gathering should be limited (M = 3.9, SD = 1.13).

We also explored if the preference for limited data gathering depended on respondents' privacy attitudes and the perceived negative consequences for social actors (the respondents themselves, minors, senior media users and society as a whole) as identified by the PCA analysis. The two predictors together explained 21% of the variance (Table 1) in preference for limited data gathering ($F(2,91) = 12.7$, $p < .001$).

Table 1. Predictors of the preference to limit data gathering

Predictor	Coefficient (Std. error)
IPC	.416*** (.006)
Perceived negative consequences for non-business users	.34** (.004)
Constant	.964
R^2	.286

Note: Linear regression coefficients; *** $p < .001$; ** $p < .005$

Finally, to answer RQ3b – whether increased awareness of data gathering practices changes digital media users' perceptions of such practices – we analyzed answers to a repeated question on preference to limit data gathering for those respondents who completed the Facebook task that required them to access their privacy options on Facebook and to report how much data Facebook gathered on them. We found that users increased their preference for limited data gathering after completing the Facebook task to an average of 4.25 (SD = .85), but the difference was not statistically significant ($p > .05$). This result might be due to the already high initial negative perception of data gathering by digital media platforms that our respondents displayed at the beginning of the survey.

We then analyzed whether they were surprised by the number of companies that sent data about them to Facebook and found that they were indeed astonished by the information Facebook accessed (M = 3.48, SD = 1.57 on a scale from 1 to 5). Moreover, most (75%) were not aware that other companies transferred data on their users to Facebook.

We also qualitatively analyzed the answers to an open question "What was your reaction when you saw how many companies send data on you to Facebook?". We received 53 answers to this question, mainly expressing a range of negative emotions.

Only 8 respondents (13%) expressed acceptance or declared that they did not care how much data from other companies is being sent to Facebook. Similarly, 9 out of 53 (17%) wrote that they were aware of the amount of data as well as the companies that send data to Facebook and that is why they used specific privacy settings to restrict access to their data. In case of those respondents, the reaction was calm as they either made sure earlier that no information was shared or they were fully conscious of the amount of sharing going on. These were usually people declaring that they actively use the privacy setting option, Facebook container or ToR to make sure they have control over their data privacy.

Most reported reactions to the information about the number of companies sharing respondent's data with Facebook were shock (32%) and anger (21%). Some respondents wrote that they felt cheated or used and had no idea about the amount of data Facebook acquires outside its platform. Others remarked that they thought collecting data in this way was unethical and that they were disgusted that some trusted institutions or companies had sent their data to Facebook. The other responses expressed feelings such as surprise (13%) and in three cases respondents openly declared fear as their dominant reaction.

Finally, we analyzed how the perception of what platform operators knew about them changed after the respondents completed the Facebook task. We found that there was no significant difference between the assessments before and after the task ($p > .05$). However, we found that those users who perceived the categories (which Facebook assigned to them based on gathered data) to fit them well, were also more likely to assess platform operators' knowledge about them as high in the second assessment ($\tau_b = .34$, $p < .005$).

4. Summary of results

We planned the study reported above as a first step towards a comprehensive research agenda to better understand digital media users' awareness and attitudes towards institutional privacy and the imbalances of power over data between citizens and institutional actors. We were prompted by inconsistency in results and explanations of the privacy paradox as well as limited research specifically on abuses of institutional privacy [19]: their perception by and impacts on citizens of digitalizing societies.

Our study confirmed the existence of the privacy paradox among Polish Facebook users. Collected data allowed us to investigate the reasons behind this discrepancy and check if it stems from the fact that users do not value their online privacy enough to put in effort to guard it, or if they have a sense of protection or anonymity in their online activities and therefore do not feel the need to intervene. As our study proved, people value their privacy, and they are often not comfortable with how data is aggregated and commodified behind their backs. Moreover, we found that even though most of our respondents scored very high on privacy concerns, they still systematically differed in the effectiveness of their privacy protection behaviors: those who had highest privacy concerns also had fewer companies sending their data on outside Facebook activities to Facebook.

