Regular Paper

The Effects of Nudging a Privacy Setting Suggestion Algorithm's Outputs on User Acceptability

Toru Nakamura[1,a) Andrew A. Adams[2] Kiyoshi Murata[2] Shinsaku Kiyomoto[3]
Nobuo Suzuki[1,4]

Received: March 12, 2019, Accepted: September 11, 2019

Abstract: In prior work, a machine learning approach was used to develop a suggestion system for 80 privacy settings, based on a limited sample of five user preferences. Such suggestion systems may help with the user-burden of preference selection. However, such a system may also be used by a malicious provider to manipulate users' preference selections through nudging the output of the algorithm. This paper reports an experiment with such manipulation to clarify the impact and users' resistance of or susceptibility to such manipulation. Users are shown to be highly accepting of suggestions, even where the suggestions are random (though less so than for nudged suggestions).

Keywords: privacy protection, machine learning, recommendation, nudge effect

1. Introduction

Much of daily life and the economy of the Information Age is dependent on the collecting, processing and use of data about individuals. This creates concern by users about their privacy (misuse of data by organisations who legitimately hold the data) and security (access to the data by individuals or organisations without legitimate authorisation). Part of the response of service providers to the users' concerns and regulations has been to define privacy policies declaring their intentions. In addition, users are often provided with some way of providing or withholding consent to usage or sharing of specific data in specific circumstances. Facebook allows users to limit who else can see the messages posted in their account. Smartphone systems such as Apple's iOS and Google's Android provide controls on what data installed applications may access. Using these controls, however, creates a burden on users [1]. Various systems have been proposed to provide a coherent way for users to express their preferences regarding use of their personal data, such as PDS (Personal Data Store) [2] and PPM (Privacy Policy Manager) [3]. Even though the centralisation of controls in such systems may reduce user burden, that burden remains significant.

One method of reducing this burden still further is the use of machine learning techniques to provide suggestions to users for all of their settings, based on their answers to an indicative subset. In Ref. [4] a system to provide settings for 80 options (16 types of data crossed by five types of usage) was developed using an SVM model based on 9,000 learning responses and 1,000 test responses. In prior testing of this model with users, acceptance rates of over 90% have been observed when users are asked to accept or change the model's suggestions. However, the model has also been observed to match at only approximately 65% with unprompted answers [5]. This raises the question of how much users are prone to accepting suggestion, and whether malicious service providers could abuse this to influence users to make privacy choices which benefit the platform as opposed to being true reflections of users' wishes [6]. Would nudging the results of the algorithm reduce the acceptance rate by users, and if so by how much? Would some users be suspicious of such nudged algorithms?

In this paper, the results of an experiment with the model from Ref. [4] is reported. Participants were split into four groups and each answered five predictor questions and then changed or accepted provided suggestions for another 75 settings. The suggestions for the four groups were: (1) Original Model; (2) Model with Suggestions Nudged to Privacy; (3) Model with Suggestions Nudged to Sharing; (4) Random Suggestions.

Participants also completed a survey about the system, giving their opinion of both the system's accuracy and their emotional response to the suggestions.

1.1 Ethical Considerations

The goals of the experiment, to compare people's acceptance of suggestions for privacy settings with various types of suggestion, required the use of a *deception experiment*. This kind of experiment is accepted practice, but must always be developed and conducted with consideration of the ethics of deceiving participants. In this case, there was no valid way to conduct the experiment without the deception. The deception itself was judged not to be harmful to the participants, since the choices are for a prototype system and not a system in actual use. Finally, participants had the deception and its justification clearly explained to them after they had participated and were given the option of

[1] Advanced Telecommunications Research Institute International, Kyoto 619–0288, Japan
[2] Meiji University, Chiyoda, Tokyo 101–8301, Japan
[3] KDDI Research, Inc., Fujimino, Saitama 356–8502, Japan
[4] Kindai University, Iizuka, Fukuoka 820–8555, Japan
[a] tr-nakamura@atr.jp

withdrawing their data from the experiment after the explanation (none did). As is standard practice in Japan, participants were compensated for their time spent (¥3,000) and this payment was made regardless of willingness to allow their results to be used after the deception was explained. None of the participants complained about the deception, though a few remarked that it made clear why the suggestions had seemed odd.

1.2 Contributions

Neither the shift towards privacy or the shift towards sharing produced a statistically significant difference in the proportion of accepted suggestions. The acceptance rate for the machine-learning based model were much higher (statistically significantly so) than for the random model, although the random model also generated clearly (and statistically significantly) higher acceptance rate than its randomness should give if users were unaffected by the suggestion. This suggests that platforms which provide suggestions for privacy preferences can easily push users towards openness or sharing, in turn suggesting that such activity should be considered for regulation.

1.3 Construction

Section 2 describes the generic privacy setting suggestion scheme using an SVM machine learning approach, followed in Section 3 by a discussion of related work. Section 4 describes the specific suggestion scheme used in this experiment. Section 5 describes the detail of this experiment. Section 6 presents the results from the experiment with the suggestion system, while Section 7 shows the results from a survey of participants after their experience. Finally, Section 8 gives some conclusions and pointers to possible further work.

2. Privacy Setting Suggestion Scheme

The experiment presented in this paper is based specifically upon the suggestion model presented in Ref. [4]. The detail is described in Appendix A.1. That model is, however, simply a specific instantiation of a more general use case presented here.

2.1 Generic Use Case for Privacy Preference Suggestions

The generic use case of privacy setting prediction is shown in **Fig. 1**.

(1) Model generation: Based on the existing privacy preference choices of a large number of users an SVM system is used to identify a representative small set of choices, from which the remaining preferences can be reasonably accurately predicted. The values of thee "predictor" items define a feature vector.

(2) Recommendation: New Users, or users wishing to amend their selections, make their selections of the predictor items. The matching set of remaining items from the model is then presented to the user for acceptance or alteration.

2.2 Suggestion Generation Algorithm Production

As shown in **Fig. 2**, this is the process for generating a privacy suggestion system:

(1) Divide users' data into training data (approx. 90% of users) and test data (the remaining approx. 10% of users).

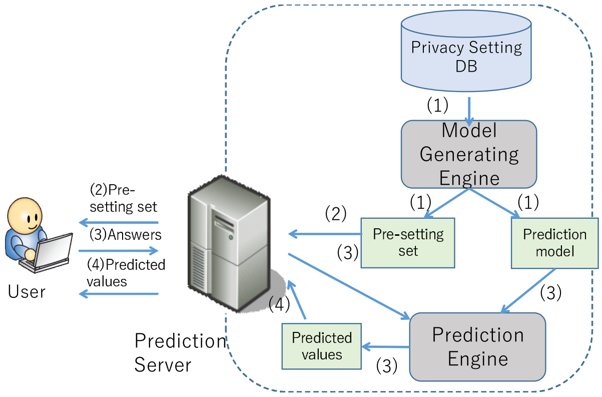


Fig. 1 Use case of privacy setting prediction.

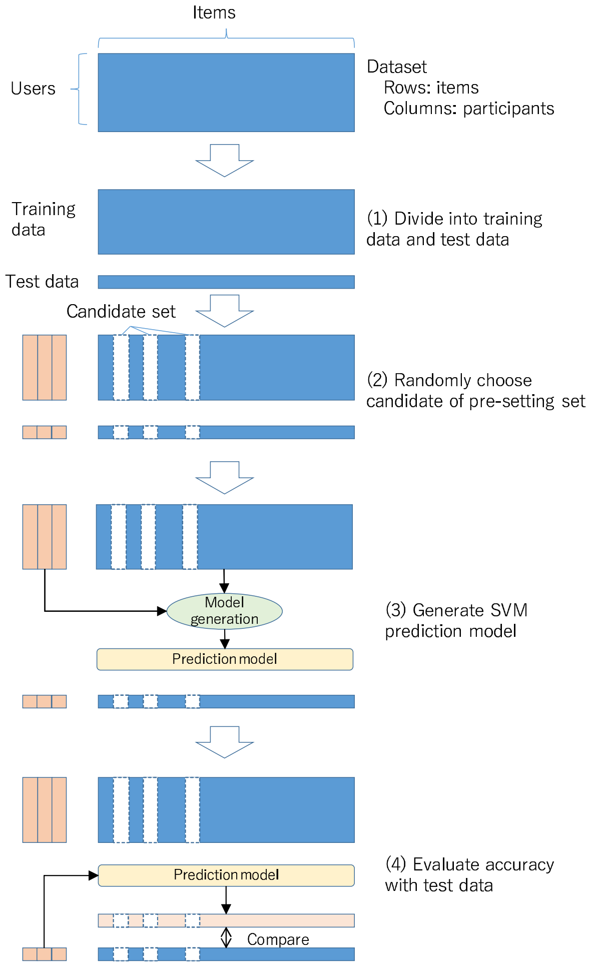

Items
Users
Dataset
Rows: items
Columns: participants
Training data
(1) Divide into training data and test data
Test data
Candidate set
(2) Randomly choose candidate of pre-setting set
Model generation
(3) Generate SVM prediction model
Prediction model
Prediction model
(4) Evaluate accuracy with test data
Compare

Fig. 2 Privacy setting prediction algorithm.

(2) Randomly choose a small number of settings as a candidate for the predictor set.

