Supporting students in making informed data sharing decisions: from comprehension to consenting

Marc Beardsley, Judit Martínez- Moreno, Patricia Santos, and Davinia Hernández-Leo
Department of Information and Communications Technologies
Universitat Pompeu Fabra
Barcelona, Spain
{marc.beardsley, judit.martinez, patricia.santos, davinia.hernandez-leo}@upf.edu

Abstract—Whether deciding to participate in a study or gain access to online applications, research suggests that individuals often fail to understand the conditions they agree to when giving consent. Limited comprehension related to the risks of big data and habituation may contribute to consent being given in a manner that counters the best interests of individuals (i.e., a privacy paradox). Who is responsible for helping students learn to make informed decisions regarding the sharing of their data? We conducted a quasi-experimental study involving 127 undergraduate students. We explored the effects on study enrolment of supporting students in making informed data sharing decisions. Prior to reading the consent form for a study involving the evaluation of a collaborative learning application, one group of participants was primed to think about data sharing risks and evaluate their participation in terms of a harm-benefit decision whereas the other group was not. The primed group scored lower on a consent form comprehension quiz yet had higher enrolment rates and were more likely to cite trust as reasoning for enrolling in the study.

Keywords—Informed Consent, Ethics, Big Data Analytics, Responsible Research

I. INTRODUCTION

Whether deciding to participate in research or gain access to an online application, studies on consent suggest that individuals often do not read [1, 2, 3] or fail to understand [4] the conditions they agree to before they give their consent. More specifically, studies on informed consent have consistently found that potential research participants do not demonstrate adequate comprehension of informed consent forms [5]. Although, most of these studies relate to biobehavioral research; similar results have been reported in educational studies [4]. A combination of limited comprehension related to risks such as data sharing risks [3]; and a form of habituation [6] may contribute to consent being given in a manner that counters the best interests of participants. This behavioural habituation in the context of data sharing decisions has been described as a privacy paradox [1, 7]. Habituation are a progressive decrease in response to a repeated stimulus and result from an individual’s accumulation of prior experiences [8]; thereby raising the possibility that experiences related to the giving of consent have contributed to increasing levels of apathy related to data sharing decisions.

A. Data Sharing Risks

When compared to biomedical data, the data collected by educational technologies is considered to be low risk and explicit harm-benefit analyses have not been expected as potential harms are deemed to be low in comparison to expected benefits. In the past, much of the data collected in educational contexts was limited to discrete exchanges of information serving efforts to better support student learning. However, ongoing exchanges of information between data collectors and students are increasingly being used and this trend is projected to continue [9]. Examples include the creation of student profiles to predict behaviours such as risk of dropping out [10]; intelligent systems that aim to track and shape the learning habits of students [11]; and community platforms that track and facilitate educator sharing of practices [12]. Furthermore, advances in technology and methodology have made it more feasible to aggregate and analyze data from multiple sources, from both inside and outside of the classroom, and to integrate physiological data. Examples of the former include ubiquitous learning approaches which merge situated and mobile learning concepts to enable “people to learn at any time and any place” [13]. Examples of the latter can be found in studies that make use of learner physiological data by means of electroencephalography (EEG) [14], galvanic skin response [15], heart rate monitors [16], and emotion recognition technology [17]. Moreover, with reference to artificial intelligence, the speed of computation has been doubling every three months since 2012 [18]. This suggests that the uses and related risks in the collected data are yet to be fully realized but are growing.

B. Big Data Risks: Beyond Security Breaches

Current policies such as General Data Protection Regulation (GDPR) highlight concerns around data privacy, cyber security, and datafication – which refers to the proliferate use of digital information “to access, understand and monitor people’s behavior” [19]. Yet, when asked about potential risks in data sharing, most students offer up risks related to unwanted advertising, security breaches and identity theft [3]. Nevertheless, additional risks have emerged out of the use of big data with practices such as microtargeting, surveillance, and the use of biased algorithms which can all work to subtly shift the behaviour
and beliefs of individuals to serve the interests of others – whether intentional or not. For example, microtargeting is the practice of profiling and targeting users with very specific attributes with the intent of influencing their beliefs and behaviours [20]. Microtargeting has been associated with discriminatory advertising, socially divisive advertising, and with strategies to create social discord [20]. Further, the proliferate use of social networks and the emergence of the Internet of Things has contributed to a level of apathy related to a loss of privacy and mass surveillance. Sharing private data, including biometric data, has become normalized. A recent report in the Times [21], brought forth concerns about the location tracking industry and emphasized how most are unaware of the pervasiveness of the data being collected about themselves and others; with little consideration being given to the potential consequences of such massive surveillance to one’s individual freedom. Additionally, case studies have demonstrated how biased algorithms can “perpetuate, exacerbate, or mask harmful discrimination” [22]. Yet, due to the obscure manner in which algorithms are often devised and deployed (e.g. lacking in interpretability and explainability), the biases and their lack of fairness are difficult to uncover [18]. Thus, a risk is not just that students become habituated to downplaying the massive collection of their personal data that is collected, and the resulting recommendations made to them by artificially intelligent systems. The more widespread adoption of learning analytics; adaptive and personalized learning environments; and mobile and ubiquitous learning approaches may contribute to increasing this aforementioned risk.