Our second goal was to see who, in the mind of our respondents, should be the main actor responsible for securing privacy for users and which actor is the most capable of doing it. One surprising result was that public administration was not perceived as an actor sufficiently capable of ensuring institutional privacy. Our respondents were under no illusion that legal solutions, for example related to GDPR, could protect their privacy in a meaningful way. Our respondents also indicated that individual users were as much responsible for protecting privacy as online platforms, even if they felt that the agency of individuals in protecting privacy was much lower than that of social platforms. Thus, they seemed to be aware of the imbalances of control over data. They also seem to be aware of disproportions between the risks and gains of data gathering that they and business actors experience. In the eyes of our respondents benefits are much higher than risks on side of business actors, while risks exceed the gains on the side of individual users.

We also wanted to verify if the sense of agency in privacy protection can be stimulated among internet users when they are confronted with information on how much data is being collected on their. We found that the overwhelming majority of our respondents expressed negative emotions - such as shock, disgust or fear - when confronted with the scope of data accumulation on social media.

In the course of our study, we could also observe some signs of increased individuals' awareness concerning the scope of data gathering and use by other commercial actors, which may increase their agency in privacy protection: e.g. an interest in updating privacy settings and tools for increased privacy protection as a result of our exercise, as expressed in open questions in the survey.

5. Limitations

Our study was carried out on a small, convenience sample of Facebook users and while it may have satisfactory ecological validity it also suffers from several limitations. First, the studied sample consisted of self-selected individuals who were sufficiently interested in privacy issues to click on the post promoting the study. This bias is reflected in the measures of privacy attitude included in the survey: all are skewed towards high privacy concerns. To investigate privacy behaviors and attitudes more comprehensively – and the discrepancy between them – studies employing representative sample collection would be beneficial.

Second, there was a high attrition rate with only 61% participants completing the survey, and only 31% completing the additional task that required accessing Facebook privacy settings. This might have exacerbated the self-selection bias for those of the analyses presented in the paper that tackled the issue of increasing awareness of data gathering and of the resulting changes in perception of such practices. Again, a good strategy for carrying out such procedures in the future is to ensure a more representative sample, with a wider range of privacy attitudes. Moreover, the difficulties involved in guiding users towards specific – often hidden – privacy features could be mitigated by using a face-to-face study design (e.g. CAPI) that would include assistance from the interviewer.

Our sample was also limited to Facebook users speaking Polish and thus generalization of the results to other countries, cultures or regulatory contexts would be farfetched. However it is worth noting that similar results on users perceptions of the categories Facebook ascribes to its users were already reported for the US population [14].

Finally, it must be noted that the bias in the sample might have been aggravated by one additional, unforeseen factor: after nine days of promoting the study on Facebook, the promoting post was banned by the platform and the account that was used to post it was indefinitely blocked from the possibility to advertise posts. Thus, data gathering was prematurely halted, resulting in sample size below the desired level. Appealing the ban did not yield any results and the justification given did not provide any specific policy breaches. Thus, we are left to presume that the topic of the study together with the privacy settings task raised the privacy options awareness of study respondents to levels that the company did not feel comfortable with. The fact that such a simple procedure might be considered endangering by data gatherers serves as a case in point that raising social media users' privacy

awareness can help mitigate the imbalances in control over data in digitalizing societies.

6. Discussion and conclusions

The imbalance of power stemming from the big tech capacity to gather and analyze data is growing, which means that the extent of privacy invasion risks increases and should be closely monitored. The main challenge is that personal data – closely related to the privacy of the data providers – becomes a commodity that can be traded [25]. Commodification of privacy goes against the assumption that, similarly to other human rights, the right to privacy is inalienable and non-negotiable [26]. As such, it should be key in developing not only better commercial platforms and services, but also human-centered smart policies [27].

Privacy requires that individuals have an area of autonomous presence and action, free from external intervention and control. The right to privacy means that individuals have the ability to determine who may have information about them and how that information is to be used [28]. This implies awareness and agency of users, as well as clear definition of responsibilities assigned to private and public institutions involved in data collection and its further processing.

However, in the reality of data-driven economy and smart policy-making, commodification of digital identities does not go hand in hand with increased awareness and agency of data subjects. To the contrary, data collection mechanisms are designed to maximize data sharing and therefore benefit the system providers at the cost of users' privacy. While users may be aware of the "imbalance of benefits" that their data sharing involves, they often remain reluctant to change their online behaviors.

