PARTIAL RESULTS OF A REVIEW OF SURVEY METHODS MEASURING E-PRIVACY CONCERNS

Anders Matre, Magnus Englund and Vanessa Ayres-Pereira
University of Bergen, Christiesgt. 12, Bergen, Norway

ABSTRACT
We present partial results of a review of survey methods used to measure e-privacy concerns and discuss the validity of the privacy paradox. We analyzed the content of 246 questionnaire items used to measure privacy concerns in 27 papers. We conclude that privacy concerns have been operationalized as a heterogeneous construct and that the instruments may be partially responsible for varying conclusions about the paradox. We discuss the necessity to enhance content validity and investigate measurement accuracy.

KEYWORDS
Privacy Concerns, Privacy Paradox, E-privacy

1. INTRODUCTION

According to some researchers, measuring one’s level of concern about the privacy of their digital data (e-privacy hereafter) is relevant to predict one’s likelihood of engaging in privacy-protective behaviors (e.g., Malhotra et al, 2014). However, a body of research emerged contradicting this view and concluding that, despite reporting high levels of e-privacy concerns, users are likely to disclose personal information even for small benefits; this phenomenon was coined the privacy paradox (cf. Kokolakis, 2017, for a review). Kokolakis (2017) suggested that the contradictions across studies (sometimes supporting, sometimes refuting the paradox) could be, among other reasons, a function of the variability in the methods. Most surveys assess different aspects of the environment that people may be concerned about, however, the aspects can vary across studies. Therefore, researchers have been arguing for more work that helps to develop instruments to measure e-privacy concerns, especially in experiments (e.g., Buck et al, 2018).

Previous works have attempted to develop reliable and valid scales to measure privacy concerns (e.g., Buck et al, 2018; Malhotra et al, 2004; Smith et al, 1996; Stewart & Segars, 2002). The scales, however, vary in terms of scope and dimensions. While some assess privacy concerns in general (e.g., Stewart & Segars, 2002; Malhotra et al, 2004), other focus on specific sectors (Buck et al, 2018). Whereas some consider concerns for data collection, errors, unauthorized access, and secondary use (e.g., Smith et al, 1996; Stewart & Segars, 2002), others assess control factors, awareness of privacy practices (Malhotra et al, 2004), the type of data and benefits for disclosure (Buck et al, 2018).

Kokolakis (2017) noted that a common framework for measuring e-privacy concerns had not yet emerged. Therefore, the present paper aims to assess how researchers interested in the privacy paradox have defined and measured e-privacy concerns recently. This is a work-in-progress. And, while preliminary, our findings show trends in the operationalization of privacy concerns and provide parameters for item design considering current practices in the literature. We also discuss how some practices could be biasing and eliciting a higher degree of agreement towards privacy concerns.
2. BODY OF PAPER

2.1 Method

A systematic paper selection process was performed in several steps conforming to PRISMA guidelines for systematic reviews (Shamseer et al., 2015). A first set of papers (n = 40) was identified through the database PsycInfo and the second set (n = 62) through the reference list of the first set of papers. We selected the papers that were published within the last five years (2016–2020), involved the expression privacy paradox, used survey methods, included the entire questionnaire transcribed, and reported privacy concerns measures (inclusion criteria). We excluded literature reviews, papers about offline privacy decisions, and not peer-reviewed (exclusion criteria). Finally, we analyzed 246 questionnaire items measuring e-privacy concerns, published in 27 papers. We analyzed the content of the items using the content analysis technique (Hsieh & Shannon, 2005), that is, by determining the existence and frequency of several categories of analysis (see Results for details).

2.2 Results

2.2.1 General Trends Across Papers

Only four out of the 27 papers included a definition of the construct “privacy concerns”. All definitions focused on the lack of control and/or potential misuse of one’s data. Our sample of papers contained little to no discussion of potential questionnaire methodology shortcomings. All papers used closed-ended questions and a battery, rather than a single item, to measure e-privacy concerns. The number of items ranged between three and 25 (M = 9.11, SD = 6.41). No questionnaire was entirely identical, but, according to a text analyser tool, six studies used near-identical wordings in some items. These items could be traced back to Smith et al. (1996). Most studies used Likert scales (88.9%), 19 encompassed agree–disagree, mostly with seven or five points.

