Who Quits Privacy-Invasive Online Platform Operators? A Segmentation Study with Implications for the Privacy Paradox

Sebastian Hermes
Technical University of Munich
sebastian.hermes@tum.de

Anela Sutanrikulu
Technical University of Munich
anela.halilovic@tum.de

Maximilian Schreieck
Technical University of Munich
maximilian.schreieck@tum.de

Helmut Krcmar
Technical University of Munich
helmut.krcmar@tum.de

Abstract

Although individuals are concerned about their privacy, it is increasingly difficult to withdraw from privacy-invasive platform operators and keep activities private. IS research has identified the privacy paradox as a phenomenon and information asymmetries as one critical reason behind users dichotomy between privacy concern and behavior. However, prior work neglected to investigate (1) the characteristics of consumers caught in the privacy paradox, (2) new areas of information asymmetries such as knowledge about alternative services, and (3) new privacy-decision processes such as quitting privacy-invasive platform operators. To close these gaps, we conducted a representative segmentation study of Google and its services across five countries guided by the theory of planned behavior. Our results identify three clusters and indicate that the privacy paradox is only prevalent in two of them. Consumers in these two clusters lack knowledge about data integration, data usage, and alternative services.

1. Introduction

Privacy concerns are one of the key challenges that organizations, policymakers, and society face in the contemporary digital era [1, 2] and are especially prevalent for digital platforms. This is why Yun et al. (2018) [3] called to "investigate the PIP [personal information privacy] concerns toward the unknown or hidden fifth parties [such as] data crawling/data mining companies, business intelligence companies, and [...] big data companies ([...] SAP, Amazon, Google, etc.)." Their call is also supported by Lowry et al. (2017) [4], stating that exciting opportunities arise when we put "privacy at the center of the IS artefact by focusing on (1) online platforms, (2) the IoT, and (3) big data." Investigating these digital platforms is a pressing matter, because from a privacy perspective, it is nowadays easy to de-anonymize a person using information from various sources [5], and big data companies are doing just that [4]. By developing extremely specific user profiles [6], big data companies also create novel, highly ambitious privacy issues. Google' services are incorporated into most systems worldwide, including mobile operating systems, search, e-mail, and mapping applications [4]. As a result, it is becoming more and more difficult to withdraw from these global systems or keep our activities private.

In this context, research has identified an inconsistency between consumers attitudes and actual behavior. This so-called privacy paradox refers to consumers indicating a high level of privacy concern while simultaneously neglecting their privacy and data disclosure protections [7]. For example, consumers emphasize their concerns about their data, the willingness to protect their data, and their control over who has access to it [8], while at the same time disclosing a variety of personal data, often without reviewing the privacy policy of the service provider [9]. Therefore, the question arose of how this was able to occur. One explanation for the privacy paradox are information asymmetries. Information asymmetries refer to information that is relevant to privacy-decision making, but not known to all actors involved in the privacy-decision process [10].

Research on information asymmetries has largely investigated information such as privacy risks (e.g. identity theft) or protection techniques (e.g. privacy-enhancing technology) [11, 10]. To better understand the role of these different areas of information asymmetries for the privacy paradox, prior work has mainly study the privacy-decision process of disclosing information (giving privacy away). From an empirical point of view prior work primarily studied students in the context of e-commerce and social networks [11].

Research has, to the best of our knowledge, neglected to investigate (1) characteristics of consumers caught in the privacy paradox (exception: [12], (2) new areas of information asymmetries such as

URI: https://hdl.handle.net/10125/71181
978-0-9981331-4-0

objective knowledge about data collection, integration, and usage as well as knowledge about alternative services, and (3) new privacy-decision processes such as quitting privacy-invasive online platform operators (taking privacy back).

To close these gaps, we conducted a representative segmentation study of Google and its services such as Search and Chrome across five countries guided by the theory of planned behavior with the aim of answering the following two overarching research questions: *Which users are willing to quit privacy-invasive online platform operators and what are the implications for the privacy paradox?*

The remainder of this paper is structured as follows. Section 2 first reviews the literature on the privacy paradox and current explanations for it and then describes the theory of planned behavior (TPB). Section 3 describes our methodology. In Section 4 we present the results of the cluster analysis and in Section 5 we present differences between clusters. Section 6 discusses the theoretical and practical implications. The paper concludes and presents limitations and future research in Section 7.

2. Theoretical Underpinnings

2.1. Privacy Paradox

Recent privacy laws in the European Union and the US have adopted the standpoint that privacy is a matter of autonomy and control over the collection, storage, and use of information [13, 14]. With the rise of online platforms and lack of transparency, users' control over their personal information has become more difficult. Thus, users may develop concerns about how their personal data are processed when they use online platforms. In fact, research has shown that users are highly concerned about their privacy [11]. Supposedly, users try to protect their privacy. However, research has also demonstrated that this might not always be the case. Despite being exposed to potential privacy threats, such as unwanted contracts or advertisements or identity theft, users are willing to disclose their data by using online platforms [15]. This dichotomy between privacy concern or privacy attitude and users' actual behavior is known as the "privacy paradox" [11].

In the effort to explain the dichotomy between privacy attitude or concern and user behavior, several theories and interpretations have been developed. The theory of information asymmetries and incomplete information indicates that missing information hinders consumers from making rational decisions. However, the concept of bounded rationality demonstrates that even if individuals had access to complete information, they might not be able to process the information to make a rational privacy decision [10]. Hence, individuals' bounded rationality limits their ability to obtain, remember, and process all information. As a consequence, individuals relay on mental models and heuristics. Furthermore, the knowledge gap hypothesis addresses privacy literacy and indicates that users' lack of privacy literacy, such as users' lack of knowledge about technical aspects of online data protection, prevents them from behaving according to their attitudes and concerns [16]. A further theory is the privacy calculus theory, which implies that consumers conduct a rational calculus of losses and gains before disclosing their personal information, wherein the final outcome is determined by the privacy trade-off [17, 11]. Thus, users might weight the gained benefits more than the risks of disclosing their personal information. There are also other interpretations and assumptions regarding human behavior that might explain the privacy paradox, such as optimism or affect bias.

2.2. Theory of Planned Behavior

Given that online platforms are fueled by user data, privacy is a concern that directly affects users. For this purpose, it is important to understand whether there is a difference in the ways different users handle these privacy concerns. Furthermore, it is important to understand what users do about their privacy concerns. To explain user behavior, we draw on the TPB. The TPB, developed by Ajzen (1985) [18], is a common theory used for developing models that explain human behavior with respect to various phenomena. It incorporates three key determinants (attitude toward the behavior, subjective norms, and perceived behavioral control) that form an intention, which, given a sufficient degree of actual control, results in behavior [19].