(3) Generate SVM models corresponding to the remaining choices of each user in the training set, creating a feature model which matches the candidate predictor set with the remaining choices with the highest number of matches across the training set.

(4) Evaluate the accuracy of the candidate predictor set and associated feature model on the test data.

(5) Iterate steps from (2) to (4) a large number of times. Select the predictor set of questions and associated feature model as the suggestion creation algorithm.

3. Related Work

In privacy policy management the burden of creating and maintaining privacy policies has been identified as a major issue. In one study, Madejski et al. [7] showed that a serious mismatch existed between users' a priori intentions with regard to their privacy and their actual settings in an online social network service. Users are commonly required to check the privacy policies of a given service offered by a service provider before starting to use it. Thus, each service provider prepares a privacy policy for each service. Because of the large number of such policies to which users are asked to agree, the burden placed on users becomes significant and many barely bother to check policies or make changes to default settings. In many cases the privacy policy is a one-sided offer with no customisation possible, so if a user does not agree with the privacy policy of a service, the user simply cannot use that service.

In this regard, Solove [8] suggested that the model of privacy self-management cannot achieve its objectives, and it has been pushed beyond its limits, while privacy law has been relying too heavily upon that model to legitimise processing. Moreover, other studies such as that conducted by Acquisti and Grossklags [9] demonstrated users' lack of knowledge about technological and legal forms of privacy protection when confirming privacy policies. Their observations suggest that several difficulties obstruct individuals in their attempts to protect their own private information, even for those users concerned about and motivated to protect their privacy. This was reinforced by Pollach in Ref. [10] whose work also supported the presumption that users are not familiar with technical and legal terms related to privacy. Moreover, it has been suggested that users' knowledge about privacy threats and technologies that help to protect against them is inadequate [11]. In response, Guo and Chen [12] proposed an algorithm to optimise privacy configurations based on the desired privacy and utility preferences of users.

Various languages to describe privacy policies have been created [13], [14], [15]. Backes et al. compared various enterprise privacy policies using formal abstract syntax and semantics to express the policy contents [16]. Tondel and Nyre [17] proposed a similarity metric for comparing machine-readable policies.

Probably the most widely known of these languages is the Platform for Privacy Preferences Project (P3P) [13], [18] which was designed to enable websites to express their privacy practices in a standard format that could be retrieved automatically and interpreted easily by user agents. The project provides user agent modules that allow users to be informed of site practices and to automate decision-making based on these practices when appropriate. Building on the P3P, the Privacy Bird [19], [20] is a web browser extension which automatically retrieves the P3P policies of a web site and compares it with a users' stated preferences. However, Kolter and Pernul [21] suggested that the available privacy settings of the Privacy Bird result in inadequate user acceptance, putting the ultimate goal of real-world use at risk. Thus, they proposed a user-friendly, P3P-based privacy preference generator [21] for service providers, including a configuration wizard and a preference summary.

In practice, P3P and derived interfaces have not been widely adopted by online and offline services [22]. One of the P3P creators, Cranor, reviewed the lack of adoption of P3P in 2012 [23], concluding that a lack of clarity of some of the concepts embedded in the P3P, poor tools for users, and a lack of compelling reasons (regulatory, market-forces, user-demand) for service providers, have led to minimal adoption. P3P's limitation to browsers in a world where a great deal of users' interaction with online services is via specific smartphone apps, has further limited the utility and adoption of P3P.

Yee proposed a privacy policy checker [24] for online services. The checker compares the user privacy policy with the provider privacy policy and then automatically determines whether the service can be used. Biswas proposed an algorithm [25] that detects conflicts in privacy settings between user preferences and the requirements of an application on a smart phone. Privacy Butler [26] is a personal privacy manager that can monitor a person's online presence and attempts to make corrections based on a privacy policy for a user's online presence in a social network. The concept of the Privacy Butler is similar to the concept of the project underlying the experiment reported here, but it focuses on modifications to content hosted by social networking services; it monitors whether the modification is a satisfactory match for the privacy policy.

Privacy Mirror [27] is a tool that is intended to show users what information about them is available online. Srivastava [28], [29] proposed a privacy settings recommender system for a specific online social network service.

Fang et al. [30], [31] have proposed a privacy wizard for social networking sites. The purpose of the wizard is to automatically configure a user's privacy settings with minimal effort required by the user. The wizard is based on the underlying observation that real users conceive their privacy preferences based on an implicit structure. Thus, after asking the user a limited number of carefully chosen questions, it is usually possible to build a machine learning model that accurately predicts the user's preferences.

This approach is very similar to that presented here. The main difference is the target dataset. Fang et al. covered data and connections in Facebook specifically, so the variety of the items was limited and the number of the potential recipients of data is small, and typically comprised of individuals within the users' social graph. The system presented here treats more general data items and the number of the potential recipients is larger because the system does not focus on a specific service such as Facebook.

There is also existing published research about learning privacy preferences. Berendt et al. [32] emphasised the importance of privacy preference generation and Sadah et al. [33] suggested that machine learning techniques have the power to generate more accurate preferences than users themselves in a mobile social networking application. Tondel et al. [34] proposed a conceptual architecture for learning privacy preferences based on the decisions a user makes in their normal interactions on the web. They suggested that such learning of privacy preferences has the potential to increase the accuracy of preferences without requiring users to have a high level of knowledge or any willingness to invest time and effort to protect their privacy. Kelley et al. [35] showed preferences for a mobile social network application. Preference modeling for eliciting preferences was studied by Buffett and Fleming [36]. Mugan et al. [37] proposed a method for generating persona and suggestions intended to help users incrementally refine their privacy preferences over time.

On the other hand, as far as the authors are aware, there is no research which focuses on the possibility of a malicious service provider's manipulating individual users into sharing their personal data with it with juggling its privacy setting suggestions. However, there are related studies to suggest the significance of this research. For example, through experiments to examine whether opt-in and opt-out create differences in terms of individual consumers' intention to be contacted with further health survey and the mechanisms underlying the differences, Johnson et al. [38] found that there are major differences between the two formats, and that defaults had a sizeable effect and even a mild or minimal framing manipulation had a significant impact on consumer choices. Wang et al. [39] set up an experimental Facebook application, and, based on Nissenbaum's idea of contextual integrity [40], investigated a relationship between the relevance of personal data being shared with the application via the default settings to it and individual users' data disclosure behavior. The results of the experiment showed that participants hesitated to install the application when they were required to provide irrelevant or too much personal data, suggesting default privacy settings are not necessarily accepted by individual users. These findings lead us to the question whether biased defaults can be used to manipulate individual user behavior. Compared with those simply-structured experiments to investigate the impacts of defaults on individual privacy settings in restricted contexts, the more complicated experiments conducted in this research are designed so that the question can properly be explored through examining the differences among non-, privacy-, open- and randomly biased defaults from a broader viewpoint.

4. Creation of The Suggestion System

This section presents the specific details of the privacy settings and associated suggestion generation algorithms used in the experiment. The suggestion generation algorithms were based on the user data from Ref. [4].

As reported in Ref. [4], 10,000 subjects indicated their willingness to share 16 types of data (see **Table 1**) for each of five purposes (see **Table 2**), on a six-point Likert scale (strongly disagree/disagree/weakly disagree/weakly agree/agree/strongly agree). An SVM machine learning analysis process with parameters shown in **Table 3** was applied to a learning set of 9,000 of these responses, five of the resulting data/purpose combinations were shown to have a high (over 85%) prediction rate on the other 75 questions when applied to the remaining 1,000 responses as a test set. See **Table 4** for the combinations that form the prediction set.

This prediction algorithm was then used to create a tool to provide suggested sharing settings for a privacy policy management system. Instead of a Likert scale, the privacy policy has three options: Always share; Share with Selected Service Providers Case-by-Case; Never Share, as shown in Fig. 5 (Note: the actual experiment was conducted with Japanese people using Japanese versions of the text shown).

Table 1 Types of personal data.

No.	Data type
1	Addresses and telephone numbers
2	Email addresses
3	Service accounts
4	Purchase records
5	Bank accounts
6	Device information (e.g., IP addresses, OS)
7	Browsing histories
8	Logs on a search engine
9	Personal info (age, gender, income)
10	Contents of email, blog, twitter etc.
11	Session information (e.g., Cookies)
12	Social Info. (e.g., religion, volunteer records)
13	Medical Info.
14	Hobby
15	Location Info.
16	Official ID (national IDs or license numbers)

Table 2 Usage purposes.

No.	Data purpose
A	Providing the service
B	System administration
C	Marketing
D	Behavior analysis
E	Recommendation

Table 3 Parameters of SVM.

#learning data	1000
#test data	9000
#combinations of items	5
γ	0.2
cost	1.0

Table 4 Predictor questions.

ID	Data type	Data usage
14-A	14. Hobby	A. Providing service
4-B	4. Purchase record	B. System administration
15-B	15. Location Info.	B. System administration
12-C	12. Social Info.	C. Marketing
6-E	6. Device information	E. Recommendation

x-y, where x is an identifier of a data type and y is an identifier of a usage purpose, is used in the following discussion to refer to a particular combination.

