Yao Li*, Alfred Kobsa, Bart P. Knijnenburg, and M-H. Carolyn Nguyen

Cross-Cultural Privacy Prediction

Abstract: The influence of cultural background on people's privacy decisions is widely recognized. However, a cross-cultural approach to predicting privacy decisions is still lacking. Our paper presents a first integrated cross-cultural privacy prediction model that merges cultural, demographic, attitudinal and contextual prediction. The model applies supervised machine learning to users' decisions on the collection of their personal data, collected from a large-scale quantitative study in eight different countries. We find that adding culture-related predictors (i.e. country of residence, language, Hofstede's cultural dimensions) to demographic, attitudinal and contextual predictors in the model can improve the prediction accuracy. Hofstede's variables - particularly individualism and indulgence - outperform country and language. We further apply generalized linear mixed-effect regression to explore possible interactions between culture and other predictors. We find indeed that the impact of contextual and attitudinal predictors varies between different cultures. The implications of such models in developing privacy-enabling technologies are discussed.

Keywords: privacy decision, culture, individualism, context

DOI Editor to enter DOI
Received ..; revised ..; accepted ...

1 Introduction

Research has shown that information privacy concerns are not uniform, but are influenced by many determinants, such as demographic differences [5, 40, 49], privacy attitudes [1], cultural dimensions [7, 11, 31], and contextual/situational factors [25, 34]. Since privacy concerns vary extensively across these factors, any social, technical and regulatory approaches to support users in exercising control over their privacy should take these factors into account. To this end, some privacy researchers have proposed to segment users based on privacy-related factors [26], such as demographics [11] or culture [13, 31]. In terms of culture, prior research has considered the country of residence [4], the native language [45], and cultural characteristics such as Hofstede's "cultural dimensions" [7, 11, 31].

Other researchers instead endeavored to predict individual users' privacy decisions based on contextual and situational factors [8, 15, 46, 50]. Preliminary efforts exist to use these predictions to provide adaptive privacy decision support [16, 27]. Privacy prediction models are usually exclusively based on variables related to the immediate decision context.

While both streams of research have had considerable success, to our best knowledge no work has tried to combine the two into an integrated, cross-cultural prediction model. Previous studies that examined how demographic differences, privacy attitudes and contextual factors individually impact user privacy behaviors have not properly attended to their possible interaction with cultural background. Such a cross-cultural model would be of considerable interest though. As more organizations run their business globally, they need to deal with information privacy issues that extend beyond cultural boundaries.

In this paper, we present a cross-cultural privacy prediction model based on an online survey with 9625 participants from 8 different countries on four continents. Our main goal is to demonstrate the importance of incorporating cultural factors in privacy prediction, e.g. in websites that cater to audiences from different countries. We apply both machine learning and regression methods to predict users' agreement when organizations request their personal data. We are thereby specifically interested in the following research questions:

RQ1: If demographic, attitudinal and contextual factors are used to predict users' acceptance of personal data collection, will the addition of culture-related features improve the predictions?

RQ2: Which measurement of culture yields better prediction accuracy: country of residence, native language, or Hofstede's cultural dimensions?

***Corresponding Author: Yao Li:** University of California, Irvine, E-mail: yao.li@uci.edu
Alfred Kobsa: University of California, Irvine, E-mail: kobsa@uci.edu
Bart P. Knijnenburg: Clemson University, E-mail: bartk@clemson.edu
M-H. Carolyn Nguyen: Microsoft Corporation, E-mail: cnguyen@microsoft.com

RQ3: Do we need to consider the effects of contextual and attitudinal factors differently in different cultures?

To answer these questions, we test the improvement in the prediction accuracy of a privacy prediction model using different cultural predictors, in addition to demographic, attitudinal and contextual factors. We also examine the statistical significance of interactions between demographic, attitudinal, and contextual factors and culture in predicting disclosure tendencies. Below, we review the existing literature on determinants of users' privacy decisions. In Section 3, we first describe a supervised machine learning approach to establish our privacy prediction model using various combinations of determinants as predictors. Thereafter we present a generalized linear mixed-effect regression model to investigate interaction effects. Section 4 discusses our results and explains the impact of each determinant. In Section 5, we discuss the implications of our findings for future privacy research and the development of cross-cultural, context-aware privacy support systems.

2 Literature Review

In this section, we review the related literature on the influence of cultural factors on privacy concerns. A number of antecedents have been shown to be associated with privacy concerns, such as demographic differences [5, 40, 49], contextual factors [25, 34], and cultural climate [7, 11, 31]. With regard to demographics, it has been shown that women, particularly older women, generally had more concerns about their personal privacy than men [5, 11], and people with higher education were more sensitive and aware of potential privacy issues [49]. In terms of contextual factors, the identity of the information inquirer, the type of the requested information [9, 28], the social ties of the requester [9], and the timing of the request have all been shown to strongly determine users' privacy preferences [8, 29]. In terms of culture, researchers have investigated the difference in personal information privacy concerns across nationalities [14], and made suggestions how cultural differences should be taken into account in regulatory approaches [31, 32].

Whereas most existing cross-cultural privacy research focuses on attitudes, our work focuses on cultural predictors of individuals' behavioral intentions regarding privacy. Specifically, we consider the influence of country of residence, native language and Hofstede's cultural dimensions [20] as possible cultural antecedents of privacy decisions. Hofstede's dimensions are the result of a factor analysis at the level of country means of a comprehensive survey instrument, aimed at identifying systematic differences in national cultures. Their purpose is to measure culture in countries, societies, subgroups, and organizations; they are not meant to be regarded as psychological traits [19].

Hofstede's dimensions were first developed in the 1960s and 1970s at IBM, and later enhanced by two new dimensions. Various studies validated the model by including other respondent groups such as students, managers, and pilots [17, 23]. Hofstede's most recent model identifies the following six cultural dimensions [21]:

- Power distance (PDI) is the degree to which the less powerful members of a society accept and expect that power is distributed unequally. A high score of PDI indicates that people accept a hierarchical order;
- Individualism (IND) is defined as a preference for a loosely-knit social framework in which individuals are expected to take care of only themselves and their immediate families. Low individualism is collectivism;
- Masculinity (MAS) represents a preference in society for achievement, heroism, assertiveness and material rewards for success. Such society is more competitive. Its opposite, femininity, stands for a preference for cooperation, modesty, caring for the weak and quality of life;
- Pragmatism (PRA) describes how a society has to maintain some links with its own past while dealing with the challenges of the present and future;
- Uncertainty Avoidance (UAI) is the degree to which the members of a society feel uncomfortable with uncertainty and ambiguity;
- Indulgence (IDL) stands for a society that allows relatively free gratification of basic and natural human drives related to enjoying life and having fun. Its opposite is restraint.