Just as intentions are held to have determinants, so do attitude, subjective norms, and perceived behavioral control. All three variables are an expectancy-value function of salient beliefs. Attitude toward the behavior, which can be either favorable or unfavorable, is composed of the multiplicative combination of the perceived likelihood that performance of the behavior will lead to a particular outcome and the evaluation of that outcome [20, 21]. Subjective norms explain a persons belief about the extent to which significant others think that a person should engage in a behavior or not, which incorporates a social pressure and the motivation to comply with these referents [22, 19]. Perceived behavioral control (PBC) denotes a subjective degree of control over the performance of a behavior

[23]. The more resources and opportunities individuals perceive to have and the fewer obstacles or impediments that they encounter, the greater their PBC over the behavior should be [20]. Figure 1 illustrate the causal model of the TPB[1]

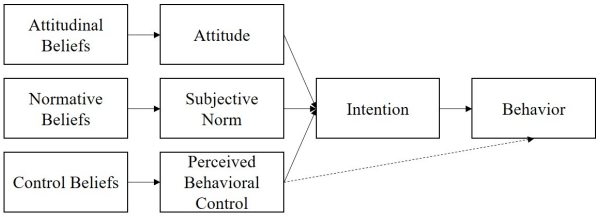

Figure 1. The Causal Model of the TPB adpated from [20, 23]

3. Methodology

3.1. Eliciting External Beliefs and Deriving the Objective Knowledge Scale about Information Privacy

We conducted a belief elicitation study (pre-study) using an open-ended questionnaire, following the approach by Ajzen (2002) [24]. The objective was to freely elicit the most salient attitudinal, normative, and control beliefs. We solicited the key drivers of each behavior from a convenience sample of 19 participants, which included faculty, staff, and students from the Technical University of Munich. Their responses are sorted based on the frequency mentioned. We then chose the beliefs that indicated a high frequency.

Regarding the knowledge scale, we generated and evaluated items and the corresponding dimensions based on a literature review, interviews, and a pretest. First, we conducted a review of information privacy and generated 23 items and 3 dimensions (data collection, data integration, and data usage) that were suitable for measuring an individuals knowledge about information privacy. Second, we interviewed two faculty members, two privacy consultants, and one online marketing executive to confirm the construct domain and dimensionality. Third, we interviewed three IS undergraduates and two consumers to evaluate face validity. Fourth, we conducted a Q-Sort with six IS doctoral students to assess content validity. During the second, third, and fourth steps we reworded and deleted items and converged on 15 multiple choice questions. Lastly, we conducted a small online questionnaire with 28 consumers and used the results to eliminate additional six questions since they were unsuitable for distinguishing between novice and expert knowledge and therefore failed discriminant validity.

[1]Note that our objective is not to reveal the causal model for quitting privacy-invasive platform operators. Instead we use the constructs of the TPB to determine clusters of individuals that are affine, torn, and non-affine across three different behaviors.

3.2. Data Collection

A representative online survey was conducted among consumers in five countries exploring their knowledge and opinions about Google and its services. The survey was distributed by a market research institute which had partner organizations in each country. The partner organizations recruited respondents and compensated them for participation. We chose Google's service ecosystem as an example of a company operating privacy-invasive online platforms [25, 26]. For example, Google disclosed search queries to third parties without user consent and merged privacy policies without user consent [27]. The survey results were obtained from 1,433 individuals: 274 in Denmark, 299 in France, 274 in Germany, 289 in the UK, and 297 in the US. 730 individuals were male and 703 female.

3.3. Measurements

The survey instrument was adopted from Conner and Sparks (2005) [28] and translated into Google's online platform context. According to the TPB, each behavior must be defined within a well-specified target, action, context, and time frame [24, 23]. In total, we used three different behaviors to better assess quitting a service provider completely, not only a specific service of that service provider. The three behaviors were (1) using (action) a different browser (target) than Google Chrome (context) in the next six months (time frame), (2) using a different search engine than Google Search in the next six months, and (3) being signed off my personal Google account while using Google Search in the next six months. The survey questions were designed to cover the TPB constructs such as behavioral beliefs (e.g. Me using a different browser than Google Chrome would increase the loading time of websites I want to access), normative beliefs (e.g. Privacy experts think I should use a different browser than Google Chrome.), control beliefs (e.g. For me to use a different browser than Google Chrome in the next 6 months will be very difficult to very easy), attitude (e.g. Me using a different browser than Google Chrome in the next 6 months would be very impractical to very practical), subjective norms (e.g. Most people who are important to me use a different browser than Google Chrome), perceived behavioral control (e.g. I am confident that I can use a different browser than Google Chrome in the next 6 months), behavioral intention,

and actual behavior. Additionally, we measured privacy concerns [29] (e.g. I am concerned that online service providers may keep my private information in a non-secure manner), subjective information privacy knowledge [30, 31] (e.g. In general, I am quite knowledgeable about how online companies collect, manage and use my personal information), and objective information privacy knowledge (self-developed) (e.g. Online companies use cookies to collect information from your hard drive). All items were measured on a 7 point Likert scale except objective information privacy knowledge which was conducted as multiple choice questionnaire.

3.4. Data Analysis

To perform a cluster analysis on the survey data, an exploratory factor analysis (EFA) was conducted first to find underlying factors and reduce the dimensionality of the dataset. Studies have shown that "given a sufficiently large number of response categories (e.g. seven), and absence of skewness, and equal thresholds across items, it seems possible to obtain reasonable results", so factor analysis can be performed without the assumption of normality within the data [32]. Thus, given that the survey data were ordinal, we neglected this assumption. The EFA processes began with a test of absence of multicollinearity and singularity within the variables. Provided that no items had a squared multiple correlation close to 0 or close to 1, the test indicated no issues. To complement these results, a Bartletts Test of Sphericity was performed. The test showed that the correlations between items were sufficiently large ($X^2(3655) = 104051.67, p < 0.01$). Next, the Kaiser-Meyer-Olkin verified the sampling adequacy of the analysis (0.96) [33]. After these tests showed the appropriateness of using EFA to process the data, the number of factors had to be chosen. To do so, using eigenvalues is suggested, based on the Kaisers criterion [34] as well as Horn's parallel analysis [35], which suggested 13 and 15 factors, respectively. After testing the suggested numbers of factors, i.e., examining how variables loaded onto factors using different factor extractions and rotation methods, it was decided to use 15 factors. For further analysis, it was decided that the maximum principal axis factor extraction method should be used, which is suggested for data that do not follow a normal distribution [36]. Furthermore, the oblique rotation method, specifically the promax rotation, was used, given that a factor correlation could not be excluded. Factors below the factor loading criterion of 0.40 were removed one by one, based on the number of factors and the loading intensity [37]. Table 5 in the Appendix demonstrates the factor loadings. For the retrieved factor solution Cronbach's alpha was evaluated to test the internal consistency reliability of each factor, which is provided in Table 1 along with the factor naming. The Root Mean Square of the Residuals of 0.02 and the factoring reliability of 0.873 indicated a good model fit. To use the EFA results for the cluster analysis, factor scores were computed using Bartletts approach. These scores were centered at zero such that a positive score indicated that the items belonging to the factor had an above average loading, while a negative score indicated that the items had a below average loading onto those factors.