In this experiment, participants were divided into 4 groups which were asked to accept/change suggestions from four different prediction models. The 4 models are following:

(1) Normal model

(2) Privacy-biased model: predictions are nudged towards privacy

(3) Open-biased model: predictions are nudged towards sharing

(4) Random model

For the normal model, the SVM models were generated with the parameters shown in Table 3, using R and the "e1071" [41] SVM package. The resulting predictor set of data type and usage purpose are shown in Table 4.

The privacy- and open-biased models were generated by adjusting the "class weights" parameter from these predictor questions. The classes are the selections: Never Share, Ask, Always Share. This parameter is a three-valued vector indicating the weight of each class, with a default value of (1, 1, 1), giving equal

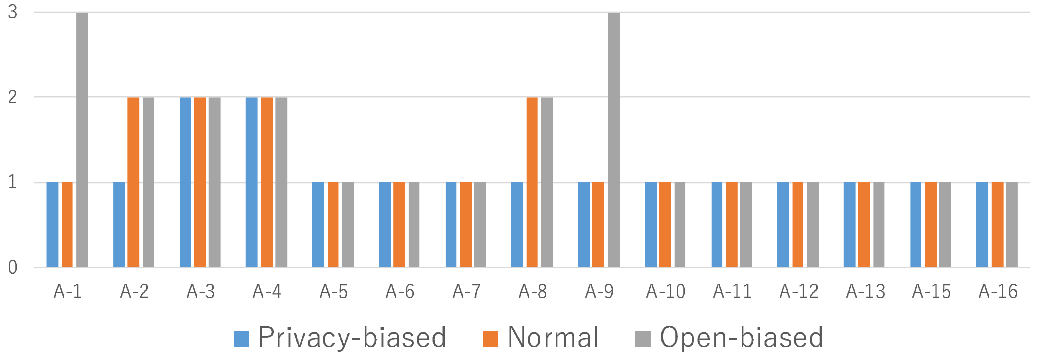


Fig. 3 Suggested values for usage purpose "A. Providing the service" given answers (1,2,3,1,2) for the prediction set.

Table 5 Distributions of suggestion values of normal model, privacy-biased model, and open-biased model.

	1	2	3	Total
Privacy-biased model	77.6%	21.1%	1.30%	100%
Normal model	57.2%	41.5%	1.30%	100%
Open-biased model	55.1%	35.9%	8.98%	100%

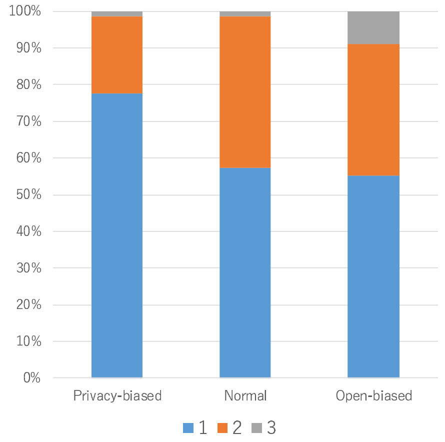


Fig. 4 Distributions of suggestion values of normal model, privacy-biased model, and open-biased model.

weighting to each class. For the privacy-biased model, suggestions are calculated using the weighting parameter (10, 1, 1) for each classes, privileging "1. Never share", over "2. Ask", and "3. Always Share". Similarly, the open-biased model used class weights of (1, 1, 10).

To show how this effects specific suggestions, consider the suggested answers for the sixteen questions in section A ("Providing the service"). For a participant who answered (Never, Ask, Always, Never, Ask) to the predictor questions, the suggestions for the 15 non-predictor questions in section A are shown in **Fig. 3** with each model. This shows the typical kinds of shifts, with the privacy nudge suggesting "Never" in cases where the normal model suggests "Ask" (e.g., for A-2) while the Open-biased model suggests "Always" instead of the normal model's suggestion of "Never" in one case (A-9).

To provide a measure of the strength of the nudging taking place, the percentage of suggested answers for all possible answers to the predictor questions are shown in **Table 5** and **Fig. 4** (1: Never; 2: Ask; 3: Always).

The random model outputs prediction values which are independent of the answers for prediction set. Different random values were generated for each participant in the random model set.

Figure 5 shows a snapshot of the answer page for the prediction questions. First Participant answered these questions. After that the other 75 item are shown with one option pre-selected with the prediction values, shown in **Fig. 6**. Participants were asked to consider all the question and change any of the suggestions they thought did not match their desires.

5. Experiment

This section describes the conduct of the experiment using the system described above.

5.1 Construction of Experiment

Figure 7 shows the process participants followed in the experiment. The experiment system has a list of pairs of an ID and a password. Each ID was assigned to one of the four group (1, 2, 3, and 4) referring to normal model, privacy-biased model, open-biased model, and random model, respectively.

A professional research participant recruitment firm was used to recruit a suitable demographic mix of participants, including age range, gender and employment status.

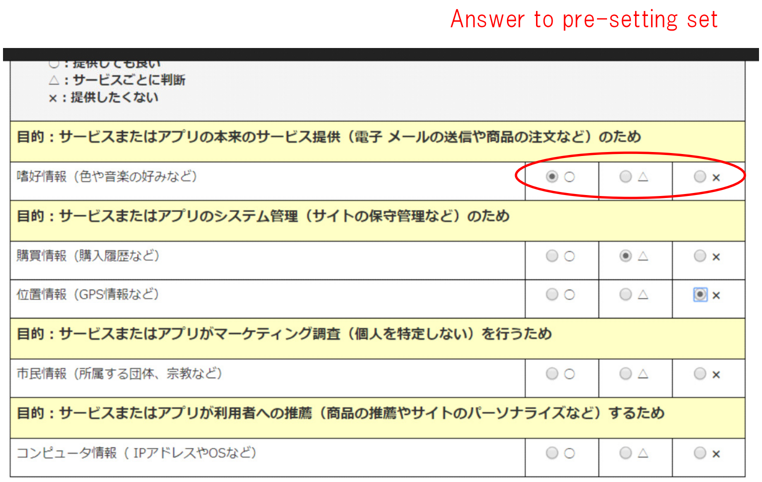
Participants were given their ID and password in advance. First the participants input their ID and password to a log-in form, following which they were shown an explanation of this experiment. This explanation included a description of how to complete the experiment, as well as suitable details about the organization behind the experiment, and the fact that the suggestion system is based on AI techniques. Note that the random group were being misled by this explanation. This is a common necessary untruth in this kind of randomly controlled experiment. The original text of explanation of this experiment is shown in Appendix A.2. The translated version of it in English is also shown in Appendix A.2.

After reading the explanation, the participants gave their unprompted answers to the predictor set of five question, the system calculated its prediction and presented the remaining 75 questions, following which participants accepted the prediction or selected a different answer.

For the rest of the paper these answers are referred to as predicted values and selection values.

5.2 Environment

The experiment was conducted from March 26 to April 2,

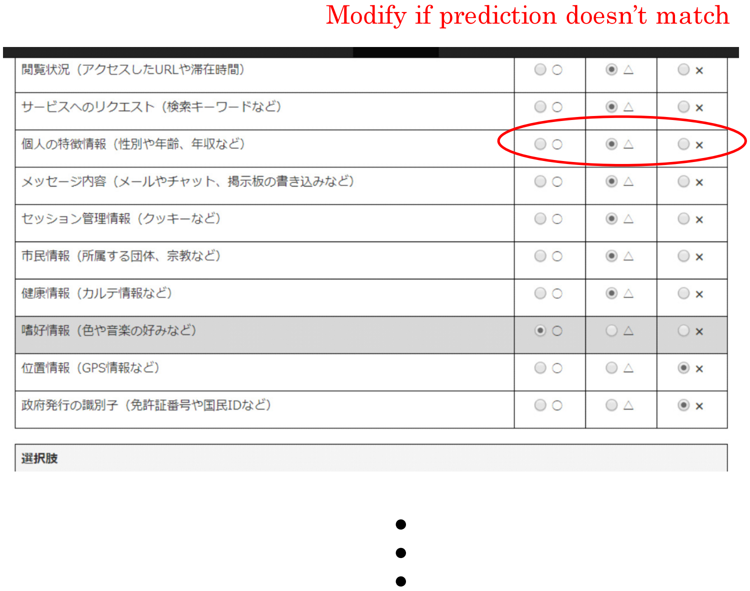


Answer to pre-setting set

○：提供しても良い
△：サービスごとに判断
×：提供したくない

目的：サービスまたはアプリの本来のサービス提供（電子 メールの送信や商品の注文など）のため			
嗜好情報（色や音楽の好みなど）	● ○	○ △	○ ×
目的：サービスまたはアプリのシステム管理（サイトの保守管理など）のため			
購買情報（購入履歴など）	○ ○	● △	○ ×
位置情報（GPS情報など）	○ ○	○ △	● ×
目的：サービスまたはアプリがマーケティング調査（個人を特定しない）を行うため			
市民情報（所属する団体、宗教など）	○ ○	○ △	○ ×
目的：サービスまたはアプリが利用者への推薦（商品の推薦やサイトのパーソナライズなど）するため			
コンピュータ情報（ IPアドレスやOSなど）	○ ○	○ △	○ ×

Fig. 5 Snapshot of answer page for prediction questions.