One of the reasons may be that the exchange of personal data via digital means is perceived a "conscious compromise", in which users voluntarily surrender private information in return for digital access to specific information, goods and services [28]. This approach assumes that users can always refuse to share data and forego the use of digital platforms. However, in the context of modern technological dependency, such "logging out" [29] would mean foregoing important services and interactions, which is not a viable option for the majority of users.

In the era of Big Data, which is dominated by data-driven economy, the awareness of users is a core element in balancing between technological innovation and individuals' rights. Existing passive defense mechanisms of privacy and personal data protection seem to be both unrealistic and ineffective. A more realistic and effective approach towards protection of users' needs to involve an active empowerment of individuals in their personal data management [26]. This could be done through addressing privacy paradox at its core – and creating an easy alternative to the "resigned pragmatism" [16].

As our study shows, creating awareness is a first step in this process, as individuals do not seem to be fully alert about the amount of data sharing, as well as its potential commercial value, and tend to underestimate their ability to control their digital identity [26]. Instead they enter an unequal exchange in which their data becomes a currency to pay for "free" digital services or discounts for online products and services. Most users seem to not realize that at the moment of this exchange both the user data and profiling algorithms are turning into a legally protected private business asset. At the same time, respondents in our study with high level of privacy concern proved to be able to protect their institutional privacy.

One approach to counteract this process is to expand the control that people have over their digital data at the individual level. Another one is a collective control approach in which the collective power of data subjects can be exercised over data commons [29]. While our ambition was not to find a perfect solution, we hoped to shed some light on the users' perception of risks and the perceived agency of actors involved in the data cycle.

Meanwhile, the EU and other international bodies discuss different ways of protecting privacy without seriously hindering the digital economy and smart policies implementation. The EU GDPR's primary aim was to give individuals control over their personal data: ensuring that the end-users' consent will be freely given, specific, informed and active. However, as we tried to illustrate here, this approach has serious limitations on the individual level: people simply do not perceive government as the most capable agent when it comes to industrial privacy protection.

In this context, further studies concerning privacy-related behaviors and data collection patterns are crucial to ensure that the users' digital rights are respected and their agency is effectively strengthened.

7. Acknowledments

A. Rychwalska's contribution was supported by a grant no. 2017/27/B/HS6/00626 from Polish National Science Centre; and A. Domaradzka-Widła's contribution was supported by grant no. 2018/30/E/HS6/00379 from Polish National Science Centre.

8. References

[1] Rychwalska, A., G. Goodell, and M. Roszczynska-Kurasinska, "Data Management for Platform-Mediated Public Services: Challenges and Best Practices", *Surveillance & Society 19*(1), 2021, pp. 22–36.

[2] Zuboff, S., *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*, Profile Books, 2019.

[3] Rychwalska, A., and M. Roszczynska-Kursinska, "Polarization on social media: when group dynamics leads to societal divides", (2018).

[4] Howard, P.N., and B. Kollanyi, *Bots, #Strongerin, and #Brexit: Computational Propaganda During the UK-EU Referendum*, Social Science Research Network, Rochester, NY, 2016.

[5] Jemielniak, D., and Y. Krempovych, "# AstraZeneca vaccine disinformation on Twitter", *medRxiv*, 2021, pp. 2021.04.08.21255107.

[6] Raynes-Goldie, K., "Aliases, creeping, and wall cleaning: Understanding privacy in the age of Facebook", *First Monday*, 2010.

[7] Marwick, A., and E. Hargittai, "Nothing to hide, nothing to lose? Incentives and disincentives to sharing information with institutions online", *Information, Communication & Society 22*(12), 2019, pp. 1697–1713.

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

[9] Spiekermann, S., J. Grossklags, and B. Berendt, "E-privacy in 2nd generation E-commerce: privacy preferences versus actual behavior", *Proceedings of the 3rd ACM conference on Electronic Commerce*, Association for Computing Machinery (2001), 38–47.

[10] Acquisti, A., and R. Gross, "Imagined Communities: Awareness, Information Sharing, and Privacy on the Facebook", *Privacy Enhancing Technologies*, Springer (2006), 36–58.

[11] Dinev, T., and P. Hart, "An Extended Privacy Calculus Model for E-Commerce Transactions", *Information Systems Research 17*(1), 2006, pp. 61–80.