Factor name	Cronbach's Alpha
control belief power	0.94
privacy concerns	0.95
behavioral belief strength	0.93
subjective knowledge	0.90
attitude toward Google Search	0.94
intention toward using alternatives	0.93
perceived behavioral control	0.90
evaluation of outcome	0.78
attitude toward Google Chrome	0.92
perceived norms	0.90
attitude towards sign-in behavior	0.93
normative belief strength	0.92
motivation to comply	0.93
control belief strength	0.78

Table 1. Factor Names and Cronbach's Alpha

To cluster the data, k-means was used. The elbow method, the average silhouette width, and the gap statistic were used to examine the number of clusters needed for the clustering algorithm. However, these methods yielded different results. After some testing, including different randomizations of cluster centroids, three clusters turned out to be the most reasonable number for further analysis. The clusters produced by k-means were then appended to the dataset of factor scores to serve as classification labels. A multiclass classification using the XGBoost algorithm was performed to find the most influential factors. XGBoost is an optimized distributed gradient boosting machine learning algorithm [38]. The most influential factor was control belief power, followed by PBC, intention toward privacy protection behavior, attitude toward Google Chrome, attitude toward sign-in, and motivation to comply with experts. K-means was then run again on the dataset, including only the most influential factors. The results of this run were taken as the final clustering solution. The overall average silhouette coefficient was equal to 0.23 and only a few observations were mis-clustered. Furthermore, analysis of variance (ANOVA) test of the factors and the clusters showed a statistical significance ($p < 0.01$).

4. Cluster Analysis Results

Cluster 1 encompasses factor scores that are above the average zero mean. The cluster included 308 users. These users had a particularly high intention to use a browser other than Google Chrome, a search engine other than Google Search, as well as to avoid being signed into their Google account during the next six months. Furthermore, these users indicated that they would find it rather easy to use services other than Google or their Google account. These users also indicated that, for them, knowing alternatives and reading about data leaks and privacy violations would make using a different browser and search engine other than those provided by Google reasonable. They also had a positive attitude about using a browser other than Google Chrome and avoiding their Google account over the next six months. Similarly, these users were highly motivated to do what privacy experts recommended. Overall, these users did not a have high affinity toward Google services, thus they were named non-affine users.

Cluster 2 contained users that had the opposite preferences and therefore shared no similarities with cluster 1. These users had, on average, lower factor scores. The cluster included 367 users; slightly more than cluster 1. Thus, these users had a relatively low intention to leave Google services over the next six months. They would also find it rather difficult to leave Google Search and Chrome and their account. These users also indicated that, for them, knowing alternatives and reading about data leaks and privacy violations would not make using a different browser and search engine other than those provided by Google likely. Moreover, they had a rather negative attitude about using a browser other than Google Chrome or to avoid being signed into their Google account over the following six months. They also had a rather low motivation to comply with the opinion of privacy experts. In summary, cluster 2 included users that had a high affinity toward Google services. Hence, they were called affine users.

Cluster 3 included users that held a rather neutral standpoint. It included 758 users, the vast majority of the survey participants. The factor scores were all centered around zero, meaning that these users represented the average response. It seems that these users had a minor positive tendency to use different services. However, in general, these users were undecided and are therefore named torn users. A summary indicating the average factor scores per cluster is given in Table 2.

Factor	Non-affine	Affine	Torn
control belief power	1.1211	-1.0013	0.02925
PBC	1.1407	-1.0896	0.06403
behavioral intention	1.1858	-0.9471	-0.02327
attitude toward Google Chrome	1.0117	-0.8672	0.008772
attitude toward sign-in behavior	1.0496	-0.8520	-0.01399
motivation to comply	0.9483	-0.9387	0.06920

Table 2. Factor Score Averages by Cluster

5. Analysis of Cluster Differences

The segmentation of clusters across countries and gender involved several tests. The chi-squared test indicated that countries and clusters had a dependency ($p < 0.01$), but that gender and the clusters are not related ($p > 0.05$). A Bonferroni test to assess which clusters contributed to the significance between the different countries indicated that only affine and non-affine users significantly differed among countries ($p < 0.05$). The distribution of each cluster across all countries is listed in Table 3. It should be noted that in Denmark, affine users are more than twice as frequent than non-affine users. In comparison to other countries, Denmark has the most affine users and the least non-affine users. Generally, platform torn users are more frequent than other users in all countries, with France having the most in comparison to the other countries.

	Non-Affine	Affine	Torn
UK	23.38%	17.44%	20.18%
USA	23.38%	20.71%	19.66%
France	20.78%	15.26%	23.61%
Denmark	11.36%	27.25%	18.34%
Germany	21.10%	19.35%	18.21%

Table 3. Proportion Table of Countries and Clusters

Differences with respect to age were tested using an ANOVA test, given that age is a continuous variable. Lavenes test for homogeneity of variances and Shapiros test for normality were also conducted. The latter test showed that age was not normally distributed. Given that there is some controversy regarding whether ANOVA should be run on non-normally distributed data, the test was run but also complemented with a Kruskal-Wallis test. The results for both tests show that there exists significance between the clusters ($p < 0.05$ and $p < 0.05$). The Tukey test was conducted to observe which clusters dragged the significance. It showed only a significance ($p < 0.05$) between affine and torn users. A box plot in Figure 2 shows the differences in the age variable across clusters. The median age of the torn users is the highest, while the median of the affine users is the lowest. This indicates that younger participants tend to neglect privacy issues and continue using Google

services.