C. Educational Responsibilities

Educational institutions, educators and researchers regularly create and evaluate formative experiences for learners both with and without the use of technology. In situations requiring students to consent to sharing their data, who is responsible for helping students learn to make informed decisions regarding the sharing of their data? Beyond those asking students to use new technologies and share their data, it is not clear who else should hold the responsibility. A past study [3] found that two thirds of high school students surveyed were not confident in their abilities to responsibly manage their data; and none of the teachers surveyed had received training in responsibly managing student data. It can be argued that, especially in educational settings, if students are not being supported in making effective data sharing decisions then their apathetic approach to data sharing is being reinforced. Yet, supporting student decision making requires an investment of resources and may come at a cost of lower study enrolment rates as risks become more apparent. This leads to our research question: Do efforts to increase awareness and consideration of data sharing risks affect student enrolment in an educational technology research study?

Previous studies have evaluated student perceptions of data privacy in an educational technology context [3, 25, 26], explored ways to improve consent form formats to increase the likelihood that they would be read [2], and investigated participant ability to recall information from consent forms [4]. Our study explores the effects of an intervention aimed at helping participants make informed decisions regarding the sharing of their data on participant comprehension of a consent form and likelihood to enrol in the study.

II. Evaluation Methodology

To explore our research question, we conducted a quasi-experimental study involving first-year undergraduate students from a public university in Spain. A total of 201 students enrolled in an academic course called ‘Introduction to ICT’ were eligible to participate in this study and a parallel study that evaluated a collaborative learning application [23]. Data for the study was used from the engineering degree students who attended the class and gave consent for their data to be used for research purposes. A total of seven different classes were involved in the study (see Table S1 in supplementary material [24]).

A. Instruments and Experimental Design

Similar to studies that evaluated perceptions of data privacy in an educational technology context [3, 25, 26], we conducted a survey study to investigate the effects of priming students to consider data sharing risks and benefits on study enrolment rates, consent form comprehension, and reasons given for enrolling in the study. Informed consent best practices were adopted such as making use of an enhanced consent form and a consent form comprehension quiz. Further, a pre-lesson questionnaire aimed at priming participants to explicitly consider the harms and benefits of their participation was devised. The enhanced consent form which included an information annex and a separate signature sheet aimed to facilitate participant understanding by adopting a participant-oriented approach [27]; used simple rather than technical or legal language; and presented the information in clear subsections each with distinct headings and concise content summaries (see Table S2). The signature sheet required participants to explicitly check a box marking their participation choice with the choices being: Yes, I give my consent to participate in the research; No, I do not give my consent to participate in the research.

The consent form comprehension quiz consisted of seven open-ended questions on key elements of the consent form including the study purpose, risks and benefits, type of personal data to be collected, and participant rights. Comprehension was calculated by identifying how many of the potential answers from the consent form were recalled (see Table S3). Full points were awarded if the central ideas were communicated. Answers that could not be clearly graded were reviewed by a second grader before a final score was determined. The use of non-standardized quizzes tailored to the content of a specific consent form to measure comprehension of participants has been used in similar studies on informed consent [28, 29]. In addition to the quiz questions, the online form contained a question to elicit the reason for consenting in the research: “Did you consent to participate in the research? If so, why? If not, why not?”

The pre-lesson questionnaire was distributed to participants with the aim of facilitating participant
consideration of data sharing risks and to help prime participants to evaluate their participation in terms of a harm-benefit decision. Prior knowledge activation has been shown to facilitate learning [30]. Further, a study by Gilad and Kliger (2008) [31] found that priming manipulations were able to affect subjects’ risk attitudes. The pre-lesson questionnaire comprised of questions and level of agreement statements aimed at having participants reflect on their knowledge of data sharing and its risks; and to explicitly retrieve benefits and harms of data sharing (see Table S4).