93 countries have meanwhile been scored along each of these dimensions, on a scale of 1 to 120. Hofstede's model has been successfully applied to studies about the influence of culture on privacy concerns. For example, Milberg et al. [32] found that power distance, individualism and masculinity had a positive effect on overall information privacy concerns, whereas uncertainty avoidance had a negative effect. Bellman et al. [7] found opposite results, namely that power distance, individualism

and masculinity are negatively associated with privacy concerns, and that uncertainty has no significant effect. Posey et al. [38] and Miltgen and Peyrat-Guillard [33] found that focus groups in individualistic societies were more hesitant to disclose information than those in collectivistic societies. Similarly, Cho et al. [11] found that Internet users from highly individualistic cultures exhibited greater concerns about online privacy. Steenkamp and Geyskens [45] found that individualistic countries give more weight to privacy protection and customization than collectivistic countries.

While research on the effect of culture on privacy concerns is plentiful, we note that only very few papers have tested the effects of country, language, and/or cultural dimensions on privacy decision prediction in concert with other non-cultural predictors. No study has yet discussed what culture measurement is more appropriate and why, nor considered whether the impact of non-cultural predictors on privacy decisions varies in different cultures. Our paper will fill these gaps.

3 Method

3.1 Data collection

Our data was collected through online surveys simultaneously distributed in Canada, China, Germany, the United States, the United Kingdom, Sweden, Australia and India. These eight countries were chosen to cover a wide range of the existing privacy/data protection landscape and social/cultural attitudes. Participants were recruited from an existing online panel maintained by Ipsos, which comprises over 4.5 million people in 46 countries and complies with ISO 20252.

The survey had been developed iteratively in earlier interviews and focus groups with 76 participants in four different countries [30]. It was conducted in the language of the participants. The translations were done by professional translators who were also native speakers. Semantic consistency between translations was ensured by qualitative pilot-testing in each country. Technical terms in the online survey were explained through mouse-over text.

The survey took between 20-25 minutes to complete. It first collected data about participants' demographics, and their attitudes on a number of topics in information privacy. Each attitudinal item was measured by multiple questions on a 7 or 5 point scale. These questions were subjected to a confirmatory factor analysis, and the factor scores were saved as input for the prediction models. The first two sections of Table A1 in the Appendix summarize the demographic and attitudinal items of the survey as well as the extracted attitudinal factors.

The next part of the survey presented descriptions of scenarios to participants in which their personal information was collected. Each scenario contained a combination of contextual factors: the type of personal data being collected, the collection method, and the device from which data is collected (these three factors are interrelated, and are therefore captured in a single contextual factor); the entity that is collecting the information; the usage purpose of the data; and value exchange from data collection. The third section of Table A1 in the Appendix summarizes the contextual items used in the survey.

192 scenarios were defined after verifying that their combination of contextual factors made sense, and a randomized subset of 6 scenarios was shown to each respondent. Example scenarios are:

- "From my mobile device, I will provide my current location and contact list [collection method] to a service provider. The service provider is a company with no locations in my country [entity]. The service provider will use the information as I agreed [purpose]. It will also save me time or money [value]."
- "A service provider can use my purchase history to generate other information about me such as buying behavior [collection method]. The service provider is a company that provides free services [entity]. The service provider will use the information to customize the choices offered to me [purpose]. It will also provide me something of unique or compelling value [value]."

For each shown scenario, participants were asked to indicate "whether the scenario as described would be acceptable or not acceptable to you" ("outcome variable" at the bottom of Table A1).

Regarding cultural factors, we consider country of residence and language (both measured as part of the demographics), as well as Hofstede's cultural dimensions (we used the most recent values for participants' country of residence from Hofstede's website). The forth section of Table A1 summarizes the cultural factors.

9,625 responses were received when the survey was administered to the multinational panel (about 1200 per country). The respondents ranged from 18 to 65 years old. Age and gender quotas were set within each country to ensure that the sample collected would be representative of each country's online adult population. Table A2 in the Appendix shows some demographics by country, and Hofstede's cultural dimension values for these countries.

3.2 Supervised learning methods

We applied supervised machine learning to examine whether adding cultural predictors will improve the prediction performance and how (RQ1 and RQ2). We analyzed the data with Decision Trees [39], a supervised learning technique which recursively separates instances in branches to form a tree structure. The construction of a decision tree is based on Information Gain (IG). IG calculates how useful a given feature is to distinguish the different categories of the outcome. Features with high IG will be the top nodes of the tree as they instantly separate the whole dataset into subsets.

Decision Trees can directly predict binary outcomes and indicate the accuracy of its predictions. It can produce a model with interpretable rules, and identify the features that most differentiate the outcomes. We used the software package WEKA [22], which implements the J48 algorithm for Decision Trees. To avoid overfitting, i.e. the problem that our model might predict less well on a new dataset than our current dataset, we chose to perform tree pruning and a 10-fold cross-validation, where we randomly split our dataset into 10 smaller sets, trained a model using 9 of them, and validated the model on the remaining set. We ran 10 rounds of validation in total, and computed the performance measures of the classifier by averaging the accuracy and F-measure over the 10 rounds. Accuracy measures the overall percentage of correctly classified instances, while the F-measure is the average of precision and recall. As Figure 1 and Equations 1-4 show, recall measures the proportion of positives that are correctly predicted as such, and precision measures the proportion of correct predictions among the positive predictions. The F-measure is then the harmonic mean of these two metrics, ranging from 0 (worst) to 1 (best).

$$Accuracy = \frac{tp + tn}{tp + fp + tn + fn} \tag{1}$$

$$Precision = \frac{tp}{tp + fp} \tag{2}$$

$$Recall = \frac{tp}{tp + fn} \tag{3}$$

$$F - measure = 2 \times \frac{precision \times recall}{precision + recall} \tag{4}$$

After classification, we also performed feature selection in WEKA based on the IG of each feature, yielding a ranked list of the most relevant predictors. The ranked list informed us of the importance of each predictor, especially the cultural predictors.

The input predictors included all 8 demographic, 7 attitudinal and 4 contextual features, and also country, language and Hofstede's 6 cultural dimensions. The outcome of our classifier was binary, indicating whether or not participants accept the collection of their personal data in a given scenario. Some of the cultural predictors were highly correlated, such as power distance and individualism. Multicollinearity was however not reduced in our analysis. Our main goal was to examine the prediction improvement using different sets of predictors, and reduction of multicollinearity would not improve the prediction performance [37].