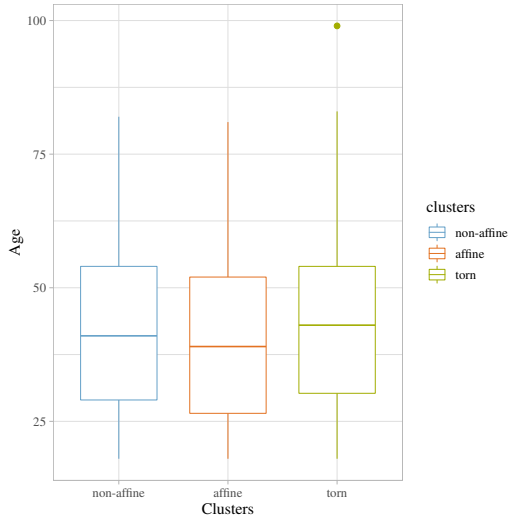

Figure 2. Boxplot for Age

The segmentation of clusters across objective knowledge of information privacy included the testing of how clusters differed based on user responses to questions regarding data collection, data integration, and data usage. A chi-squared test indicated a significant dependency between the individual items and the clusters. The Bonferroni test for the data collection dimension indicated that only the difference between non-affine users (mean of 0.32) and torn users (mean of 0.28) was significant ($p < 0.05$). The same type of test indicated that data integration and data usage had differences between non-affine users (mean of 0.39 and 0.76) and affine users (mean of 0.32 and 0.70) ($p < 0.01$ and $p < 0.05$) and non-affine users and torn users (mean of 0.31 and 0.69) ($p < 0.01$ and $p < 0.01$). The proportion contingency in Table (4) for the three variables shows that affine users almost twice as frequent as non-affine users answered all three question blocks related to objective knowledge incorrectly.

The segmentation of clusters based on knowledge about alternatives was performed to see if users from different clusters considered themselves knowledgeable about alternatives to Googles services. To do so we draw on the construct control belief strength which basically covers items such as I know alternatives to Google Chrome. Both a Kruskal-Wallis ($p < 0.01$) and ANOVA ($p < 0.01$) test demonstrated that there exists significant difference between clusters. The Bonferroni test indicated that only the difference between non-affine users (mean of 5.8) and torn users (mean of 4.9) as well as non-affine users (mean of 5.8) and affine users (mean of 4.8) were significant ($p < 0.01$). To visualize the

Data Collection	Non-affine	Affine	Torn
0%	19.81%	29.97%	29.16%
33%	65.91%	54.22%	58.84%
66%	12.66%	14.44%	10.82%
100%	1.62%	1.36%	1.19%
Data Integration	Non-affine	Affine	Torn
0%	17.21%	30.79%	29.95%
33%	50.97%	45.23%	48.28%
66%	29.87%	22.34%	19.92%
100%	1.95%	1.63%	1.85%
Data Usage	Non-affine	Affine	Torn
0%	3.90%	6.81%	9.23%
33%	13.96%	15.53%	15.17%
66%	32.14%	38.96%	35.22%
100%	50.00%	38.69%	40.37%

Table 4. Proportion Table of Objective Knowledge and Clusters

differences, a box plot is displayed in Figure 3.

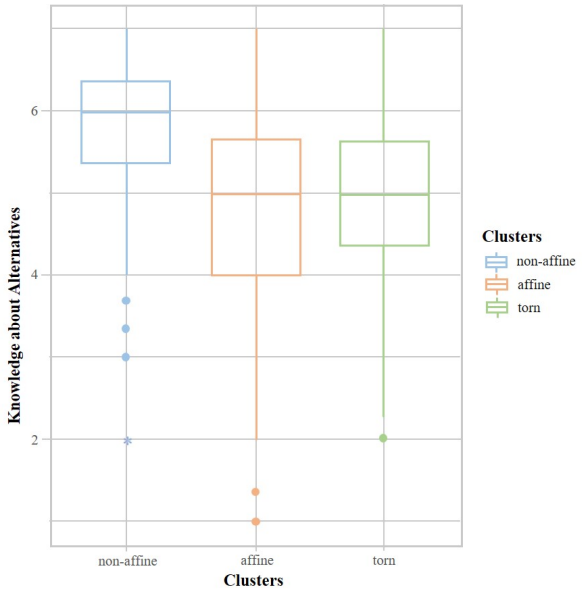

Figure 3. Boxplot for Knowledge about Alternatives

The segmentation of clusters based on general privacy concerns was performed to see if there is any deviance in the way different users were concerned about online providers' practices. The factor included concerns about keeping private information in a non-secure manner, not taking measures to prevent unauthorized access to user information, divulging user information to unauthorized parties without user consent, using and selling user information for other purposes without authorization or notification, and using user information for other purposes. From the descriptive statistics, it was already evident that all participants were rather concerned. Thus, the data were skewed and therefore were not normal. This was confirmed by a Shapiros test. Lavenes test indicated

homogeneity of variances between clusters. Thus, a Kruskal-Wallis and an ANOVA test were performed and indicated significance between the clusters ($p < 0.01$). To visualize the differences, a box plot is provided in Figure 4. It can be seen that non-affine users are the most worried, which is in accordance with their negative attitude toward Google. In comparison to the other two clusters, affine users are less worried. Given that the

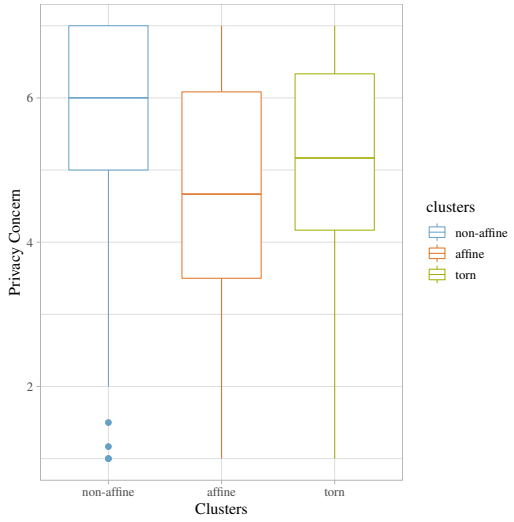

Figure 4. Boxplot for General Privacy Concern

three **behavioral items** related to Google services, i.e. *"How often do you use a different browser than Google Chrome?"*, *"How often do you use a different search engine than Google Search?"*, and *"How often do you avoid to be signed in to your personal Google account before using Google service that don't require to be signed in?"* were not used in the final cluster analysis, differences were tested for these items. Both chi-square and Kruskal-Wallis tests were performed and yielded statistical significance ($p < 0.01$) for all three items). Furthermore, the post-hoc Bonferroni test also showed significant differences across all clusters for all variables ($p < 0.01$). There were no unexpected patterns found in the contingency tables for these variables. Non-affine users tend to use browsers other than Google Chrome and search engines other than Google Search, as well as sign out of their Google account. Affine users behave the opposite, while torn users do not commit to using either different services or Google services.