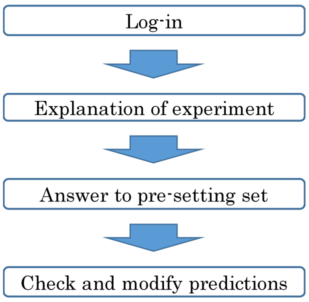


Modify if prediction doesn't match

閲覧状況（アクセスしたURLや滞在時間）	○ ○	● △	○ ×
サービスへのリクエスト（検索キーワードなど）	○ ○	● △	○ ×
個人の特徴情報（性別や年齢、年収など）	○ ○	● △	○ ×
メッセージ内容（メールやチャット、掲示板の書き込みなど）	○ ○	● △	○ ×
セッション管理情報（クッキーなど）	○ ○	● △	○ ×
市民情報（所属する団体、宗教など）	○ ○	● △	○ ×
健康情報（カルテ情報など）	○ ○	● △	○ ×
嗜好情報（色や音楽の好みなど）	● ○	○ △	○ ×
位置情報（GPS情報など）	○ ○	○ △	● ×
政府発行の識別子（免許証番号や国民IDなど）	○ ○	○ △	● ×

選択肢

Fig. 6 Snapshot of result page with prediction values.



Log-in

Explanation of experiment

Answer to pre-setting set

Check and modify predictions

Fig. 7 Construction of experiment.

2018. Participants accessed the web-based experiment system via their own PCs. The number of participants was 552. The distribution of age and gender for participant is shown in **Table 6**.

5.3 Basic Result of Experiment

The basic result of the experiment is shown in **Table 7**. The acceptance rate is the percentage of all 75 non-predictor questions for all members of each group where the participant did not change the suggested answer. This shows that the acceptance rate for the normal model, privacy-biased model and open-biased model shows effectively no difference, while the acceptance rate for random model is much less than the other 3 models. These results are discussed in more detail in the next section.

Table 6 Distribution of participants.

		Group				
Gender	Age	1	2	3	4	Total
Male	25-34	40	41	39	35	155
	35-44	34	29	38	31	132
Female	25-34	43	31	36	44	154
	35-44	37	22	27	25	111
	Total	154	123	140	135	

Table 7 Result of experiment.

Group	Acceptance Rate
1. Normal	89.6%
2. Privacy-biased	91.6%
3. Open-biased	89.2%
4. Random	71.8%

6. Experimental Results

6.1 Comparison Between Groups

Table 10 shows the percentages of suggestions versus selections for each group, giving both the spread of suggestions and selections and the acceptance/rejection of the suggestion. Considering Groups 1–3, the spread of suggestions (right hand total columns) varies considerably between each group, showing a considerable level of variation in the suggestions presented. Nevertheless, the rate of acceptance of suggestions in Groups 1–3 is very similar at almost 90%. The acceptance rate for the random group is markedly lower than that for the groups whose suggestions were based around a model of their preferences.

So, it appears that the nudging had no overall impact on the acceptability of the suggestions to participants. The strength of the nudging experienced by participants can be measured by looking at the differences between the suggestion that were generated from their predictor questions by the "normal model" and the nudged models. This is shown in **Table 8**. As expected the variation between suggestions from the "normal model" and the "random model" is approximately two-thirds. The Privacy-nudged (Group 2) were given a different suggestion in just over one third of cases, while the Open-nudged (Group 3) were given a different suggestion in one quarter of cases. A more detailed breakdown of the suggestions and selections for each group is given in **Table 9** showing the acceptance and alternate selections for each sharing choice (1: Never; 2: Ask; 3: Always). In particular, the widely different results for Group 4 (Random) show that using a model which gives suggestions which are broadly close to participants' instincts produces a much higher acceptance rate than the random model.

Table 8 Distribution of each group.

	Group 2	Group 3	Group 4
Acceptance	0.918	0.888	0.704
Difference between nudged/random and normal model	0.374	0.25	0.638

Table 9 Distribution of each group.

	Group 2	Group 3	Group 4
Normal suggestion	0.106	0.177	0.324
Nudged/random suggestion	0.87	0.765	0.597
Other option	0.024	0.0579	0.079

6.2 Statistical Tests on the Differences in Recommendation Acceptance

Single factor analyses of variance were conducted to ver-

Table 10 Distribution of each group.

Group 1 (Normal)

	Selected_1	Selected_2	Selected_3	Total
Suggested_1	33.7%	0.339%	0.105%	34.1%
Suggested_2	7.05%	49.4%	2.6%	59.1%
Suggested_3	0.0887%	0.234%	6.52%	6.85%
Total	40.8%	50%	9.23%	100%
Acceptance	33.7%	49.4%	6.52%	89.62%

Group 2 (Privacy-biased)

	Selected_1	Selected_2	Selected_3	Total
Suggested_1	57.1%	2.04%	0.823%	60%
Suggested_2	2.5%	29%	2.21%	33.8%
Suggested_3	0.0496%	0.526%	5.67%	6.25%
Total	59.7%	31.6%	8.71%	100%
Acceptance	57.1%	29%	5.6%	91.7%

Group 3 (Open-biased)

	Selected_1	Selected_2	Selected_3	Total
Suggested_1	26.8%	0.49%	0.0699%	27.4%
Suggested_2	3.99%	32.8%	1.22%	38%
Suggested_3	2.44%	2.57%	29.6%	34.6%
Total	33.3%	35.9%	30.9%	100%
Acceptance	26.8%	32.8%	29.6%	89.2%

Group 4 (Random)

	Selected_1	Selected_2	Selected_3	Total
Suggested_1	29%	2.44%	1.62%	33.1%
Suggested_2	8.27%	25.1%	1.62%	35%
Suggested_3	8.65%	5.57%	17.7%	31.9%
Total	46%	33.1%	20.9%	100%
Acceptance	29%	25.1%	17.7%	71.8%

Table 11 All-pairs comparison tests.

			Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
	I	J				Lower Bound	Upper Bound
Tamhane T2	Group 1	Group 2	−1.71	1.52	0.839	−5.74	2.32
		Group 3	0.424	1.51	0.100	−3.58	4.43
		Group 4	14.3*	1.93	1.68E-11	9.13	19.4
	Group 2	Group 1	1.71	1.52	0.839	−2.32	5.74
		Group 3	2.13	1.57	0.688	−2.04	6.30
		Group 4	16.0*	1.98	2.25E-13	10.7	21.2
	Group 3	Group 1	−0.424	1.51	1.00	−4.43	3.58
		Group 2	−2.13	1.57	0.688	−6.30	2.04
		Group 4	13.8*	1.97	1.45E-10	8.60	19.1
	Group 4	Group 1	−14.3*	1.93	1.68E-11	−19.4	−9.13
		Group 2	−16.0*	1.98	2.25E-13	−21.2	−10.7
		Group 3	−13.8*	1.97	1.45E-10	−19.1	−8.60
Dunnett T3	Group 1	Group 2	−1.71	1.52	0.837	−5.74	2.32
		Group 3	0.424	1.51	1.00	−3.58	4.43
		Group 4	14.3*	1.93	1.67E-11	9.13	19.4
	Group 2	Group 1	1.71	1.52	0.837	−2.32	5.74
		Group 3	2.13	1.57	0.685	−2.04	6.30
		Group 4	16.0*	1.98	2.8E-13	10.7	21.2
	Group 3	Group 1	−0.424	1.51	1.00	−4.43	3.58
		Group 2	−2.13	1.57	0.685	−6.30	2.04
		Group 4	13.8*	1.97	1.45E-10	8.60	19.1
	Group 4	Group 1	−14.3*	1.93	1.67E-11	−19.4	−9.13
		Group 2	−16.0*	1.98	2.8E-13	−21.2	−10.7
		Group 3	−13.8*	1.97	1.45E-10	−19.1	−8.60

*: The mean difference is significant at the 0.05 level.

Table 12 Acceptance Rate (Adjusted).

Group	Acceptance Rate
1. Normal	80.1%
2. Privacy-biased	83.4%
3. Open-biased	81.0%
4. Random	63.5%

ify whether or not there are significant differences in the mean numbers of recommendation acceptance among the four groups. Since Levene's test for equality of variance between the groups failed (Levene's Statistic $(3, 548) = 30.682, p < .000$), Welch's t-test applied to the the null hypothesis that there is no significant difference between the mean numbers of recommendation acceptance of the four groups (Welch's Statistic $(3, 298.146) = 24.013, p < .000$). Tamhane's T2 all-pairs comparison test, as well as Dunnett's T3 test, demonstrates that (a) there is no significant difference of the mean numbers of recommendation acceptance between any pair of Groups 1–3, and (b) the mean number of recommendation acceptance of Group 4 is significantly lower than one of any other group (**Table 11**).

There are two likely explanations for the high level of acceptance in Groups 1–3. The first is that people are highly susceptible to suggestions. This interpretation is re-inforced by the results from Ref. [5] in which the unbiased model showed, again, an acceptance rate of around 90% compared with a match of only 60% between the prediction and unprompted selection. The second possibility is that participants are not really paying attention to the suggestions. The acceptance rate of 221 participants was in fact 100%, which on 75 questions shows either very weak preferences/very high susceptibility to suggestion, or not paying very much attention to the study. In Group 4, those presented with random suggestions, 31 participants accepted 100% of the suggestions. This suggests that perhaps one quarter of participants do not pay proper attention and simply accept all answers. Removing those participants who accepted all suggestions, the acceptance rates for the groups are shown in **Table 12**. These remain very high (and very similar) for Groups 1–3 who were shown (sometimes uniformly nudged) suggestions generated from the algorithm, and still much higher than random but noticeably lower for the group presented with random suggestions.