3.3 Generalized linear regression methods

Aside from the decision tree models, we also built Generalized Linear Mixed Effects Models (GLMMs) to explore whether we need to consider the effects of contextual and attitudinal predictors differently in different cultures (RQ3). Since the responses of participants in the same country may be associated in our dataset, and since each participant's responses to the 6 scenarios may also be associated, our GLMM accounts for the within-group/subject association via the introduction of random effects at both the country and the individual level. We used the software package R and specifically the lme4 library for GLMM. We specified a logit link function since our outcome variable is binary. We converted the categorical variables into dummies and normalized the continuous variables. To select models, we used analysis of deviance to examine the model fit. The outcome variable is the acceptability of the scenarios. Due to the high correlation among Hofstede's dimensions, we did not include all 6 dimensions in one single model. Instead, only one cultural dimension was entered at a time. Contextual factors and attitudinal factors

		Prediction	
		Positive	**Negative**
Truth	**Positive**	True positive (tp)	False negative (fn)
	Negative	False positive (fp)	True negative (tn)

Fig. 1. Confusion Matrix

were entered separately. Thus, we developed 6 regression models containing interaction effects between one of the 6 cultural values and all 4 contextual factors, and another 6 regression models containing interaction effects between one of the 6 cultural values and all 7 attitudinal factors. We controlled demographic factors in the models.

4 Results

4.1 Supervised learning results

In this section, we answer RQ1 and RQ2 using supervised machine learning.

4.1.1 Performance of culture-related factors

Our main results regarding the performance of culture-related factors are presented in Table 1. The 7 baseline models in our experiments are those without any culture-related predictors. We add country, language and cultural dimensions one by one to the baseline models to test whether the addition of culture-related factors improves the overall performance of the classifiers.

When country of residence is added as a predictor, the accuracy and F-measure generally improve. Country has the same top-ranked information gain as some of Hofstede's cultural dimensions (see Table 2).

The language variable has five values, English, Chinese, French, Swedish and German. It distinguishes between the two languages spoken in Canada (French and English), but not between countries that share a language (e.g. UK, US, Australia, India and a part of Canada). The language variable thus provides different information compared to the country variable. When language is added as a predictor, the accuracy and F-measure also improve compared to the baseline, but language provides less information gain.

When Hofstede's cultural dimensions are added to the baseline features, the accuracy and F-measure are overall higher than when either language or country is added. The two exceptions are when the baseline features are purely demographic or purely attitudinal: Hofstede's cultural dimensions then do not outperform country or language as predictors. Most importantly, we find that adding Hofstede's cultural dimensions to all the baseline features is better than adding country or language. Adding country, language, or even both, does not further improve the accuracy and F-measure when Hofstede's cultural dimensions are already among the baseline features.

Baseline Feature	Culture-Related Feature	Acc. (%)	F
demographic		74.5463	0.637
	+country	**75.1385**	**0.691**
	+language	75.0139	0.681
	+Hofstede	75.0368	0.687
contextual		75.2056	0.690
	+country	75.8658	0.715
	+language	75.8745	0.708
	+Hofstede	**76.3333**	**0.733**
attitudinal		75.4680	0.693
	+country	75.9364	0.706
	+language	**76.1416**	**0.714**
	+Hofstede	75.9771	0.712
demographic +contextual		75.2965	0.688
	+country	75.9329	0.714
	+language	75.8779	0.708
	+Hofstede	**76.3247**	**0.735**
demographic +attitudinal		75.568	0.695
	+country	76.032	0.710
	+language	**76.1541**	0.710
	+Hofstede	75.9827	**0.714**
contextual +attitudinal		76.0831	0.721
	+country	76.3424	0.729
	+language	76.5571	0.729
	+Hofstede	**76.7229**	**0.740**
demographic +contextual +attitudinal		76.1580	0.722
	+country	76.3545	0.730
	+language	76.5286	0.731
	+Hofstede	**76.7978**	**0.741**
	+country+language	76.5333	0.731
	+country+Hofstede	**76.7978**	**0.741**
	+language+Hofstede	**76.7978**	**0.741**
	+country+language+ Hofstede	**76.7978**	**0.741**

Note: Best performance for each category is in boldface.

Table 1. Accuracies and F-measures for each model.

Table 2 describes the results from feature selection. First, if there are no contextual features in the baseline (i.e. demographic + attitudinal + all culture-related features), the five most relevant features are all culture-related (i.e. they have the largest information gain). Second, if contextual features are included, the most relevant feature is the combination of the type of personal data being collected, the collection method and the device from which data is collected, followed by the same five culture-related features. Inspecting our Decision Tree models, we also observe that the culture-related features are at the top levels of the tree structures. Overall, culture-related features are very impactful on the classification.

With Contextual Features		Without Contextual Features	
IG	**Feature Name**	**IG**	**Feature Name**
0.043	*collection method*	0.042	uncertainty avoidance
0.042	uncertainty avoidance	0.042	country
0.042	country	0.042	individualism
0.042	individualism	0.042	indulgence
0.042	indulgence	0.041	power distance
0.042	power distance	0.035	*importance of notice and control*
0.035	*importance of notice and control*	0.033	pragmatism
0.033	pragmatism	0.031	*perception of social privacy threats*
0.031	*perception of social privacy threats*	0.029	*technology dependence*

*Note: features that are **not** cultural factors are italicized.*

Table 2. Most relevant features from feature selection.

4.1.2 Comparison of Hofstede's cultural dimensions

In Section 4.1.1, we found that Hofstede's cultural dimensions outperform language and country as predictors of acceptability in almost all cases. In this section, we analyze which of the 6 cultural dimensions perform(s) best. Our results are detailed in Table 3. We include all the baseline features and add Hofstede's cultural dimensions one by one. Adding power distance, individualism, uncertainty avoidance, and indulgence to the baseline improves the accuracy and F-measure of the model. Particularly, adding either individualism or indulgence results in the same performance as adding all of Hofstede's cultural dimensions, and is thus sufficient for achieving top performance. Additionally, feature selection (Table 2) shows that uncertainty avoidance, individualism, indulgence, and power distance are more relevant (i.e. have a larger information gain) than pragmatism and masculinity.

Baseline Feature	Culture-Related Feature	Acc. (%)	F
demographic +contextual +attitudinal		76.1580	0.722
	+power distance (PDI)	76.7394	0.740
	+individualism (IND)	**76.7978**	**0.741**
	+masculinity (MAS)	76.1641	0.721
	+uncertainty avoidance (UAI)	76.5861	0.729
	+pragmatism (PRA)	76.6104	0.734
	+indulgence (IDL)	**76.7978**	**0.741**
	+IND+IDL	**76.7978**	**0.741**
	+PDI+IND	**76.7978**	**0.741**
	+PDI+IND+MAS	**76.7978**	**0.741**
	+PDI+IND+MAS+UAI	**76.7978**	**0.741**
	+PDI+IND+MAS+UAI +PRA	**76.7978**	**0.741**
	+PDI+IND+MAS+UAI +PRA+IDL	**76.7978**	**0.741**

Note: Best performance is in boldface.