2.2.2 General Trends Across Items

The analysis of “e-privacy scenarios” typically involve the description of a) a technological context (e.g., social media), b) one or two actors—the user and a second agent, c) the type of information disclosed by the user (e.g., location), and d) an undesired consequence of data usage for the user (i.e., e-privacy risk). The user is any person who provides personally identifiable data for an audience. The second agent is anyone who acts upon the users’ data (e.g., collects, accesses, shares). Considering this perspective, we determined the aspects of the “e-privacy scene” described in the items and their frequency. We also evaluated how many among the 246 items inquired about attitudes or beliefs.

Most of the items (58.1%) inquired about negatively valanced attitudes, often using literally the word “concern” (38.6%). Other items inquired about different types of attitudes (e.g., trust, importance) or beliefs. Most items regarded actions of a second agent upon the user’s data (78.5%). The type of action varied: a quarter (25.2%) covered data access, usually unauthorized, and 19.9% addressed data collection. Although the majority inquired about the actions of a second agent, only 74.2% specified who the agent was. Those items that specified the agent, often treated of companies (e-commerce and social network sites, SNSs, 35.8%) or individuals (referred to as “people”, “others” or “somebody”, 11.4%). Nearly 60% specified the technological context (mostly SNSs, 42.3%). Only 24.8% referred to which type of data was processed, but it was often generally worded as “personal information”. The items rarely specified actions taken by the users or personal negative consequences that could arise from data disclosure.

2.3 Discussion

Discussions on the privacy paradox usually assume that the different studies measure the same constructs. However, we found almost a lack of definition of the core construct and variations in the content of most of the analyzed measurement aspects. Still, the items do appear to have some commonalities. A typical privacy concern item will likely contain a reference to an attitude of concern, in a specific technological context,
about an action of data collection of access to the user’s data, and answers will be rated on an agree-disagree scale. The superficial agreement across items increases the chances of replicability, but some of the current design practices, if overlooked, might leave questionnaires subject to measurement error and bias. Questions with unprecise information—that is, input information different from the one present when deciding whether or not to disclose data—can lead to a reduction in measurement accuracy and attitude-behavior consistency (cf. Schwarz & Bohner, 2001). The content of most items was general and rarely specified the type of data processed, how it could be misused, and who the second agent was. General terms can make respondents infer that the wordings refer to different objects or draw their answers on unintended contextual features. Also, vague information added to items framed with a focus on the presence of concern (and rarely on the absence of concern or both) could bias responses towards the selection of agreement rather than disagreement options. Our considerations call for future experiments to evaluate items’ effects on responses and the instruments’ ability to relate the construct to privacy-protective behaviors.

3. CONCLUSION

Our analyses, while preliminary, contribute to the literature by revealing recent trends on the operationalization of privacy concerns in the privacy paradox literature. In light of empirical discussions, we offer a brief examination of how some practices could be influencing agreement towards privacy concerns, reducing attitude-behavior consistency and, consequently, conclusions on the paradox. Our analyses are still in progress and must be expanded to evaluate associations between categories of analysis, as much as with the papers’ varied conclusions on the paradox. While preliminary, the results can be applied to design items measuring e-privacy concerns considering current practices. The findings also reveal the potential to conduct studies evaluating contextual effects to enhance measurement accuracy and the value of privacy concerns in predicting privacy behaviors.

ACKNOWLEDGEMENT

A.M. and M.E. conducted the data collection, analyses, and drafted the first version of the manuscript. V.A.P. conceived and supervised the project; she also commented on the manuscript. This research is part of the scientific program of the ALerT project, supported by the Norwegian Research Council. We thank Dr. Gisela Böhm for comments and encouragement throughout the development of this paper.

REFERENCES