7. Analysis of Survey Results

7.1 Quantitative Analyses

After engaging with the suggestion system experiment, participants had the details of the experiment, including the nudging/randomness of the suggestions they were given. After having read this explanation they gave their impression of the system and experiment via a survey. Participants were requested to indicate their level of agreement with the following statements on a four-point Likert scale (0: not at all; 1: not much; 2: to an extent; 3: very much) (this is an English translation of the survey which was in Japanese).

Q1 This system would be convenient to help control the disclosure and protection of my personal data,

Q2 This system would be useful to help control the disclosure and protection of their personal data,

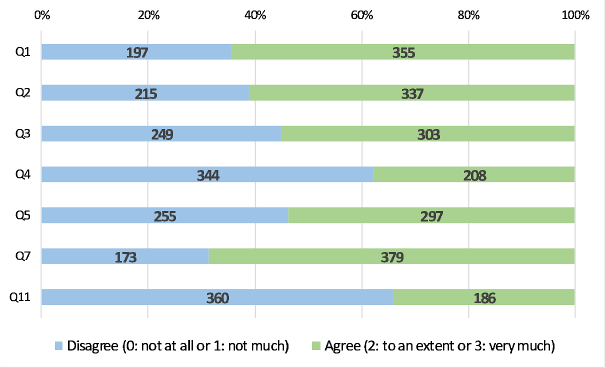


Fig. 8 Disagree/agree ratio in each question.

Q3 I would use this system if it were available.

Q4 I would recommend this system to my friends and acquaintances if it were available.

Q5 It is socially desirable that such a system is broadly available.

Q7 I accepted the system's recommendation for my personal data protection settings.

Q9 The recommendations seemed strange.

Q10 The group I was in (privacy-nudged, openness-nudged, random, non-nudged) matched my feelings about how well or poorly the suggestions matched my desires.

Q11 The fact that the system was presented as using "Artificial Intelligence (Machine Leaning)" to create the recommendations made me more likely to accept the suggestions.

Respondents were asked to indicate the maximum number of predictor questions they would be willing to answer to provide a basis for the suggestions (Q6). They were also asked to give their impression of what percentage of the systems recommendations they accepted (Q8). Finally, there were two open-ended questions with text boxes for answers: (Q14) How do you feel about the system? (Q15) How do you feel about the experiment?

As shown in **Fig. 8**, in total, a majority of respondents gave positive responses for the convenience, usefulness, intention to use, social desirability, and accuracy of our suggestion system (Qs 1, 2, 3, 5, 7). On the other hand, over sixty percent of respondents indicated that they would hesitate to recommend the system to others (Q4). This is consistent with prior studies by the authors in Japan where people are reticent about recommending systems to other people even where they have very positive evaluations themselves. Similarly, over sixty percent claimed that they were not affected in their acceptance of the recommendations by use of the terms "AI" and "machine learning" in the original explanation of the experiment (Q11).

Were these evaluations well-founded? To check this, independent sample t-tests were performed to compare participants' individual acceptance rates (percentage of suggestions left unchanged) and their answers to Qs 1, 2, 3, 4, 5, 7, 11 (taken as binary answers: 0/1 disagree; 2/3 agree). The results of these tests are shown in **Table 13**, with significant differences for Qs 3 (I would use this system), 4 (I would recommend this system), 7 (I accepted the system's recommendations). There were no significant differences for the other questions tested.

The answers to Q8 asking participants to estimate their own

Table 13 Independent samples t-tests.

	diff of means	t	df	Sig	Outcomes
Q1	−1.92	−1.33	370	0.185	n.s.
Q2	−2.00	−1.43	430	0.154	n.s.
Q3	−3.64	−2.67	491	0.0079	Significant at 0.01 level
Q4	−3.99	−2.99	484	0.0029	Significant at 0.01 level
Q5	−1.83	−1.36	550	0.176	n.s.
Q7	−13.4	−8.68	244	0.0000	Significant at 0.01 level
Q11	−2.01	−1.41	544	0.160	n.s.

Table 14 Actual and perceived acceptance rates.

Group		Actual acceptance rate	Q8: Perceived acceptance rate
1	N	154	154
	Mean	88.8	61.0
	Std. Deviation	17.1	26.2
	% of Total Sum	29.3%	28.7%
2	N	123	123
	Mean	91.1	65.5
	Std. Deviation	16.5	25.8
	% of Total Sum	24.0%	24.6%
3	N	140	140
	Mean	88.3	61.4
	Std. Deviation	17.5	26.9
	% of Total Sum	26.5%	26.3%
4	N	135	135
	Mean	69.8	49.7
	Std. Deviation	25.3	26.7
	% of Total Sum	20.2%	20.5%
Total	N	552	552
	Mean	84.5	59.4
	Std. Deviation	21.1	26.9
	% of Total Sum	100.00%	100.00%

Table 15 Correlations between actual and perceived acceptance rates.

Group 1

		Actual acceptance rate	Q8: Perceived acceptancerate
Actual acceptance rate	Pearson Correlation	1	.189*
	Sig. (2-tailed)		0.0187
	N	154	154
Q8: Perceived acceptance rate	Pearson Correlation	.189*	1
	Sig. (2-tailed)	0.0187	
	N	154	154

Group 2

		Actual acceptancer ate	Q8: Perceived acceptancerate
Actual acceptance rate	Pearson Correlation	1	.327**
	Sig. (2-tailed)		0
	N	123	123
Q8: Perceived acceptance rate	Pearson Correlation	.327**	1
	Sig. (2-tailed)	0	
	N	123	123

Group 3

		Actual acceptance rate	Q8: Perceived acceptancerate
Actual acceptance rate	Pearson Correlation	1	.234**
	Sig. (2-tailed)		0.0053
	N	140	140
Q8: Perceived acceptance rate	Pearson Correlation	.234**	1
	Sig. (2-tailed)	0.0054	
	N	140	140

Group 4

		Actual acceptance rate	Q8: Perceived acceptancerate
Actual acceptance rate	Pearson Correlation	1	.494**
	Sig. (2-tailed)		0
	N	135	135
Q8: Perceived acceptance rate	Pearson Correlation	.494**	1
	Sig. (2-tailed)	0	
	N	135	135

Total

		Actual acceptance rate	Q8: Perceived acceptancerate
Actual acceptance rate	Pearson Correlation	1	.371**
	Sig. (2-tailed)		0
	N	552	552
Q8: Perceived acceptance rate	Pearson Correlation	.371**	1
	Sig. (2-tailed)	0	
	N	552	552

**. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).

acceptance rate when compared with their individual actual acceptance rate shows that participants mostly underestimated their acceptance. Participants in Group 2 (Privacy-nudged model), tended to give higher estimates of their acceptance rate than those in Group 1 (Normal model) (**Table 14**). In total and in each group, there are significant positive correlations between actual and perceived acceptance rates at the five percent level ($p \leq 0.05$). Participants in Group 4 (Random model) had a much lower estimate of their acceptance rate (**Table 15**). The apparently closer estimate of their acceptance rate to their actual acceptance rate was not statistically significant (see below).

Table 16 shows the statistics of the differences and distances (the absolute values of the differences) between actual and perceived acceptance rates. The results of Levene's test confirms the homoscedasticity of the differences among the four groups (Levene's Statistic $(3, 548) = 1.353, p = 0.256$), while rejecting one of the distances (Levene's Statistic $(3, 548) = 3.315, p = 0.020 < 0.05$). Analysis of variance of the differences demonstrates that there is no significant difference of the means of the differences among the four groups at five percent level ($F(3, 548) = 2.220, p = 0.085$). For the distances, similar results are obtained through a Welch test (Welch's Statistic $(3, 301.603) = 1.026, p = 0.381$).

Did the different models produce different reported levels of concern regarding the suggestions? **Table 17** shows that generally those who belonged to Groups 1–3 (Normal, Privacy-nudged, Open-nudged) did not report feeling that the suggestions were

Table 16 Statistics of differences and distances between actual and perceived acceptance rates.

Group	N	Difference/distance between actual and perceived	Mean	Std. Dev.	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
						Lower Bound	Upper Bound		
1	154	Difference	27.8	28.4	2.29	23.3	32.4	−40	100
		Distance	29.4	26.8	2.16	25.1	33.6	0	100
2	123	Difference	25.6	25.7	2.32	21	30.2	−28.67	100
		Distance	26.6	24.6	2.22	22.2	31	0	100
3	140	Difference	26.8	28.4	2.4	22.1	31.6	−32.67	100
		Distance	28.5	26.7	2.26	24.1	33	0	100
4	135	Difference	20.1	26.1	2.25	15.7	24.6	−42.67	100
		Distance	24.8	21.7	1.87	21.1	28.5	0	100
Total	552	Difference	25.2	27.4	1.17	22.9	27.5	−42.67	100
		Distance	27.4	25.1	1.07	25.3	29.5	0	100

Table 17 Q9. Feeling of strangeness during the experiment.

	Group	1	2	3	4	Total
No (0/1)	Count	90	81	87	65	323
	Adj. Resid.	−0.053	1.848	0.980	−2.732	
Yes (2/3)	Count	64	42	53	69	228
	Adj. Resid.	0.053	−1.848	−0.980	2.732	
Total	Count	154	123	140	134	551

Table 18 Pearson's chi square test.