Table 3. Accuracies and F-measures when using Hofstede's cultural dimensions.

4.1.3 Robustness analysis

We conclude our supervised learning study with three robustness analyses. First, to ensure that the above results are not limited to Decision Trees, we also try other algorithms, such as Naive Bayes, Logistic Regression and k Nearest Neighbors (k-NN). The results are shown in Table A3 in the Appendix. The results from Logistic Regression coincide with the results from Decision Tree, in that Hofstede's cultural dimensions outperform country and language when added to the baseline features. Using Naive Bayes and k-NN yields different results: for these algorithms, the performance gets worse when adding country, language or Hofstede's dimensions. This somewhat surprising result may be due to overfitting in these models. Second, to ensure that our results are not unduly influenced by a single country, we run each model 8 more times, each time with one country excluded. We find that the results stay the same. Third, to ensure that our results are not limited to Hofstede's cultural dimensions, we also try Schwartz's framework of cultural values [41] and the Global Leadership and Organizational Behavior Effectiveness (GLOBE) cultural competencies [24]. In both cases we can replicate our result that adding cultural factors improves prediction accuracy more than adding country of residence or language spoken. When using Hofstede's dimensions, optimal prediction accuracy was reached by adding either individualism or indulgence. Similarly, embeddedness or egalitarianism can be added for optimal accuracy when using Schwartz's framework, and a combination of assertiveness, gender egalitarianism, and either future orientation or humane orientation when using GLOBE. This matches Schwartz's finding that Hofstede's individualism is positively correlated with egalitarianism and negatively with embeddedness [41], as well as Hofstede's finding that GLOBE's future orientation and gender egalitarianism are positively correlated with individualism [18]. Thus, individualism, egalitarianism, embeddedness, future orientation or gender egalitarianism can better predict acceptability than other cultural dimensions.

4.2 Regression results

To explore whether the impact of context and attitude on acceptability of personal data collection varies along cultural dimensions, we examine the interaction effects between individualism and contextual/attitudinal factors (RQ3). While almost all the six cultural dimensions have significant interaction effects with contextual and attitudinal factors, we only report models using individualism, because: 1) the effect size of these interaction effects is comparatively large; 2) individualism, power distance, and uncertainty avoidance are highly correlated [21]; 3) it is in line with our supervised learning result that individualism is one of the most relevant cultural factors; and 4) individualism has been well studied to distinguish culture and societies [47].

The general structure of the regression models including interaction effects between individualism and contextual factors is the following:

$$Logit(Y_{ijk}) = \beta_0 + \sum \beta_a CX_a + \beta_b IND + \sum \beta_c CX_a IND + s_k + r_i + e_n$$

where,
$Y_{ijk} = j^{th}$ response of i^{th} participant in k^{th} country
$\beta_0 =$ intercept
$CX_a = a^{th}$ contextual factors (a = 1, 2, 3, 4)
$\beta_a =$ coefficients of contextual factors CX_a
$IND =$ value of individualism
$\beta_b =$ coefficient of individualism
$\beta_c =$ coefficients of interaction effects
$s_k =$ country-level random effect
$r_i =$ subject-level random effect
$e_n =$ residual

The results from analysis of deviance in Table 4 show that adding individualism significantly improves the model fit, and so does adding the interactions of individualism with the contextual predictors.

	Dev	Chisq	DF	p
Contextual predictors	198256		17	<.001
IND	196814	1442.2	18	<.001
Contextual predictors × IND	186530	10284	58	<.001

Table 4. Analysis of deviance for context × IND.

In the next sections, we will interpret the interaction effects in more detail. To compare the effects of contextual variables in countries with different individualism levels, we compute the odds ratios (OR) for China (CN), India (IN), Germany (DE), Canada (CA) and US, since Australia and the UK have similar individualism as US, and Sweden as Germany (see "Hofstede's cultural values" in Table A2).

4.2.1 Interaction effects between individualism and usage purpose of the data

We find that the effect of usage purpose is significantly different between countries with different levels of individualism. Table 5 and Figure 2 show that the probabilities of acceptability decrease more strongly in individualistic countries than in less individualistic countries when the usage purpose changes from "as I agreed" to "to autonomously make decisions for me" or "to customize the options presented to me". For example, in more individualistic countries like US, the odds of accepting data collection in order "to autonomously make decisions for me" are predicted to be 88.9% lower than when the usage purpose is "as I agreed". But in less individualistic countries like China, the odds are only 50.6% lower. In the US, the odds of accepting data collection in order "to customize the options presented to me" are predicted to be 73.6% lower than when the usage purpose is "as I agreed". But in China, the odds are only 44.1% lower.

Usage Purpose	β	$\beta(\times \text{IND})$			
As I agreed	-	-			
To autonomously make decisions	-1.75***	-0.49***			
To customize the options	-1.11***	-0.24***			
	OR				
	US	CA	DE	IN	CN
Compared to "as I agreed":					
To autonomously make decisions	0.111	0.141	0.186	0.279	0.494
To customize the options	0.264	0.299	0.344	0.423	0.559

*Note:**$p<.05$, **$p<.01$, ***$p<.001$.

Table 5. Regression coefficients and odds ratios of usage purpose × IND.

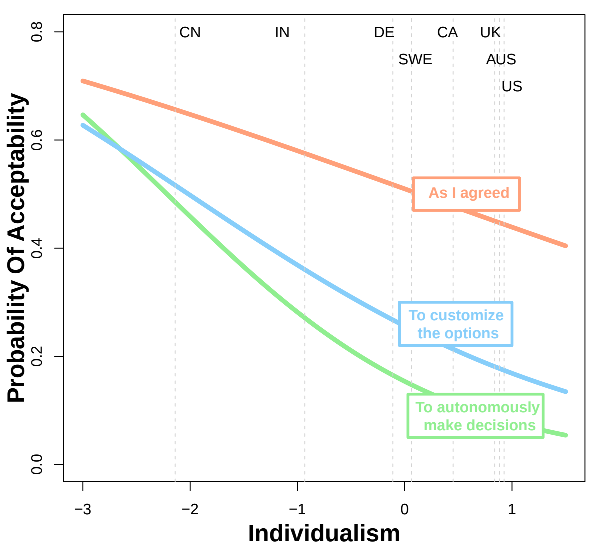

Fig. 2. Interaction between usage purpose and individualism on probability of acceptability.