The participants were distributed in two groups that followed different session flows: Flow A and Flow B (see Table 1). In Flow A, students answered the pre-lesson questionnaire at the start of the session; they then read the consent form and decided whether to give their consent or not; and then, they answered the consent form comprehension quiz. In Flow B, students started by reading the consent form; decided whether to give their consent or not; answered the consent form comprehension quiz; and finally completed the pre-lesson questionnaire. Descriptive statistics were used to report results of the closed questions in the questionnaires. An inductive thematic analysis [32] was conducted on the open question, "Did you consent to participate in the research? If so, why? If not, why not?" The analysis was used to identify the main reasons (themes) for giving consent (see Table 2). Statistical analysis was performed using SPSS (IBM Corp.) [33] to evaluate the effects of the session flows on consent form comprehension quiz scores and participant likelihood to consent.

### III. Results

157 students attended the 'Introduction to the ICT' classes and 129 of them (82.17%) consented to participate in the research. However, data from two participants were incomplete and removed from the study. The results include data from the remaining 127 participants (34 female, 90 male and 3 preferring not to identify their gender; age: $M = 18.65, SD = 1.95$).

A. Enrollment rates

In Flow A, there were 77 participants and 80 in Flow B. In Flow A, 90.91% ($n = 70$) of participants consented while only 73.75% ($n = 59$) from Flow B did so. Data from two consenting participants, one from Flow A and one from Flow B, were incomplete and removed.

B. Consent form comprehension

The participants in Flow A that completed the pre-lesson questionnaire before reading the consent form had lower consent form quiz scores ($M = 35.49\%, SD = 14.77$) in comparison to participants from Flow B ($M = 43.00\%, SD = 14.70$; $t(127) = 2.89, p = .005, d = .51$). Furthermore, in comparing answers to the consent form quiz question, “What are the risks and potential consequences of participating in the research?” participants in Flow A ($M = 0.73, SD = 0.448$) performed worse on the risk-related question than participants in Flow B ($M = 0.90, SD = 0.307$; $t(121.99) = 2.51, p = 0.014, d = 0.443$) (see Table S5).

#### C. Reasons for enrolment

In conducting an inductive thematic analysis of participant answers to the survey question, “Did you consent to participate in the research? If so, why? If not, why not?” four categories of answers were generated and compared between groups (see Table 2). Answers that were relevant to more than one category, were marked for each category to which they were relevant. From the 127 participants, 221 code labels were generated from the answers for an average of 1.7 per participant. Table 3 shows the percentage of participants assigned to each categorization. The most common categorizations were beneficence (73.23% of participants), trust (24.60%), and convenience (14.96%). In comparing the two flows (see Table 3), Flow A had a significantly higher number of trust categorizations (31.88%) when compared to Flow B (15.51%; $X^2 (1, n = 127) = 4.575, p = .032$). On the other hand, Flow A had a significantly lower number of beneficence categorizations (63.77%) when compared to Flow B (84.48%; $X^2 (1, n = 127) = 6.897, p = .009$).

#### Table I. Order of activities in the two session flows

<table>
<thead>
<tr>
<th>Flow A</th>
<th>Flow B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-lesson questionnaire</td>
<td>Consent form</td>
</tr>
<tr>
<td>Consent form</td>
<td>Consent form quiz</td>
</tr>
<tr>
<td>Consent form quiz</td>
<td>Pre-lesson questionnaire</td>
</tr>
</tbody>
</table>

#### Table II. Thematic categories of reasons for consenting

<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
<th>Example Statement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beneficence</td>
<td>Willingness to contribute to research and/or educational improvements</td>
<td>Yes, because, in my opinion, everything that can improve the education is worth it.</td>
</tr>
<tr>
<td>Trust</td>
<td>Trust of the institution and/or measures taken to reduce risks of participation.</td>
<td>Yes, because it was from the university and they wouldn't do anything bad to me.</td>
</tr>
<tr>
<td>Convenience</td>
<td>No inconvenience perceived in participating</td>
<td>Yes, I did, because I guess that there is nothing to lose if I accept it.</td>
</tr>
<tr>
<td>Interest</td>
<td>A personal interest in the topic or study</td>
<td>Yes, I do, it sounds interesting.</td>
</tr>
</tbody>
</table>

#### Table III. Comparison of reasons for consenting

<table>
<thead>
<tr>
<th>Category</th>
<th>Flow A+B</th>
<th>Flow A</th>
<th>Flow B</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beneficence</td>
<td>73.2%</td>
<td>63.8%</td>
<td>84.5%</td>
<td>.009</td>
</tr>
<tr>
<td>Trust</td>
<td>24.6%</td>
<td>31.9%</td>
<td>15.5%</td>
<td>.032</td>
</tr>
<tr>
<td>Convenience</td>
<td>15.0%</td>
<td>15.9%</td>
<td>13.8%</td>
<td>.735</td>
</tr>
<tr>
<td>Interest</td>
<td>7.9%</td>
<td>8.7%</td>
<td>6.9%</td>
<td>.708</td>
</tr>
</tbody>
</table>
IV. DISCUSSION

The main findings of our study were that primed participants performed worse on the informed consent comprehension test; but had higher enrolment rates and were more likely to cite ‘trust’ as a factor influencing their decision to enrol in the study. As the study was explorative, it is difficult to draw concrete conclusions, but a key takeaway is that the efforts made to increase potential participant consideration of data sharing risks and benefits did not negatively affect enrolment rates in the study.