6. Discussion

6.1. Key Findings

Our results indicate three clusters. The first cluster often uses alternative services to those offered by Google. Cluster members mainly come from the United Kingdom (UK) and the United States (US) and, the fewest, from France. Members rate themselves highly knowledgeable about information privacy. However, the objective multiple choice test demonstrated that the cluster lacks knowledge about data collection in contrast to data integration and data usage. Moreover, the cluster claims to largely know alternatives to Google services. Members demonstrate high privacy concerns and a high-level of using alternative services to Google which indicates that this cluster is characterized by a low degree of privacy paradox.

The second cluster does not use alternative services to those offered by Google. Cluster members stem from Denmark and the fewest are from France. This cluster has the lowest average age. While cluster two lacks knowledge about data collection and integration, it performed good about data usage. However, compared to cluster one, cluster two is significantly less knowledgeable about data integration and usage. Moreover, the cluster does somewhat know alternatives to Google services. As it demonstrates a medium level of privacy concerns and a low level of using alternative services this cluster is characterized by a medium degree of privacy paradox.

The third cluster somewhat uses alternative services to those provided by Google. Cluster members mainly come from France and the fewest from Germany and Denmark. This cluster has the highest average age. While cluster three lacks knowledge about data collection and integration, it performed good about data usage. However, compared to cluster one, cluster three is significantly less knowledgeable about data integration and usage (just as cluster two). Moreover, cluster two only somewhat knows alternatives to Google services. As it demonstrates high privacy concerns and medium levels of using alternative services this cluster is characterized by a medium degree of privacy paradox.

6.2. Theoretical and Practical Implications

Our study makes two theoretical contributions to the privacy literature. The results show that all clusters have medium to high general privacy concerns and low to high levels of using alternatives to Googles services. Hence, the degree of privacy paradox is not primarily influenced by variations in privacy concerns (or variations in disagreeing with Google's practices according to [12]), but largely by variations in the use of alternative services (the actual behavior). As a consequence, we demonstrate that the privacy paradox can exhibit varying degrees and is not a dichotomous phenomenon. Second, we extend research

on information asymmetries [10] by demonstrating that objective knowledge about data integration and data usage as well as knowledge about alternative services are new areas of information asymmetries that contribute to consumers privacy paradox.

The second theoretical contribution also triggers two practical implications for regulators. First, regulators need to enforce online service provider to better inform consumers about their data integration and data usage practices. To cope with these regulations, service providers usually develop more transparent privacy policies. However, as consumers' dont read privacy policies , we argue that the enforcement should focus on triggering service provider to develop new tools instead. Tools that can be easily accessible, readable, and comprehensible by consumers such as visual signs on the initial screen (e.g. certifications or warning labels). Second, as Google can easily deny its competitors access to customers (e.g. by pre-installing Search on Android or setting it as default on Chrome), consumers are dissuaded from finding, and therefore knowing about, alternative services. We encourage regulators to level the playing field (such as triggering Google to allow other search engine to be available during Android setups) and help consumers get to know alternative services.

7. Conclusion

Consumers indicating a high level of privacy concern while simultaneously neglecting their privacy and data disclosure protections are defined as being caught in the privacy paradox [7]. However, prior work neglected to investigate (1) the characteristics of consumers caught in the privacy paradox, (2) new areas of information asymmetries such as knowledge about alternative services, and (3) new privacy-decision processes such as quitting privacy-invasive platform operators. To close these gaps, we conducted a representative segmentation study of Google and its services across five countries guided by the theory of planned behavior. Our results identify three clusters and indicate that the privacy paradox is only prevalent in two of them. Consumers in these two clusters lack knowledge about data integration, data usage, and alternative services.

We contribute to the privacy literature by identifying clusters with varying degrees of the privacy paradox (in contrast to assuming that it is a dichotomous phenomenon) and by demonstrating that knowledge about alternative services to privacy-invasive ones are a new area of information asymmetry that contributes to consumers privacy paradox.

Our study has several limitations. First, the results reflect consumer attitudes towards Google which limits the generalizability of our findings. We encourage future research to explore other cases such as Facebook to enhance the generalizability of our findings. Second, the research context is limited to the US and some European countries. Thus, results might differ when assessing other countries or continents such as Asia. Third, the authors decided that three clusters were most suitable to make sense of the data. However, statistical tests also identified other cluster solutions and therefore, our results might differ when choosing a different number of clusters. Lastly, we encourage future research to assess the effect of regulation on the lack of knowing alternatives. Especially the recent regulation of Google, which forces the company to allow other search engines to be selectable as default when initially setting up an Android phone, reflects a promising case.

References

[1] S. Conger, J. H. Pratt, and K. D. Loch, "Personal information privacy and emerging technologies," *Information Systems Journal*, vol. 23, no. 5, pp. 401–417, 2013.

[2] H. Smith, T. Dinev, and H. Xu, "Information privacy research: An interdisciplinary review," *MIS Quarterly*, vol. 35, no. 4, pp. 989–1016, 2011.

[3] H. Yun, G. Lee, and D. J. Kim, "A chronological review of empirical research on personal information privacy concerns: An analysis of contexts and research constructs," *Information and Management*, vol. 56.

[4] P. Lowry, T. Dinev, and R. Willison, "Why security and privacy research lies at the centre of the information systems (is) artefact: Proposing a bold research agenda," *European Journal of Information Systems*, vol. 26, no. 6, pp. 546–563, 2017.

[5] Y.-A. De Montjoye, L. Radaelli, V. Singh, and A. Pentland, "Unique in the shopping mall: On the reidentifiability of credit card metadata," *Science*, vol. 347, no. 6221, pp. 536–9, 2015.

[6] T. H. Davenport, J. G. Harris, G. L. Jones, K. N. Lemon, D. Norton, and M. B. McCallister, "The dark side of customer analytics," *Harvard Business Review*, vol. 85, p. 37, 05 2007.

[7] C. Phang, J. Sutanto, C.-H. Tan, and E. Palme, "Addressing the personalizationprivacy paradox: An empirical assessment from a field experiment on smartphone users," *MIS Quarterly*, vol. 37, no. 4, pp. 1141–1164, 2013.

[8] G. Bansal, F. Zahedi, and D. Gefen, "The impact of personal dispositions on information sensitivity, privacy concern and trust in disclosing health information online," *Decision Support Systems*, vol. 49, no. 2, pp. 138–150, 2010.

[9] R. Chakraborty, C. Vishik, and R. Rao, "Privacy preserving actions of older adults on social media: Exploring the behavior of opting out of information sharing," *Decision Support Systems*, vol. 55, no. 4, pp. 948–956, 2013.

[10] A. Acquisti and J. Grossklags, "Privacy and rationality in individual decision making," *IEEE security & privacy*, vol. 3, no. 1, pp. 26–33, 2017.

[11] S. Kokolakis, "Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon," *Computers & Security*, vol. 64, pp. 122–134, 2017.