4.2.2 Interaction effects between individualism and value exchange from data collection

Table 6 shows that the effect of value exchange is significantly different between individualistic and collectivistic countries. The effects of different types of value exchange are stronger in countries with lower levels of individualism. For example, in the US, the odds of accepting data collection when it "benefits the community" are predicted to be 23.6% higher than with "no other benefit". But in China, they are 128.3% higher. Likewise, the odds of accepting data collection when it "saves me time or money" are 57.8% higher than with "no other benefit" in the US, while they are 142% higher in China. Finally, the comparative odds for "a unique or compelling value" are 57.7% higher in the US but 153.3% higher in China.

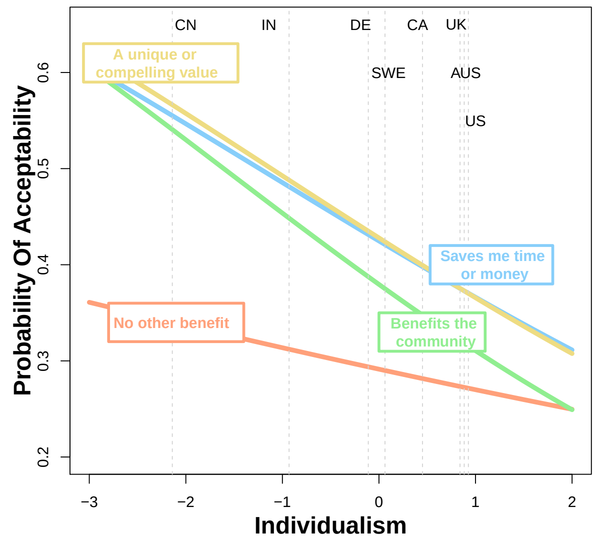

Fig. 3. Interaction between value exchange and individualism on probability of acceptability.

As shown in Figure 3, the probabilities for "benefits the community", "saves me time or money" and "a unique or compelling value" are fairly similar in less individualistic countries. When the value exchange switches from any of the three to "no other benefit", the probabilities of acceptability decreases pronouncedly in less individualistic countries but stays roughly the same in individualistic countries. The overall tendencies of "saves me time or money" and "a unique or compelling value" are similar. "Benefits the community" is a significantly less acceptable value exchange in highly individualistic countries.

4.2.3 Interaction effects between individualism and collecting entity

Table 7 shows that there is a significant interaction between individualism and the type of entity that collects users' personal data. Compared to "a service provider that provides free services" (increasingly the most com-

Value Exchange	β	$\beta(\times$IND)
No other benefit	-	-
Benefits the community	0.40***	-0.20***
Saves me time or money	0.59***	-0.14***
A unique or compelling value	0.60***	-0.15***

	OR				
	US	CA	DE	IN	CN
Compared to "no other benefit":					
Benefits the community	1.236	1.360	1.521	1.792	2.283
Saves me time or money	1.578	1.686	1.824	2.045	2.420
A unique or compelling value	1.577	1.697	1.851	2.101	2.533

*Note:*p<.05, **p<.01, ***p<.001.*

Table 6. Regression coefficients and odds ratios of value exchange $\times$ IND.

mon type of service provider on the web), certain collecting entities have a negative interaction with individualism (italicized in Table 7): people in collectivist cultures such as China and India are relatively more accepting of data collection performed by the government, an employer or a foreign service provider than people in individualist cultures such as the US and Canada. Other recipients have a positive interaction with individualism (bolded in Table 7): people in individualist cultures are relatively more accepting of data collection when they either pay for, or already have an existing relationship with the service provider. Compared to "a service provider of free services, the acceptability of "unfamiliar" and "well-known" service providers as well as ones "in my country" is not affected by individualism.

4.2.4 Interaction effects between individualism and collection methods

Collection methods include 15 different categorical values. The interaction between collection method and individualism is significant (see Table 8). In a more detailed comparison, the odds of accepting data collection of bank account numbers, government issued ID, and medical history are lower from mobile devices than from computers. In more individualistic countries, however, this decrease of odds is much stronger than in collectivist countries (see Table 9).

Entity	β	$\beta(\times$IND)
A SP of free services	-	-
A SP in my country	-0.03	0.01
My employer	*-0.20****	*-0.07***
A SP not in my country	*-0.35****	*-0.09****
A SP that I pay for services	**0.13*****	**0.07****
A SP I have an existing relationship with	**0.25*****	**0.06***
An unfamiliar SP	-0.52***	-0.04.
A well-known SP	0.25***	0.01
The government	*0.05**	*-0.08***

	OR				
	US	CA	DE	IN	CN
Compared to "a SP of free services":					
A SP in my country	0.987	0.980	0.972	0.960	0.943
My employer	*0.764*	*0.790*	*0.823*	*0.873*	*0.953*
A SP not in my country	*0.645*	*0.675*	*0.711*	*0.768*	*0.860*
A SP that I pay for services	**1.220**	**1.178**	**1.131**	**1.066**	**0.977**
A SP I have an existing relationship with	**1.359**	**1.323**	**1.281**	**1.223**	**1.141**
An unfamiliar SP	0.570	0.581	0.595	0.617	0.649
A well-known SP	1.300	1.291	1.282	1.268	1.248
The government	*0.983*	*1.019*	*1.064*	*1.133*	*1.243*

*Note: *p<.05, **p<.01, ***p<.001; SP stands for "Service Provider".*

Table 7. Regression coefficients and odds ratios of entity $\times$ IND.

When the data type is location and contact list, the odds of accepting data collection are predicted to be 24.1% higher in the US if the data collection occurs on a computer rather than a mobile device, but 10.6% lower in China. This indicates that users in more individualistic countries are more willing to provide location and contact list through computers while users in less individualistic countries preferably through mobile devices.

When the data type is a photo or video image taken of them, users in more individualistic countries are relatively more accepting if it is taken in public, while users in collectivist countries accept it more at work. In all countries, users find it relatively unacceptable if a person that they know can provide their photo or video image.

Collection Method	β	$\beta(\times$IND)
I will provide my bank account number (to entity) from:		
• my computer	-	-
• my mobile device	-0.28***	-0.08.
I will provide my government issued ID (to entity) from:		
• my computer	0.53***	-0.11**
• my mobile device	0.38***	-0.16***
My current location and contact list:		
• I will provide from my mobile device	0.17***	-0.13***
• I will provide from my computer	0.36***	-0.03
• (Entity) can collect when I am using my mobile device	0.24***	-0.13**
I will provide my medical history (to entity) from:		
• my computer	0.53***	-0.11*
• my mobile device	0.12**	-0.27***
A photo or video image of me is taken:		
• at home	-1.05***	-0.34***
• in a public space	-0.77***	-0.28***
• at work	-0.72***	-0.37***
• by a person that I know (to entity)	-1.65***	-0.34***
My purchase history that entity can:		
• collect when I am using a computer	0.13***	-0.11**
• use to generate other information about me	0.17***	-0.16***

*Note: *p<.05, **p<.01, ***p<.001.*

Table 8. Regression coefficients of collection methods × IND.