In educational technology research, there has been a call for more empirical studies related to ethics [34] as studies in the field are increasingly incorporating techniques and technologies that are increasing risks to participants (e.g. data aggregation, machine learning, artificial intelligence, natural language processing, contextual learning, physiological data). Likewise, it is important to discuss where the responsibility for helping students learn to make informed decisions regarding the sharing of their data lies. It could be argued that educational institutions, educators and researchers – those who ask students to use new technologies and share their data in formal educational contexts – hold some of the responsibility. Moments in which consent is requested are valuable opportunities for student learning. Proponents of situated learning have argued that students acquire the skill to perform by actually engaging in the process. For example, Lave and Wenger [35] state that meaning, understanding and learning are all defined relative to actional contexts. The moments in which consent is requested are actional contexts that, if occurring in a formal educational setting, would be unfortunate to disregard. These moments provide educators with the opportunity to have students critically reflect on and make meaning out of the data sharing experience. Finally, it could be argued that the risks associated with big data and data collected by educational technologies are far below those of biomedical data and interventions and do not warrant the additional efforts to inform potential participants. However, it can also be argued that the net benefits of technology and big data in education are not as apparent at the level of the individual. Especially with the negative effects of excessive technology use on student learning [36] and well-being [37] increasingly being documented.

A. Limitations of the study

The study took place in an authentic setting which may limit the generalizability of the findings. Moreover, the students in the course were able to decide whether or not to share their data with researchers. As a result, student responses were limited to only those participants who had consented to sharing their data. The study was not able to make use of the data from students who did not consent nor evaluate their reasons for not consenting. Further, the lower comprehension scores in Flow A may be attributable to cognitive fatigue as participants in Flow A completed a task prior to reading the consent form whereas participants in Flow B did not. Future iterations of the study should address the imbalance of tasks; reduce the variance between classes by having the same instructor for all classes; and improve the validity and reliability of the consent form comprehension quiz. The quiz used in this study consisted of open questions and were analyzed using an inductive thematic approach. A closed format quiz such as the Quality of Informed Consent (QuIC) used in bioethical research [38] may result in more reliable comprehension scores.

V. CONCLUSION

We conducted an empirical study involving ethics in educational technology research by exploring the question: Do efforts to increase awareness and consideration of data sharing risks affect students’ decisions to enrol in an educational technology research study? In our study, we made use of a pre-lesson questionnaire to prime participants to consider data sharing harms and benefits; explored the impact on study enrolment, consent form comprehension, and on reasons given by participants for enrolling in the study. In the context of our study, primed participants showed lower levels of consent form comprehension. This negative effect is possibly attributable to cognitive fatigue resulting from the sequencing of tasks. However, priming participants to think about harms and benefits may have contributed to higher enrolment rates as evidenced by such participants being more likely to identify trust as a critical factor that influenced their decision to give consent. Overall, the study aims to encourage more discussion related to the responsibilities of educational institutions, educators and researchers – those who ask students to use new technologies and share their data in formal educational contexts – to more explicitly integrate and help students understand the critical views of their technological approaches.

ACKNOWLEDGMENT

This work has been partially funded by the European Regional Development Fund and the National Research Agency of the Spanish Ministry of Science, Innovation and Universities, under project grants TIN2017-85179-C3-3-R and MDM-2015-0502. D. Hernández-Leo acknowledges the support by ICREA under the ICREA Academia programme.

REFERENCES


[18] Raymond Perrault, Yoav Shoham, Erik Brynjolfsson, Jack Clark, John Etchemendy, Barbara Grosz, Terah Lyons, James Manyika, Saurabh Mishra, and Juan Carlos Niemela, “The AI Index 2019 Annual Report”, AI Index Steering Committee, Human-Centered AI Institute, Stanford University, Stanford, CA, December 2019


[22] Executive Office of the President, Munoz, C., Director, D. P. C., Megan (US Chief Technology Officer Smith (Office of Science and Technology Policy)), & DJ (Deputy Chief Technology Officer for Data Policy and Chief Data Scientist Patil (Office of Science and Technology Policy)). (2016). Big data: A report on algorithmic systems, opportunity, and civil rights. Executive Office of the President.