Chi-Square Tests

	Value	df	P value (2-sided)
Pearson Chi-Square	9.021[a]	3	0.029
Likelihood Ratio	8.991	3	0.029
Linear-by-Linear Association	3.875	1	0.049
N of Valid Cases	551		

[a]. 0 cells (.0%) have expected count less than 5. The minimum expected count is 50.90.

Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Phi	0.128	0.029
	Cramer's V	0.128	0.029
N of Valid Cases		551	

strange. On the other hand, members of Group 4 (Random) did tended to report a feeling of strangeness relatively more than the other groups. Pearson's chi square test demonstrated that the relative feelings of of strangeness were different among all four groups at five percent significant level (Pearson's Chi-sq (3) = 9.021, p = 0.029 < 0.05). The value of phi coefficient (0.128) suggests a significant relationship between groups and reported level of strangeness at the five percent level (**Table 18**). The adjusted residuals in Table 17 shows that the participants in Group 4 reported a greater feeling of strangeness compared to the other three groups (significant at five percent level).

7.2 Qualitative Analyses: Feelings About our Suggestion System

Unfortunately, only a small number of participants responded to the open-ended questions. Their feelings about our suggestion system are shown in **Fig. 9**. In total, a majority of these respondents expressed positive feelings. Although the headcount is small, respondents who belonged to Group 4 are the most split on their feelings. This is consistent with the previous statistical analysis of Likert scale responses.

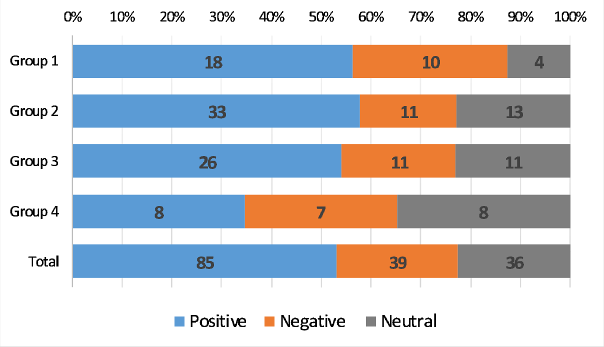


Fig. 9 Feelings about the system.

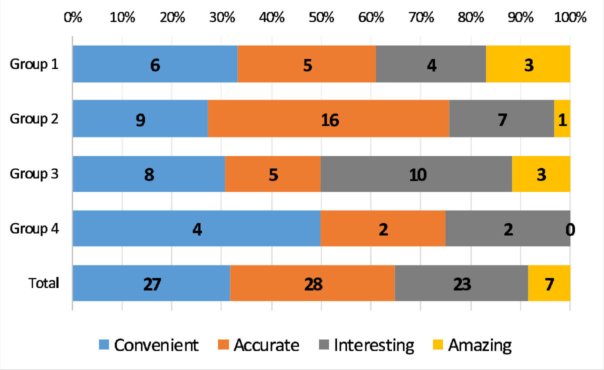


Fig. 10 Positive views on the system.

The system received positive evaluations in terms of its potential convenience and for the accuracy of its recommendations by some (**Fig. 10**). On the other hand, the system also received negative evaluations from others due to the inaccuracy of recommendations (**Fig. 11**). These views, of course, are based more on their perception of their acceptance rate than their actual acceptance rate (Table 14, **Table 19**).

There are not many differences in answers to questions (Qs 1, 2, 3, 4, 5, 7, 9) between those respondents who left all suggestions unchanged (All-accepted group (n = 221)) and those who did not (Not-all-accepted group (n = 331)). Only the accuracy of the system's suggestions (answers to Q7) was more highly evaluated by All-accepted group than Not-all-accepted group at 0.1% significant level (Welch's t(471.510) = 7.016, p < 0.000). Any statistically significant differences in the means of answers to other questions between the two groups were not found (**Table 20**).

When comparing answers to the questions responded by All-accepted group (n = 221) with those by Bottom-tertile group in terms of actual acceptance rate (n = 185), significant differ-

Table 19 Actual and perceived accept rates of those who answered to Q14 that the recommendations were accurate/inaccurate.

"The recommendations are accurate"

	Group 1 (N=5)		Group 2 (N=16)		Group 3 (N=5)		Group 4 (N=2)		Total (N=28)	
	Actual accept rate	Q8: Perceived accept rate	Actual accept rate	Q8: Perceived accept rate	Actual accept rate	Q8: Perceived accept rate	Actual accept rate	Q8: Perceived accept rate	Actual accept rate	Q8: Perceived accept rate
Max	1	1	1	1	0.975	0.95	1	1	1	1
Min	0.2875	0.2	0.3125	0.5	0.875	0.4	0.7375	0.5	0.2875	0.2
Mean	0.8525	0.72	0.8875	0.76125	0.915	0.78	0.86875	0.75	0.884821	0.756429
Median	1	0.9	0.96875	0.775	0.925	0.85	0.86875	0.75	0.975	0.8

"The recommendations are inaccurate"

	Group 1 (N=3)		Group 2 (N=5)		Group 3 (N=7)		Group 4 (N=4)		Total (N=19)	
	Actual accept rate	Q8: Perceived accept rate	Actual accept rate	Q8: Perceived accept rate	Actual accept rate	Q8: Perceived accept rate	Actual accept rate	Q8: Perceived accept rate	Actual accept rate	Q8: Perceived accept rate
Max	0.6125	0.65	1	0.98	1	0.9	0.7875	0.7	1	0.98
Min	0.45	0.3	0.8375	0.3	0.4625	0.2	0.4	0.1	0.4	0.1
Mean	0.533	0.483	0.938	0.566	0.836	0.514	0.559	0.45	0.757	0.509
Median	0.5375	0.5	0.95	0.5	0.9375	0.5	0.525	0.5	0.7875	0.5

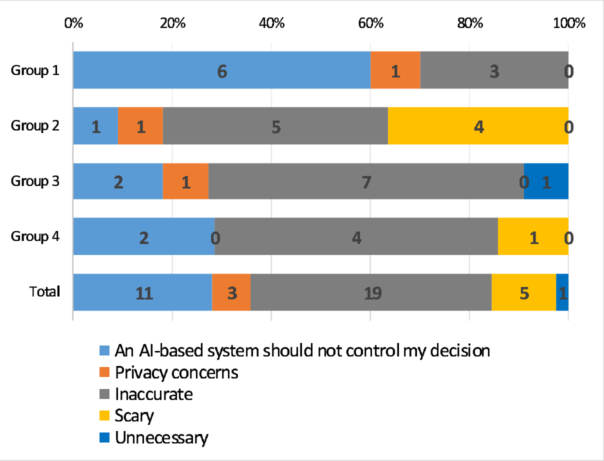


Fig. 11 Negative views on the system.

Table 20 Independent samples *t*-test.

	Mean difference	t	df	Sig	Outcomes
Q1	−0.037	−0.612	550	0.541	n.s.
Q2	−0.031	−0.496	550	0.62	n.s.
Q3	0.013	0.18	550	0.857	n.s.
Q4	0.113	1.593	445.539	0.112	n.s.
Q5	−0.022	−0.325	550	0.745	n.s.
Q7	0.4	7.016	471.51	0	Significant at 0.001 level
Q9	−0.104	−1.639	549	0.102	n.s.

Table 21 Independent samples *t*-test.

	Mean difference	t	df	Sig	Outcomes
Q1	0.017	0.24	404	0.81	n.s.
Q2	0.017	0.228	404	0.82	n.s.
Q3	0.086	1.022	404	0.307	n.s.
Q4	0.161	1.935	404	0.054	n.s.
Q5	−0.003	−0.041	404	0.967	n.s.
Q7	0.576	8.578	384.957	0	Significant at 0.001 level
Q9	−0.23	−3.077	403	0.002	Significant at 0.05 level

ences in the means of answers to the questions concerning the accuracy of the system's suggestions (Q7: Welch's $t(384.957) = 8.578, p < 0.000$) and the feeling of strangeness provided by the suggestions (Q9: $t(403) = -3.077, p = 0.0022 < 0.0071$ (Bonferroni corrected 5% significance level)) were discovered at 0.1% significant level and 5% significant level given the multiplicity of testing, respectively (**Table 21**).

8. Conclusion

8.1 Summary

This paper reported on an investigation into the impact on acceptability of recommendations in a privacy settings context of a uniform nudge towards privacy or openness. The experiment showed no change in acceptability of moderately strong nudges. This result was underlain by a parallel random suggestion model which demonstrated a susceptibility to suggestion amongst participants but nevertheless a significantly lower acceptance rate for the random suggestions than the model with or without nudges. The results of this suggest that malicious operators of suggestion systems can potentially influence users in their privacy choices without the users awareness of being manipulated.

8.2 Discussion

From the results above, we can see that many, though not all, users are highly susceptible to suggestion when give shortcuts to process time-consuming tasks. This experiment is a low risk scenario for users to accept the results, given that it is presented as an abstract task, not a real system which would have actual consequences for their privacy. However, in the survey questions, many participants expressed interest in this kind of suggestion system for real world applications. This is not surprising given the existing work showing the real world burden on users trying to manage their own digital footprints and protect their privacy [8]. The lack of difference between the average acceptance rates of the nudged towards sharing, neutral and nudged towards privacy groups suggests also that people's attitudes are not rigid, but depend on current mood, attention, recent stimuli and how questions are presented. The significant difference between the acceptance rate of the machine-learning based suggestions and the random suggestions indicates that modestly sophisticated pressures are more successful at influencing users than generic ones.