When the data type is purchase history, the acceptance odds in the US are 9% lower when it is collected to generate other information about the user than when the data is only collected as a byproduct. The odds are 17.3% higher in China though.

4.2.5 Interaction effects between individualism and attitudinal factors

Finally, we also built models with interaction terms between attitudinal factors and individualism, to explore whether the impact of privacy attitudes on privacy decisions varies in different cultures. The equation structure is similar to that of contextual factors:

$$Logit(Y_{ijk}) = \beta_0 + \sum \beta_a ATT_d + \beta_b IND + \sum \beta_c ATT_d IND + s_k + r_i + e_n$$

where,
$ATT_d = d^{th}$ attitudinal factor (d = 1..7)
β_a = coefficients of contextual factors ATT_d
Other notations remain the same.

Collection Method	OR				
	US	CA	DE	IN	CN
I will provide my bank account number (to entity) from:					
Compared to "my computer":					
my mobile device	0.701	0.730	0.764	0.818	0.904
I will provide my government issued ID (to entity) from:					
Compared to "my computer":					
my mobile device	0.824	0.842	0.865	0.900	0.953
My current location and contact list:					
Compared to "my mobile device":					
• I will provide from my computer	1.241	1.179	1.111	1.017	0.894
• (Entity) can collect when I am using my mobile device	0.936	0.934	0.931	0.927	0.922
I will provide my medical history (to entity) from:					
Compared to "my computer":					
my mobile device	0.570	0.616	0.674	0.771	0.938
A photo or video image of me is taken:					
Compared to "at home":					
• in a public space	1.394	1.356	1.313	1.252	1.167
• at work	1.348	1.369	1.394	1.432	1.490
• by a person that I know (to entity)	0.554	0.552	0.551	0.548	0.545
My purchase history that entity can:					
Compared to "collect when I am using a computer":					
use to generate other information about me	0.991	1.017	1.049	1.097	1.173

Table 9. Odds ratios of collection methods × IND.

As shown in Table 10, adding individualism and the interaction terms significantly improves the model fit. The interaction effects between 7 attitudinal variables and individualism are significant except for "perception of personal information (narrow)" (see Table 11).

We find the interactions between "perception of third-party accountability" and individualism, and "perception of personal information (broad)" and individualism, particularly interesting (bolded in Table 12). For example, for a one-unit increase in participants' perception that a third party is accountable for personal data collection, the odds of acceptability are predicted to be 38% higher in the US, whereas they are 33% lower in China. For a one-unit increase in participants' perception that broad personal data, such as photos, or

	Dev	Chisq	DF	p
Attitudinal predictors	216022		10	<.001
IND	215067	954.21	11	<.001
Attitudinal predictors × IND	214946	120.95	18	<.001

Table 10. Analysis of deviance for attitude × IND. Including individualism (and its interaction with attitudinal predictors) in our regression model significantly improves the model fit.

Attitude	β	β(×IND)
Importance of notice and control	0.42***	0.07**
Perception of third-party accountability	0.10***	0.23***
Technology dependence	-0.42***	0.08***
Perception of threat to personal information	0.26***	-0.10*
Perception of social threat	-0.19***	0.13**
Perception of personal information (narrow)	0.05	-0.06
Perception of personal information (broad)	0.07	-0.10*

*Note: *p<.05, **p<.01, ***p<.001.*

Table 11. Regression coefficients of attitude × IND.

Attitude	OR				
	US	CA	DE	IN	CN
Importance of notice and control	1.63	1.57	1.51	1.42	1.30
Perception of third-party accountability	**1.38**	**1.23**	**1.08**	**0.89**	**0.67**
Technology dependence	0.70	0.68	0.65	0.61	0.56
Perception of threat to personal information	1.18	1.24	1.32	1.43	1.62
Perception of social threat	0.93	0.88	0.82	0.74	0.63
Perception of personal information (narrow)	1.00	1.03	1.06	1.12	1.20
Perception of personal information (broad)	**0.98**	**1.02**	**1.08**	**1.17**	**1.31**

Table 12. Odds ratios of attitude × IND.

employment history, can be used to identify them by service providers, the odds of acceptability are 31% higher in China, but there is no appreciable effect for the US (a mere 2% decrease). This indicates that the attitudes towards third-party accountability regarding personal data collection and the acceptable use of broad personal data have different consequences for the acceptability of data collection in different cultures.

5 Discussion

In this study, we examine the effect of culture on privacy decisions by building an integrated, cross-cultural prediction model of users' acceptability regarding the collection of their personal data. We find that incorporating culture-related predictors, such as country of residence, language, and Hofstede's cultural dimensions, can aid in improving the prediction accuracy when added to demographic, attitudinal and contextual predictors. Hofstede's dimensions - particularly individualism and indulgence - outperform country and language predictors in terms of prediction performance. Uncertainty avoidance, power distance and pragmatism also significantly differentiate users' acceptability.

Additionally, we find that the impact of contextual and attitudinal predictors varies in different cultures. Although further studies with more countries are needed to create a comprehensive list of culturally-dependent effects, we highlight the following interesting cultural contrasts (see Table 13):

Our study confirms previous findings that contextual factors are important predictors of users' privacy decisions. Previous research has shown that collection method [43], device, and data type [36, 42] relate to consumers' beliefs and behavioral responses to privacy threats. Our study also largely confirms that culture-related factors have a significant impact on users' privacy decisions. However, compared to previous work, our supervised learning approach is unique in its ability to test the improvement in predicting users' acceptability by adding cultural predictors in an integrated model. In contrast, Milberg et al.'s [31, 32] work focuses on the particular effect of cultural predictors on users' privacy concerns (data collection, unauthorized secondary use, errors and improper access) without considering contextual and attitudinal predictors.

Altman's work provides a possible explanation for the improvement that culture makes to the prediction of users' privacy behaviors: according to Altman's theory, the behavioral mechanisms to control the accessibility to oneself are "unique to the particular phys-

Factors	Individualistic Countries (e.g. US)	Collectivistic Countries (e.g. CN)
Usage Purpose		+Autonomously make decisions +Customize the options
Value Exchange		+Benefit the community +Saving time or money +Unique or compelling value
Entities	+Paid service +Existing relationship	+Government +Employer +Foreign service provider
Collection Method	+Through computer +Photo or video image in public	+Through mobile +Photo or video image at work
Attitude	+Perception of third-party accountability -Perception of personal info broad	-Perception of third-party accountability +Perception of personal info broad

Note: + means "increase acceptability more strongly" and - means "decrease acceptability more strongly".