[12] S. Hermes, E. K. Clemons, D. Wittenzellner, A. Hein, M. Bhm, and H. Krcmar, "Consumer attitudes towards firms that monetize personal information: A cluster analysis and regulatory implications," *24th Pacific Asia Conference on Information Systems, Dubai, United Arab Emirates*, 2020.

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

[14] F. T. Beke, F. Eggers, and P. C. Verhoef, "Consumer informational privacy: Current knowledge and research directions," *Foundations and Trends in Marketing*, vol. 11, no. 1, pp. 1–71, 2018.

[15] H.-T. Chen, "Revisiting the privacy paradox on social media with an extended privacy calculus model: The effect of privacy concerns, privacy self-efficacy, and social capital on privacy management," *American Behavioral Scientist*, vol. 62, no. 10, pp. 1392–1412, 2018.

[16] S. Trepte, D. Teutsch, P. K. Masur, C. Eichler, M. Fischer, A. Hennhfer, and F. Lind, *Do People Know About Privacy and Data Protection Strategies? Towards the "Online Privacy Literacy Scale" (OPLIS)*, pp. 333–365. Springer, Dordrecht, 01 2015.

[17] R. Bandara, M. Fernando, and S. Akter, "The privacy paradox in the data-driven marketplace," *Procedia Computer Science*, vol. 121, pp. 562–567, 2017.

[18] I. Ajzen, *From intentions to action: A theory of planned behavior*, pp. 11–39. Springer, 1985.

[19] I. Ajzen, "Martin fishbeins legacy: The reasoned action approach," *ANNALS, AAPSS*, vol. 640, 2012.

[20] I. Ajzen, "The theory of planned behavior," *Organizational Behavior and Human Decision Processes*, vol. 50, pp. 179–211, 1991.

[21] I. Ajzen, "The theory of planned behavior," in *Handbook of theories of social psychology* (P. A. M. Lange, A. W. Kruglanski, and E. T. Higgins, eds.), vol. 1, pp. 438–459, Sage, 2012.

[22] E. Kim, S. Ham, I. S. Yang, and J. G. Choi, "The roles of attitude, subjective norm, and perceived behavioral control in the formation of consumers behavioral intentions to read menu labels in the restaurant industry," *International Journal of Hospitality Management*, vol. 35, pp. 203–213, 2013.

[23] P. A. Pavlou and M. Fygenson, "Understanding and prediciting electronic commerce adoption: an extension of the Theory of Planned Behavior," *MIS Quarterly*, vol. 30, pp. 115–143, 2006.

[24] I. Ajzen, "Constructing a Theory of Planned Behavior Questionnaire: Conceptual and Methodological Considerations," working paper, University of Massachusetts, 2002.

[25] S. Zuboff, *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. London, UK: Profile Books, 2019.

[26] E. K. Clemons, *New Patterns of Power and Profit: A Strategist's Guide to Competitive Advantage in the Age of Digital Transformation*. Switzerland: Palgrave Macmillan, 2018.

[27] R. F. Jorgensen and T. Desai, "Right to privacy meets online platforms: Exploring privacy complaints against facebook and google," *Nordic Journal of Human Rights*, no. 2, pp. 106–12, 2017.

[28] M. Conner and P. Sparks, "Theory of planned behaviour and health behaviour," in *Predicting health behaviour*, Open University Press, 2005.

[29] H. Xu and H. Teo, "Alleviating consumers' privacy concerns in location-based services: A psychological control perspective.," *Proceedings of the 25th International Conference on Information Systems*, pp. 793–806, 2004.

[30] B. Morrison, "Do we know what we think we know? an exploration of online social network users' privacy literacy," *Paper presented at the Proceedings of the 42nd Atlantic Schools of Business Conference, Halifax, Nova Scotia*, 2012.

[31] J. P. Carlson, D. Hardesty, and W. Bearden, "Influences on what consumers know and what they think they know regarding marketer pricing tactics," *Psychology & Marketing*, vol. 24, pp. 117–142, 2007.

[32] G. Lubke and B. Muthen, "Factor-analyzing likert-scale data under the assumption of multivariate normality complicates a meaningful comparison of observed groups or latent classes," 2002.

[33] H. F. Kaiser, "A second generation little jiffy," *Psychometrika*, pp. 401–415, 1970.

[34] H. F. Kaiser, "The application of electronic computer to factor analysis," *Educational and Psychological Measurement*, pp. 141–151, 1960.

[35] J. L. Horn, "A rationale for the number of factors in factor analysis," *Psychometrika*, pp. 179–185, 1965.

[36] A. B. Costello and J. Osborne, "Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis," *Practical Assessment, Research & Evaluation*, vol. 10, pp. 1–9, 2005.

[37] F. J. Floyd and K. F. Widaman, "Factor analysis in the development and refinement of clinical assessment instruments," *Psychological Assessment*, pp. 286–299, 1995.

[38] "Xgboost." `https://github.com/dmlc/xgboost`, 2020. Accessed: 2020-03-12.

8. Appendix

GC stands for Google Chrome. GS for Google Search. SI for being signed-in into one's Google account. SK for subjective knowledge. BI for behavioral intention. ATT for attitude. SN for subjective norm and II for injunctive influence and DI for descriptive influence. BB for behavioral beliefs and BS for belief strength and EoO for evaluation of outcome. PBC for perceived behavioral control. NB for normative belief and MtC for motivation to comply. CB for control beliefs and BP for belief power. PC for privacy concerns.