8.3 Recommendations

The results of this experiment lead to some suggestions for various groups.

Users should be aware that they are susceptible to manipulation of their privacy attitudes and should be wary of accepting default settings, or of accepting suggestions, unless they trust the source of those suggestions.

Providers of services who wish to help their users' protect their privacy while exploiting the benefits of sharing should push their users more towards privacy, which is likely to be acceptable to those users. If, over time, those users then loosen privacy controls in certain circumstances, then they are more likely to be doing so for good specific reasons, than because of limited time or understanding.

Regulators should require at a minimum transparency from providers who make setting suggestions, but could also consider requiring providers to institute privacy-nudged suggestions, to counteract the tendency of platforms (from a position of power) to push users in the direction of over-sharing.

Digital Rights Activists should be aware that users are susceptible to suggestion in this area and where possible should be providing users with both countervailing rhetoric against over-sharing, but also providing tools to help users easily limit sharing. Such tools are likely to be well-received even where explicitly privacy-nudged, provided that they are based on decent individualised suggestions.

Researchers should be aware of these issues when conducting empirical research on users' expressed privacy preferences, particularly when discussing the aparent "privacy paradox" wherein users express a desire for privacy but then accept or embrace broad sharing of their personal data. Future work on tools to help users exert control over their personal data should take the potential misuse by self-interested parties into account when making such tools available.

References

[1] Liu, Y., Gummadi, K.P., Krishnamurthy, B. and Mislove, A.: Analyzing facebook privacy settings: User expectations vs. reality, *Proc. 2011 ACM SIGCOMM Conference on Internet Measurement*, pp.61–70 (2011), available from ⟨https://doi.org/10.1145/2068816.2068823⟩.

[2] Bell, G.: A personal digital store, *Comm. ACM*, Vol.44, No.1, pp.86–91 (2001).

[3] Kiyomoto, S., Nakamura, T., Takasaki, H., Watanabe, R. and Miyake, Y.: PPM: Privacy Policy Manager for Personalized Services, *Proc. Security Engineering and Intelligence Informatics*, pp.377–392 (2013), available from ⟨https://link.springer.com/chapter/10.1007/978-3-642-40588-4_26⟩.

[4] Nakamura, T., Kiyomoto, S., Tesfay, W.B. and Serna, J.: Personalised Privacy by Default Preferences - Experiment and Analysis, *the 2nd International Conference on Information Systems Security and Privacy* (*ICISSP2016*), pp.53–62 (2016).

[5] Nakamura, T., Adams, A.A., Murata, K. and Kiyomoto, S.: Effectiveness and Acceptability Evaluation for a Machine Learning Based Privacy Setting Prediction Scheme, *Proc. 2017 Symposium on Cryptography and Information Security* (*SCIS2017*), in Japanese, 1C1–2 (2017).

[6] Adams, A.: Facebook code: Social network sites platform affordances and privacy, *Journal of Law, Information and Science*, Vol.23, No.1, pp.158–168 (2014), available from ⟨http://www.jlisjournal.org/abstracts/adams.23.1.html⟩.

[7] Madejski, M., Johnson, M. and Bellovin, S.M.: A study of privacy settings errors in an online social network, *2012 IEEE International Conference on Pervasive Computing and Communications Workshops* (*PERCOM Workshops*), pp.340–345, IEEE (2012).

[8] Solove, D.J.: Privacy self-management and the consent paradox, *Harvard Law Review*, Vol.126, pp.1880–1903 (2013).

[9] Acquisti, A. and Grossklags, J.: Privacy and rationality in individual decision making, *IEEE, Security & Privacy*, Vol.3, No.1, pp.26 –33 (2005).

[10] Pollach, I.: What's wrong with online privacy policies?, *Comm. ACM*, Vol.50, No.9, pp.103–108 (2007).

[11] Jensen, C., Potts, C. and Jensen, C.: Privacy practices of internet users: Self-reports versus observed behavior, *Int. J. Hum.-Comput. Stud.*, Vol.63, No.1-2, pp.203–227 (2005).

[12] Guo, S. and Chen, K.: Mining privacy settings to find optimal privacy-utility tradeoffs for social network services, *Privacy, Security, Risk and Trust* (*PASSAT*), *2012 International Conference on and 2012 International Confernece on Social Computing* (*SocialCom*), pp.656–665, IEEE (2012).

[13] Cranor, L.F.: P3P: Making privacy policies more useful, *IEEE Security & Privacy*, Vol.1, No.6, pp.50–55 (2003).

[14] Dehghantanha, A., Udzir, N.I. and Mahmod, R.: Towards a pervasive formal privacy language, *2010 IEEE 24th International Conference on Advanced Information Networking and Applications Workshops* (*WAINA*), pp.1085–1091, IEEE (2010).

[15] Bekara, K., Mustapha, Y.B. and Laurent, M.: Xpacml extensible privacy access control markup language, *2010 2nd International Conference on Communications and Networking* (*ComNet*), pp.1–5, IEEE (2010).

[16] Backes, M., Karjoth, G., Bagga, W. and Schunter, M.: Efficient comparison of enterprise privacy policies, *Proc. 2004 ACM Symposium on Applied Computing*, pp.375–382, ACM (2004).

[17] Tondel, I.A. and Nyre, A.A.: Towards a similarity metric for comparing machine-readable privacy policies, *Open Problems in Network Security*, Vol.7039 *Lecture Notes in Computer Science*, pp.89–103, Springer (2012).

[18] Cranor, L. et al.: The platform for privacy preferences 1.1 (P3P1.1) specification (2002).

[19] Cranor, L.F., Arjula, M. and Guduru, P.: Use of a p3p user agent by early adopters, *Proc. 2002 ACM Workshop on Privacy in the Electronic Society*, *WPES '02*, pp.1–10, ACM (2002).

[20] Cranor, L.F., Guduru, P. and Arjula, M.: User interfaces for privacy agents, *ACM Trans. Comput.-Hum. Interact.*, Vol.13, No.2, pp.135–178 (2006).

[21] Kolter, J. and Pernul, G.: Generating user-understandable privacy preferences, *ARES '09, International Conference on Availability, Reliability and Security*, pp.299–306, IEEE (2009).

[22] Pedersen, A.: P3 - problems, progress, potential, *Privacy Laws & Business International Newsletter*, Vol.2, pp.20–21 (2003).

[23] Cranor, L.F.: Necessary but not sufficient: Standardized mechanisms for privacy notice and choice, *J. Telecomm. & High Tech. L.*, Vol.10, No.2, pp.273–308 (2012).

[24] Yee, G.O.M.: An automatic privacy policy agreement checker for e-services, *ARES '09, International Conference on Availability, Reliability and Security*, pp.307–315, IEEE (2009).

[25] Biswas, D.: Privacy policies change management for smartphones, *2012 IEEE International Conference on Pervasive Computing and Communications Workshops* (*PERCOM Workshops*), pp.70–75, IEEE (2012).

[26] Wishart, R., Corapi, D., Madhavapeddy, A. and Sloman, M.: Privacy butler: A personal privacy rights manager for online presence, *2010 8th IEEE International Conference on Pervasive Computing and Communications Workshops* (*PERCOM Workshops*), pp.672–677, IEEE (2010).

[27] Bylund, M., Karlgren, J., Olsson, F., Sanches, P. and Arvidsson, C.-H.: Mirroring your web presence, *Proc. 2008 ACM Workshop on Search in Social Media, SSM '08*, pp.87–90, ACM (2008).

[28] Srivastava, A. and Geethakumari, G.: A framework to customize privacy settings of online social network users, *2013 IEEE Recent Advances in Intelligent Computational Systems* (*RAICS*), pp.187–192, IEEE (2013).

[29] Srivastava, A. and Geethakumari, G.: A privacy settings recommender system for online social networks, *Recent Advances and Innovations in Engineering* (*ICRAIE 2014*), pp.1–6, IEEE (2014).

[30] Fang, L. and LeFevre, K.: Privacy wizards for social networking sites, *Proc. 19th International Conference on World Wide Web*, pp.351–360, ACM (2010).

[31] Fang, L., Kim, H., LeFevre, K. and Tami, A.: A privacy recommendation wizard for users of social networking sites, *Proc. 17th ACM Conference on Computer and Communications Security*, pp.630–632, ACM (2010).

[32] Berendt, B., Günther, O. and Spiekermann, S.: Privacy in e-commerce: Stated preferences vs. actual behavior, *Comm. ACM*, Vol.48, No.4, pp.101–106 (2005).

[33] Sadeh, N., Hong, J., Cranor, L., Fette, I., Kelley, P., Prabaker, M. and

Rao, J.: Understanding and capturing people's privacy policies in a mobile social networking application, *Personal Ubiquitous Comput.*, Vol.13, No.6, pp.401–412 (2009).

[34] Tondel, I.A., Nyre, A.A. and Bernsmed, K.: Learning privacy preferences, *2011 6th International Conference on Availability, Reliability and Security* (*ARES*), pp.621–626, IEEE (2011).