Table 13. Summary of interaction effects.

ical, psychological and social circumstances of a culture" [3]. Altman argues that different cultures have developed distinct mechanisms for regulating behaviors with strangers, acquaintances, and families, which enables people to control their accessibilities [3]. Hofstede also suggests that "culture" is comprised of firmly held beliefs that guide people's attitudes and behaviors [20]. Thus, culturally distinct beliefs may govern users' privacy decisions, creating the culturally distinct patterns of behavior that are captured by our models.

This study also demonstrates the superiority of Hofstede's cultural dimensions over country of residence and language in predicting privacy behaviors. No prior work has empirically found this result; most cross-cultural privacy studies apply Hofstede's cultural dimensions without making such comparisons [7, 11, 13, 31].

We also confirm the effect of each cultural dimension in improving the prediction accuracy. This is to some degree similar to Milberg et al.'s [32] results that power distance, individualism and uncertainty avoidance have a significant effect on people's overall level of information privacy concerns. Power distance may be a relevant predictor of privacy decisions because it influences how tolerant people are when authorities violate their privacy. Similarly, individualism may be a relevant factor because people in individualistic cultures have a relatively higher regard for their personal privacy [51]. Indulgence, however, is an understudied cultural predictor in previous privacy research, not the least since it was only introduced around 2010. But it turns out to be a strong predictor in our supervised learning results (see Section 4.1.2). Indulgence is characterized by "personal life control" and "importance of freedom". To protect their freedom, individuals may resist having their personal space or privacy invaded [10]. Thus, privacy protects freedom and control over personal life. This may possibly explain why indulgence has impact on acceptability.

The interaction effects between contextual factors and culture are our unique contributions, as no previous research has proposed them. Contextual predictors have different impact on acceptability between individualistic and collectivistic countries. This may be due to the different cultural traditions and beliefs. For example, people in individualistic countries value their individual rights and privacy, thus they are more reluctant to accept the data collection even if it brings them benefits such as saving me time or money, presenting compelling value or benefitting the community. On the other hand, people in collectivistic societies find it acceptable to disclose personal information to the government or employers. Research has shown that members of collectivist societies are guided more by group norms and traditional authority [47]. They tend to subjugate individual rights and goals for a sense of commitment to the group and of self-sacrifice for the common good. Thus people in such cultures will agree to data collection performed by the government or employers. However, in individualistic societies this is unacceptable. In Elazar's [48] description of an individualistic culture, government is viewed as "instituted for strictly utilitarian reasons, to handle those functions demanded by the people it is created to serve". Thus, people in such cultures may feel less obligated to disclose personal information to the authority if there is no specific demand for it. People in individualist societies are more willing to disclose their data to paid services or services that they have an existing relationship with. This may be because people in individualist societies are more comfortable with contract-based relations. They are more likely to accept the data collection if an agreement has been formally established.

Another highlighted interaction effect is that users in individualistic countries prefer data collection through computers, while users in collectivist countries prefer it through mobile devices. Zhang et al. [52] have found that perceived usefulness is more important in western culture while perceived ease of use is more important in eastern culture. This finding may give rise to an explanation, namely that users in individualistic countries may find computers more useful and more suited to control the disclosure of personal data, while users in collectivistic countries may view mobile devices as easier to use in data disclosure. Additionally, people in individualistic countries are more comfortable if their photo or video image is taken in public, while users in collectivist countries accept it more at work. The latter may again be explained by the fact that people in collectivist countries feel more group commitment at their workplace and are willing to sacrifice their own privacy for the group norms and goals.

One limitation of our study is that we only have data from 8 different countries. Although they are representative to some extent of the existing privacy/data protection landscape and social/cultural attitudes, our study would have been more complete if more countries were surveyed. A second limitation is that the outcome variable measures participants' acceptability of the different information disclosure scenarios. Although people's perception or intent is closely related with actual behavior [2], especially when both are measured in a specific context, studies have shown that users may behave in ways that seemingly contradict their stated personal information disclosure intentions [35]. However, since we cannot implement real-world scenarios to collect participants' actual behavioral response due to time and cost, this is the closest reaction that we can get. Our work is in fact a step beyond previous research on privacy and culture, which has mainly considered general privacy attitudes or concerns rather than context-specific behaviors or behavioral intentions. We will attempt to examine actual behavior in future research to get more precise insights into how culture influences privacy decisions.

Hofstede's cultural dimensions have been criticized for equating culture with nation, lacking representativeness, not being exhaustive, and reducing culture to numbers and dimensions, while culture may be more accurately viewed as an integrated system of ideas and values [6]. Hofstede's cultural framework is however one of the most widely used cultural frameworks to measure culture in psychology, sociology, marketing, management, etc. His sample is the most comprehensive to date. The dimensions emerged from factor analysis on data related to work goals. Studies that replicated his work were on different domains of values in different samples, but reached significant convergence with most of Hofstede's dimensions, especially individualism and power distance [12, 44]. Other studies have proposed different sets of cultural dimensions, but were also shown to be significantly correlated with Hofstede's dimensions [41]. We tested Schwartz's framework of cultural values and the Global Leadership and Organizational Behavior Effectiveness (GLOBE) cultural competencies in our machine learning studies. They did not exhibit major differences from using Hofstede's dimensions. Additionally, Hofstede's cultural dimensions have been validated by establishing the significance of their correlations with geographic and social indicators as well as measures of economic development.

Unfortunately, hardly any cultural dimension data have been collected as yet at a regional level. Three of the countries in our survey have areas with distinct native languages, and those areas may possess somewhat different cultures. Studying privacy prediction and cultural interaction at such a finer granularity would be an interesting extension of our work.

6 Implications for Privacy Research and PET

This study provides several implications for future privacy research and privacy-enabling technology. It highlights the need to consider the cultural background of participants when examining the determinants of information privacy concerns or defining privacy-related policies. Privacy studies should either be explicitly restricted to a single culture, or the cultural background should be controlled for. It is also better to use cultural frameworks to measure cultural background than using country or language.

Another implication for practitioners, especially organizations that operate globally, is that users' acceptance of personal data collection is different in different cultures, even given the same context. For example, users in individualist countries may find it acceptable to disclose personal data to a service provider that they pay for. The same may be unacceptable in collectivist countries though, and service providers will need to take this into account when establishing privacy policies. Similarly, users in individualist countries may find it acceptable to share their current location through a computer, while users in collectivistic countries through a mobile device. Thus, organizations may need to tailor their data collection methods to different cultures in an effort to make users feel more comfortable.