	PA1	PA3	PA2	PA7	PA11	PA8	PA12	PA5	PA6	PA10	PA13	PA9	PA4	PA15	PA14
SK1	0.04	-0.03	-0.02	0.03	-0.01	-0.01	-0.02	0.01	-0.04	0.02	0.09	0.02	**0.80**	0.05	-0.01
SK2	0.02	0.01	0.05	0.00	-0.04	-0.01	0.04	-0.00	-0.04	0.01	0.00	-0.01	**0.91**	0.02	-0.02
SK3	0.01	-0.00	0.03	0.04	-0.07	-0.01	0.02	-0.01	-0.01	0.01	0.08	-0.05	**0.91**	0.04	0.00
BI-GC1	0.01	-0.01	-0.02	**0.81**	0.02	0.01	-0.04	-0.01	-0.12	0.25	-0.04	0.00	0.03	-0.02	-0.02
BI-GC2	0.01	-0.00	-0.03	**0.79**	0.05	0.05	-0.02	0.01	-0.12	0.22	-0.03	-0.04	-0.00	-0.01	-0.04
BI-GC3	-0.01	0.00	-0.04	**0.79**	0.05	-0.01	-0.05	0.03	-0.12	0.27	-0.04	0.00	-0.00	-0.04	0.00
BI-GS1	0.02	-0.01	0.01	**0.72**	0.07	0.07	-0.08	0.04	0.25	-0.13	-0.04	0.06	0.01	-0.03	-0.03
BI-GS2	0.05	0.01	-0.03	**0.79**	0.08	-0.03	-0.04	0.02	0.24	-0.17	-0.02	0.00	0.01	-0.01	0.02
BI-GS3	0.01	0.00	-0.01	**0.74**	0.07	0.03	-0.02	0.03	0.21	-0.14	-0.04	0.01	0.04	-0.02	0.03
BI-SI2	-0.05	0.04	0.01	**0.46**	0.11	0.06	0.35	0.03	-0.10	-0.07	-0.08	0.06	-0.00	-0.06	0.03
ATT-GC1	-0.01	0.02	0.03	0.13	-0.05	-0.05	0.08	0.01	0.14	**0.71**	0.07	-0.01	0.03	-0.02	-0.00
ATT-GC2	-0.01	-0.02	0.04	0.04	-0.07	0.02	0.08	0.01	0.15	**0.77**	-0.05	-0.03	0.05	-0.02	0.04
ATT-GC3	0.04	0.04	0.03	0.02	-0.01	-0.13	0.09	0.01	0.25	**0.60**	0.07	0.05	-0.04	0.07	-0.01
ATT-GC4	-0.04	-0.01	0.01	0.04	0.04	0.06	0.13	-0.00	0.14	**0.70**	-0.07	-0.00	-0.02	0.02	0.00
ATT-GS1	0.04	-0.03	-0.01	0.07	-0.05	-0.02	0.19	0.00	**0.74**	0.12	0.12	-0.00	-0.00	-0.00	-0.03
ATT-GS2	0.01	-0.03	0.01	0.01	-0.06	0.08	0.19	0.02	**0.71**	0.17	0.02	-0.02	-0.00	-0.01	0.01
ATT-GS3	0.01	0.02	-0.02	0.02	-0.06	-0.07	0.20	0.01	**0.70**	0.16	0.13	0.02	-0.06	0.07	-0.01
ATT-GS4	0.01	-0.03	-0.00	-0.02	0.01	0.10	0.21	0.02	**0.67**	0.16	-0.07	-0.00	-0.03	0.01	-0.02
ATT-SI1	-0.01	0.02	-0.00	-0.00	-0.02	-0.03	**0.80**	-0.04	0.18	0.08	0.06	-0.02	0.01	-0.00	0.03
ATT-SI2	0.01	-0.00	0.01	-0.07	0.03	0.03	**0.80**	-0.05	0.19	0.11	-0.04	-0.02	0.05	-0.02	-0.02
ATT-SI3	0.00	0.03	-0.01	-0.01	-0.00	-0.13	**0.74**	-0.01	0.21	0.08	0.11	0.01	-0.03	0.06	-0.01
ATT-SI4	0.04	-0.03	0.01	-0.08	0.13	0.06	**0.77**	-0.05	0.18	0.07	-0.12	-0.01	0.01	-0.02	-0.06
SN-II-GC	-0.06	0.03	0.08	0.19	**0.84**	-0.15	0.02	-0.10	-0.07	0.05	0.08	-0.00	0.02	0.07	0.01
SN-II-GS	-0.03	0.04	0.07	0.14	**0.85**	-0.18	0.02	-0.06	0.05	-0.08	0.08	0.01	0.00	0.07	-0.00
SN-II-SI	-0.09	0.03	0.09	0.11	**0.79**	-0.12	0.16	-0.06	-0.07	-0.07	0.09	0.01	0.04	0.03	0.06
SN-DI-GC	0.10	-0.02	0.01	0.00	**0.69**	0.14	-0.09	0.01	-0.11	0.16	0.10	-0.05	-0.11	0.08	-0.06
SN-DI-GS	0.09	-0.04	-0.02	-0.05	**0.71**	0.09	-0.05	0.05	0.12	-0.07	0.10	-0.00	-0.07	0.04	-0.04
SN-DI-SI	0.15	-0.02	-0.04	-0.07	**0.69**	0.11	0.16	0.03	-0.11	-0.06	0.08	-0.04	-0.04	0.01	-0.05
BB-BS-GC1	0.07	**0.76**	0.01	0.01	0.03	0.05	-0.04	-0.07	0.06	0.00	0.06	-0.05	0.02	-0.02	0.05
BB-BS-GC2	0.00	**0.65**	0.00	-0.13	0.17	-0.08	-0.06	0.10	0.05	-0.04	-0.03	-0.00	0.04	-0.03	0.01
BB-BS-GC3	-0.03	**0.74**	0.03	-0.03	0.09	-0.05	-0.12	0.05	0.04	-0.04	-0.02	0.03	0.01	-0.01	-0.05
BB-BS-GS1	0.05	**0.78**	-0.04	0.06	-0.04	0.04	0.02	-0.05	0.05	-0.02	0.01	-0.04	0.00	0.01	0.05
BB-BS-GS2	-0.00	**0.71**	0.00	-0.05	0.04	-0.06	-0.04	0.03	-0.13	0.09	-0.02	-0.01	0.01	0.05	-0.09
BB-BS-GS3	0.11	**0.74**	0.01	0.01	-0.03	0.04	-0.02	-0.04	0.04	0.06	0.01	-0.03	0.00	-0.03	-0.01
BB-BS-SI1	-0.07	**0.81**	-0.00	0.06	-0.15	0.11	0.14	-0.01	-0.05	-0.01	0.04	-0.04	-0.03	-0.01	0.05
BB-BS-SI2	-0.01	**0.80**	0.01	0.07	-0.10	0.05	0.09	0.01	-0.04	-0.05	0.02	0.01	-0.05	0.02	-0.01
BB-BS-SI3	-0.03	**0.80**	-0.04	0.04	-0.03	0.03	0.09	-0.03	-0.04	-0.02	-0.00	0.05	-0.03	0.03	0.01