[12] Acquisti, A., "Privacy in electronic commerce and the economics of immediate gratification", *Proceedings of the 5th ACM conference on Electronic commerce*, Association for Computing Machinery (2004), 21–29.

[13] Marlinspike, M., "Why 'I Have Nothing to Hide' Is the Wrong Way to Think About Surveillance", *Wired*, 2013. https://www.wired.com/2013/06/why-i-have-nothing-to-hide-is-the-wrong-way-to-think-about-surveillance/

[14] Hitlin, P., and L. Rainie, *Facebook Algorithms and Personal Data*, Pew Research Center, 2019.

[15] Sujon, Z., "The Triumph of Social Privacy: Understanding the Privacy Logics of Sharing Behaviors Across Social Media", *International Journal of Communication 12*(0), 2018, pp. 21.

[16] Hargittai, E., and A. Marwick, "'What Can I Really Do?' Explaining the Privacy Paradox with Online Apathy", *International Journal of Communication 10*(0), 2016, pp. 21.

[17] Wood, D.M., and T. Monahan, "Editorial: Platform Surveillance", *Surveillance & Society 17*(1/2), 2019, pp. 1–6.

[18] Gellman, B., and S. Adler-Bell, *The Disparate Impact of Surveillance*, The Century Foundation, 2017.

[19] Gerber, N., P. Gerber, and M. Volkamer, "Explaining the privacy paradox: A systematic review of literature investigating privacy attitude and behavior", *Computers & Security 77*, 2018, pp. 226–261.

[20] Cavoukian, A., "Privacy by design: The 7 foundational principles", *Information and privacy commissioner of Ontario, Canada 5*, 2009, pp. 12.

[21] Acquisti, A., I. Adjerid, R. Balebako, et al., "Nudges for Privacy and Security: Understanding and Assisting Users’ Choices Online", *ACM Computing Surveys 50*(3), 2017, pp. 44:1-44:41.

[22] Acquisti, A., L.K. John, and G. Loewenstein, "What Is Privacy Worth?", *The Journal of Legal Studies 42*(2), 2013, pp. 249–274.

[23] Stutzman, F.D., R. Gross, and A. Acquisti, *Silent Listeners: The Evolution of Privacy and Disclosure on Facebook*, Social Science Research Network, Rochester, NY, 2013.

[24] Hong, W., and J.Y.L. Thong, "Internet Privacy Concerns: An Integrated Conceptualization and Four Empirical Studies", *MIS Quarterly 37*(1), 2013, pp. 275–298.

[25] Davies, S.G., "Re-engineering the right to privacy: how privacy has been transformed from a right to a commodity", In *Technology and privacy: The new landscape*. Cambridge, MA: MIT Press, 1997, 143–165.

[26] Malgieri, G., and B. Custers, "Pricing privacy – the right to know the value of your personal data", *Computer Law & Security Review 34*(2), 2018, pp. 289–303.

[27] Foth, M., M. Brynskov, and T. Ojala, "Citizen's right to the digital city", *Berlin: Springer. doi 10*, 2015, pp. 978–981.

[28] Ronen, Y., "Big Brother's Little Helpers: The Right to Privacy and the Responsibility of Internet Service Providers", *Utrecht Journal of International and European Law 31*(80), 2015, pp. 72–86.

[29] Prainsack, B., "Logged out: Ownership, exclusion and public value in the digital data and information commons", *Big Data & Society 6*(1), 2019, pp. 2053951719829773.

[30] Dienlin, T., and S. Trepte, "Is the privacy paradox a relic of the past? An in-depth analysis of privacy attitudes and privacy behaviors", *European Journal of Social Psychology 45*(3), 2015, pp. 285–297.

[31] Baek, Y.M., "Solving the privacy paradox: A counter-argument experimental approach", *Computers in Human Behavior 38*, 2014, pp. 33–42.

[32] Human, S., and F. Cech, "A Human-Centric Perspective on Digital Consenting: The Case of GAFAM", *Human Centred Intelligent Systems*, Springer (2021), 139–159.

[33] Christakis, T., and K. Propp, "How Europe's Intelligence Services Aim to Avoid the EU's Highest Court—and What It Means for the United States", *Lawfare*, 2021.