[35] Kelley, P.G., Drielsma, P.H., Sadeh, N. and Cranor, L.F.: User-controllable learning of security and privacy policies, *Proc. 1st ACM Workshop on Workshop on AISec, AISec '08*, pp.11–18, ACM (2008).

[36] Buffett, S. and Fleming, M.W.: Applying a preference modeling structure to user privacy (2007), available from ⟨http://secml.otago.ac.nz/privacy2007/Main/SPACE07.pdf⟩.

[37] Mugan, J., Sharma, T. and Sadeh, N.: Understandable learning of privacy preferences through default personas and suggestions (2011).

[38] Johnson, E.J., Bellman, S. and Lohse, G.L.: Defaults, framing and privacy: Why opting in-opting out, *Marketing Letters*, Vol.13, No.1, pp.5–15 (2002).

[39] Wang, N., Wisniewski, P., Xu, H. and Grossklags, J.: Designing the default privacy settings for facebook applications, *Proc. 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*, pp.249–252 (2014), available from ⟨https://doi.org/10.1145/2556420.2556495⟩.

[40] Nissenbaum, H.: Privacy in Context: Technology, Policy, and the Integrity of Social Life, Stanford University Press (2010).

[41] Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., Leisch, F., Chang, C.-C. and Lin, C.-C.: Package 'e1071' (2015), available from ⟨https://cran.r-project.org/web/packages/e1071/e1071.pdf⟩.

Appendix

A.1 Suggestion Generation Algorithm

This section introduces the detail of suggestion algorithm in Ref. [4]. The algorithm was based on SVM, which is considered a powerful learning system. Although SVM is mainly for binary-class problems, it can be extended for high-dimensional feature spaces through a nonlinear mapping chosen a priori. Therefore, for the purpose of these experiments, a multilabel and multiclass SVM approach were used.

The algorithm consisted of two phases; the *learning phase* and *guessing phase*.

[Learning Phase]

- n questions are selected where $1 \leq n \leq Max$. Max equals the total number of questions and n equals the number of selected questions used for training the corresponding answers.
- Using the selected n questions, the SVM privacy preference model is generated. In this model, the class labels represent the acceptance level for each of the unselected $Max-n$ questions using a combination of answers for n as sample points in the training data.

[Guessing Phase]

- For each unknown point, i.e., a combination of answers to selected n questions, the SVM models generated in the learning phase is used for each unselected question and calculate the guessed values of the answers to those $Max - n$ unselected questions.

This approach was implemented with R. The trials of evaluation were repeated for 10 times while the data samples were randomly chosen, and were randomly split into training data and testing data. **Table A·1** shows the summary of parameters used in our experimental setup.

There were two different experiments. The first experiment was for selecting top combinations, $TC = 15$ of n questions that achieved the highest accuracy considering 150 entries randomly selected; i.e., 100 entries for the training data, 50 entries for the testing data for decreasing the running time when evaluating all possible combinations. Each parameter of the SVM model was optimised by a grid search on the parameters C and γ. In the second experiment, the same top combinations, $TC = 15$ of n questions were used for evaluating the algorithm using 10,000 entries (i.e., 9,000 for training data, and 1,000 for testing data).

Table A·1 Experimental settings.

Parameter	Value
Max	80
n	5
Top Combinations (TC)	$TC = 15$
Training Data (TRD)	$TRD = 100$, $TRD = 9{,}000$
Test Data (TED)	$TED = 50$, $TED = 1{,}000$

The results of a guessing accuracy are more than 83% for all top 15 combinations, and 85% for 9 of the 15 top combinations.

A.2 Explanation of Experiment

The following is the original text of the explanation of our experiment in Japanese.

本日は実験にご参加いただき，誠にありがとうございます．この実験は，株式会社KDDI総合研究所が，明治大学ビジネス情報倫理研究所の協力を得て開発を進めている，個人情報開示設定支援システムの有効性を評価するために行われるものです．

本システムでは，インターネット上のサービス（たとえばソーシャルメディアやオンラインショッピング）を利用する際の細かな個人情報開示設定（どの程度自分の個人情報を企業に提供してよいか）を，少数（現在のシステムでは5つ）の質問に回答することによって自動的に行うことができます．これにより簡単に，しかも短い時間で，自分の好みに合った個人情報の開示と保護の設定をすることができます．

本システムの開発にあたっては個人情報の開示と保護の設定に関する約1万人のデータを，人工知能技術を使って機械学習させ，個人情報の開示と保護の好みに関するパターンを抽出しています．そしてこのシステムは，その抽出されたパターンを使って，あなたにおすすめの個人情報の開示と保護の設定を表示します．

本日の実験では，個人情報の開示と保護についての，全部で80項目の質問を用意しています．質問はすべて，特定の個人情報（たとえば住所）をインターネット上のさまざまなサービスに提供してもよいと考えるかどうかに関するものです．回答の仕方は三者択一で，

- ○：この個人情報を提供してもよい
- △：この個人情報の提供に関しては，サービスの内容によって個別に判断する
- ×：この個人情報を提供したくない

のいずれかを選んで（クリックして）ください．

最初に5つの質問に答えてもらいます．回答が終わって次の画面に進むと，この5つの質問への回答結果に基づいて本システムが推定した，残り75項目の質問に対するあなたの好みの回答が表示されます．これを見て，自分の考えと同じである場合はそのままにし，自分の考えと異なる場合には自分の考えに近いものに修正してください．75項目の質問に対する修正作業が終了した後で，アンケート調査の画面に進みますので，必ずそちらにも答えるようにしてください．

ご協力の程，どうぞよろしくお願いいたします．

The following is the explanation translated into English.

Thank you for your participation in today's experiment. This experiment is conducted for evaluating the effectiveness of a privacy protection setting support system, which has been developed by KDDI Research, Inc. in collaboration with the Centre for Business Information Ethics at Meiji University.

Using this system, individual users can automatically make detailed settings for controlling the revelation of personal information when using online services such as social network services and online shopping services through responding to a small number of (five, this time) questions. The system allows individual users to easily set up their preferences for revealing or withholding their personal information in a short time.

Machine learning, an artificial intelligence technology, was applied to the data of preferences for revealing or withholding their personal information of about ten thousand people to extract the patterns of the settings, which are then used to produce suggestions.

Eighty questions in total are posed in today's experiment. Each question is related to whether you would provide a certain kind of personal information such as your home address to online service providers. When answering a question, you are required to choose one from the following three options:

- ○: I will provide this kind of personal information to any online service provider;
- △: I may share this kind of personal information with an online service provider when requested;
- ×: I won't provide this kind of personal information to any online service provider.

First, you are required to respond to five questions. After that, suggested responses to the remaining seventy-five questions generated by the system, based on your responses to the first five questions, will be shown to you. If the suggested responses are the same as your preferences, please leave them unchanged. If not, please change them. After the experiment, you are required to answer a questionnaire.

Thanks for your cooperation.

Toru Nakamura was born in 1983. He received the B.E., M.E., and Ph.D. degrees from Kyushu University, in 2006, 2008, and 2011, respectively. In 2011, he joined KDDI and in the same year he moved KDDI R & D Laboratories, Inc. (currently renamed KDDI Research, Inc.). Since 2018, he is a researcher in Advanced Telecommunications Research Institute International (ATR). He received CSS2016SPT Best Paper Award. His current research interests include security and privacy, especially privacy enhanced technology and analysis of privacy attitudes. He is a member of IEICE and IPSJ.

Andrew A. Adams was born in 1969. He has a Ph.D. in Computer Science from the University of St Andrews (1997) and an LLM (Masters in Law) from the University of Reading (2005). He is Deputy Director of the Centre for Business information Ethics at Meiji University. He is a member of the BCS, IEEE and ACM.

Kiyoshi Murata was born in 1957. He is a director of the Centre for Business Information Ethics and a professor of MIS at the School of Commerce, Meiji University. He has studied information ethics including privacy, surveillance, ICT professionalism and gender and computing issues since 1997 following his career of research on economics, operational research, business administration and management information systems. He is a member of the IEEE Computer Society, ACM and International Society for Ethics and Information Technology.

Shinsaku Kiyomoto received his B.E. in engineering sciences and his M.E. in Material Science from Tsukuba University, Japan, in 1998 and 2000, respectively. He joined KDD (now KDDI) and has been engaged in research on stream ciphers, cryptographic protocols, and mobile security. He is currently a senior manager at the Information Security Laboratory of KDDI Research, Inc. He was a visiting researcher of the Informatiopn Security Group, Royal Holloway University of London from 2008 to 2009. He received his doctorate in engineering from Kyushu University in 2006. He received the IEICE Young Engineer Award in 2004, Distinguished Contributions Awards in 2011, and Achievement Award in 2016. He is a member of IEICE and JPS.

Nobuo Suzuki was born in 1963. He received his Ph.D. from Tsukuba University in 2007. He became Senior Manager of Smart Wireless Technology Group at KDDI R&D Laboatories Inc. in 2011, Head of Department of Broadband Wireless Communications at Advanced Telecommunications Research Institute International in 2016, and a professor at Kindai University in 2019. His current research interests are IoT security and Cognitive wireless system. He is a Senior member of the IPSJ, and a member of the IEICE and IEEJ.