BB-EoO-GC/GS/SI1	0.13	0.03	0.13	0.04	-0.19	-0.05	0.01	-0.04	-0.06	-0.06	0.17	**0.58**	-0.04	0.12	-0.05
BB-EoO-GC2	-0.02	0.02	-0.07	-0.08	0.16	0.10	-0.07	0.03	0.07	0.07	-0.14	**0.46**	0.08	-0.16	0.00
BB-EoO-GC3	-0.05	0.02	-0.04	-0.06	0.14	0.10	-0.05	0.01	0.02	0.08	-0.12	**0.56**	0.00	-0.07	0.06
BB-EoO-GS2	-0.10	0.05	-0.11	-0.04	0.24	0.03	-0.10	0.04	0.08	0.10	-0.15	**0.46**	0.02	-0.10	0.07
BB-EoO-GS3	0.09	-0.04	0.14	0.04	-0.13	-0.07	0.02	-0.02	-0.03	-0.04	0.12	**0.59**	-0.07	0.10	-0.06
BB-EoO-SI2	0.01	-0.05	0.01	0.04	0.03	-0.07	0.10	-0.02	0.00	-0.09	0.06	**0.69**	-0.02	0.09	-0.02
BB-EoO-SI3	0.03	-0.01	0.02	0.05	-0.07	0.01	-0.03	-0.02	-0.04	0.00	0.15	**0.65**	-0.01	0.04	-0.01
PBC-GC1	0.01	0.03	0.04	0.09	-0.07	**0.72**	-0.11	-0.02	-0.06	0.21	0.04	-0.01	-0.04	0.06	0.00
PBC-GS1	0.02	0.03	0.06	0.09	-0.05	**0.72**	-0.10	-0.04	0.22	-0.09	-0.03	0.02	-0.01	0.06	0.03
PBC-SI1	-0.07	0.02	0.02	-0.02	-0.02	**0.66**	0.28	0.01	-0.11	-0.07	0.00	0.09	0.00	-0.01	0.03
PBC-GC2	0.09	-0.04	0.02	0.08	-0.09	**0.71**	-0.13	-0.04	-0.04	0.14	0.14	-0.04	-0.01	0.08	-0.02
PBC-GS2	0.04	-0.00	0.06	0.05	-0.03	**0.74**	-0.10	-0.04	0.26	-0.12	0.10	-0.04	0.01	0.07	-0.03
PBC-SI2	0.07	0.01	0.01	-0.12	0.03	**0.71**	0.26	0.01	-0.09	-0.09	0.08	-0.00	0.03	-0.04	-0.04
NB-BS-GC	-0.03	0.00	-0.07	-0.06	0.17	0.10	-0.06	0.04	0.08	0.05	**0.89**	0.02	0.06	-0.13	0.01
NB-BS-GS	-0.05	0.02	-0.07	-0.06	0.18	0.09	-0.05	0.05	0.12	-0.02	**0.89**	0.06	0.06	-0.13	0.02
NB-BS-SI	-0.04	0.04	-0.06	-0.07	0.13	0.09	0.11	0.04	-0.02	-0.04	**0.77**	0.04	0.04	-0.09	0.06
NB-MtC-GC	0.10	-0.03	0.02	0.01	-0.05	0.01	-0.06	-0.01	-0.02	0.06	0.04	-0.02	-0.01	0.02	**0.87**
NB-MtC-GC	0.12	-0.02	0.02	-0.01	-0.01	0.02	-0.06	0.00	0.04	-0.02	0.04	-0.02	-0.01	0.03	**0.83**
NB-MtC-SI	0.09	0.00	0.01	-0.03	0.01	-0.05	0.07	0.00	-0.09	0.01	0.04	0.03	-0.02	0.05	**0.81**
CB-BP-GC1	**0.63**	-0.01	-0.03	0.00	0.14	0.02	-0.10	-0.04	-0.07	0.17	0.04	0.03	0.02	0.02	0.00
CB-BP-GC2	**0.68**	-0.02	-0.01	0.04	0.00	0.02	-0.08	-0.03	-0.10	0.20	0.11	-0.03	-0.01	0.01	0.03
CB-BP-GC3	**0.68**	0.04	0.02	-0.06	0.00	0.01	-0.03	0.07	-0.01	0.09	0.03	0.02	-0.02	0.03	-0.01
CB-BP-GS1	**0.85**	-0.00	0.01	0.01	0.08	0.01	-0.08	-0.03	0.12	-0.05	-0.10	-0.01	0.04	-0.03	0.01
CB-BP-GS2	**0.84**	-0.00	0.02	0.05	0.01	-0.00	-0.11	0.01	0.17	-0.09	-0.02	-0.02	-0.02	0.01	-0.00
CB-BP-GS3	**0.86**	-0.00	0.03	-0.01	0.04	-0.00	-0.04	0.00	0.13	-0.12	-0.01	-0.06	-0.01	0.03	0.01
CB-BP-SI1	**0.76**	0.02	-0.07	-0.01	-0.04	0.01	0.17	0.01	-0.06	-0.01	-0.08	0.08	0.04	-0.08	0.01
CB-BP-SI2	**0.80**	0.02	-0.01	0.05	-0.05	-0.01	0.19	-0.02	-0.08	-0.06	-0.10	0.05	0.04	-0.09	0.02
CB-BP-SI3	**0.79**	0.04	-0.03	-0.02	-0.07	-0.01	0.19	0.03	-0.04	-0.05	-0.05	-0.00	-0.01	-0.03	0.03
CB-BS-GC	-0.09	0.01	-0.06	0.05	0.04	0.13	-0.06	0.02	-0.01	0.08	-0.05	0.01	-0.03	**0.82**	-0.00
CB-BS	0.02	0.04	0.05	-0.10	0.02	-0.12	-0.02	-0.01	0.02	0.01	-0.03	0.04	0.04	**0.53**	0.03
CB-BS-GS	-0.02	-0.02	-0.08	0.04	0.06	0.12	-0.02	0.01	0.05	-0.06	-0.11	-0.01	0.01	**0.85**	0.03
CB-BS-SI	-0.02	-0.02	-0.06	-0.07	0.12	0.12	0.12	0.05	0.00	-0.01	-0.15	0.03	0.08	**0.61**	0.01
PC1	0.01	-0.03	**0.91**	0.00	0.05	-0.02	-0.01	0.02	0.03	0.03	-0.05	-0.02	0.00	-0.03	0.01
PC2	-0.02	0.00	**0.84**	-0.03	-0.00	0.04	0.01	0.06	-0.02	0.01	-0.03	0.01	0.05	-0.05	0.04
PC3	-0.01	-0.01	**0.93**	-0.04	0.03	0.06	-0.02	0.02	-0.01	0.05	-0.05	-0.01	-0.01	-0.03	0.02
PC4	-0.01	0.01	**0.93**	0.00	0.06	0.04	-0.01	-0.03	-0.02	0.03	-0.05	0.02	0.00	-0.01	-0.00
PC5	-0.04	0.03	**0.94**	-0.03	0.01	0.09	0.02	-0.01	-0.00	-0.00	-0.04	0.01	0.03	-0.04	-0.02
PC6	0.01	-0.00	**0.87**	-0.02	0.09	0.01	0.01	0.03	0.01	0.01	-0.03	-0.00	-0.02	-0.01	-0.01

Table 5. Factor Loadings