Moreover, our results also show that the links between attitudes/opinions and behaviors are culturally different. Certain beliefs (e.g. about accountability and identifiability) have opposite effects in individualist and collectivist cultures. This cultural variability of attitudinal effects may be yet another reason for the existence of the privacy paradox [35], i.e. the surprisingly weak link between privacy attitudes and behaviors. It also warns researchers that designs, policies, or technical interventions that are created to increase the acceptability of disclosure by countering such beliefs in one culture may be completely counter-productive when implemented in another culture. For example, informing users that third-party companies are accountable for the collection, access and use of personal data in individualistic countries like US will make them more likely to disclose. But such intervention will bring counter-productive effect in collectivistic countries like China.

Our study also implicates the development of privacy decision-support systems [16, 27]. As users' information disclosure decisions are highly dependent on the context, context-aware privacy decision-support systems assist users based on contextual cues, such as location, device, etc. Our study suggests that cultural background needs to be incorporated as an additional factor if these systems operate globally. For example, the system can determine the users' country based on IP address, and get their cultural values. Together with other contextual cues, such as location, device, etc., it can make a customized recommendation to users tailored to their specific cultural background and context, when some third-party is requesting the users' personal information.

7 Conclusion

To examine the impact of cultural background on privacy decisions, we built models based on survey data of 9625 participants from 8 different countries to predict the acceptability of personal data disclosure. Using supervised machine learning, we demonstrated that adding cultural indicators significantly improves prediction accuracy. Using GLMM regression, we also demonstrated that the impacts of contextual and attitudinal factors vary by culture. We suggest that users' cultural background should be taken into account in future privacy research and practice.

8 Acknowledgment

We would like to thank our anonymous journal reviewers as well as Gloria Mark and Judith Olson for their comments on earlier versions of this article. The described research was in part supported by the NSF under grant SES-1423629.

References

[1] A. Acquisti and J. Grossklags. Privacy Attitudes and Privacy Behavior. In L. Camp and S. Lewis, editors, *Economics of Information Security*, volume 12 of *Advances in Information Security*, pages 165–178. Springer US, 2004.

[2] I. Ajzen. The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50(2):179–211, 1991.

[3] I. Altman. Privacy Regulation: Culturally Universal or Culturally Specific? *Journal of Social Issues*, 33(3):66–84, July 1977.

[4] A. Anton, J. Earp, and J. Young. How internet users' privacy concerns have evolved since 2002. *IEEE Security Privacy*, 8(1):21–27, Jan. 2010.

[5] K. Bartel Sheehan. An investigation of gender differences in on-line privacy concerns and resultant behaviors. *Journal of Interactive Marketing*, 13(4):24–38, 1999.

[6] R. F. Baskerville. Hofstede never studied culture. *Accounting, Organizations and Society*, 28(1):1–14, Jan. 2003.

[7] S. Bellman, E. J. Johnson, S. J. Kobrin, and G. L. Lohse. International Differences in Information Privacy Concerns: A Global Survey of Consumers. *The Information Society*, 20(5):313–324, Nov. 2004.

[8] M. Benisch, P. G. Kelley, N. Sadeh, and L. Cranor. Capturing Location-Privacy Preferences: Quantifying Accuracy and User-Burden Tradeoffs. Technical Report CMU-ISR-10-105, 2010.

[9] I. Bilogrevic, K. Huguenin, B. Agir, M. Jadliwala, and J.-P. Hubaux. Adaptive Information-sharing for Privacy-aware Mobile Social Networks. In *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp '13, pages 657–666, New York, NY, USA, 2013. ACM.

[10] S. S. Brehm and J. W. Brehm. *Psychological Reactance: A Theory of Freedom and Control.* Academic Press, Oct. 2013.

[11] H. Cho, M. Rivera-Sánchez, and S. S. Lim. A multinational study on online privacy: global concerns and local responses. *New Media Society*, 11(3):395–416, May 2009.

[12] T. C. C. Connection. Chinese Values and the Search for Culture-Free Dimensions of Culture. *Journal of Cross-Cultural Psychology*, 18(2):143–164, June 1987.

[13] T. Dinev, M. Bellotto, P. Hart, V. Russo, and I. Serra. Internet Users' Privacy Concerns and Beliefs About Government Surveillance: An Exploratory Study of Differences Between Italy and the United States. *Journal of Global Information Management*, 14(4):57–93, 2006.

[14] T. Dinev, M. Bellotto, P. Hart, V. Russo, I. Serra, and C. Colautti. Privacy calculus model in e-commerce – a study of Italy and the United States. *Eur J Inf Syst*, 15(4):389–402, 2006.

[15] C. Dong, H. Jin, and B. P. Knijnenburg. Predicting Privacy Behavior on Online Social Networks. In *Ninth International AAAI Conference on Web and Social Media*, pages 91–100. AAAI Publications, Apr. 2015.

[16] L. Fang and K. LeFevre. Privacy Wizards for Social Networking Sites. In *Proceedings of the 19th International Conference on World Wide Web*, WWW '10, pages 351–360, Raleigh, NC, 2010. ACM.

[17] R. L. Helmreich and A. R. Merritt. *Culture at work in aviation and medicine: national, organizational and professional influences*. Ashgate, 2001.

[18] G. Hofstede. What did GLOBE really measure? Researchers' minds versus respondents' minds. *J Int Bus Stud*, 37(6):882–896, Sept. 2006.

[19] G. Hofstede. Dimensionalizing Cultures: The Hofstede Model in Context. *Online Readings in Psychology and Culture*, 2(1), Dec. 2011.

[20] G. Hofstede and M. H. Bond. Hofstede's Culture Dimensions An Independent Validation Using Rokeach's Value Survey. *Journal of Cross-Cultural Psychology*, 15(4):417–433, Dec. 1984.

[21] G. H. Hofstede and G. Hofstede. *Culture's Consequences: Comparing Values, Behaviors, Institutions and Organizations Across Nations*. SAGE, Apr. 2001.

[22] G. Holmes, A. Donkin, and I. H. Witten. WEKA: a machine learning workbench. In *Proceedings of the 1994 Second Australian and New Zealand Conference on Intelligent Information Systems,1994*, pages 357–361, Nov. 1994.

[23] M. H. Hoppe. *A comparative study of country elites: International differences in work-related values and learning and their implications for management training and development*. PhD thesis, University of North Carolina at Chapel Hill, 1990.

[24] R. J. House, P. J. Hanges, M. Javidan, P. W. Dorfman, and V. Gupta. *Culture, Leadership, and Organizations: The GLOBE Study of 62 Societies*. SAGE Publications, Apr. 2004.

[25] L. K. John, A. Acquisti, and G. Loewenstein. Strangers on a Plane: Context-Dependent Willingness to Divulge Sensitive Information. *Journal of Consumer Research*, 37(5):858–873, Feb. 2011.

[26] B. P. Knijnenburg. Information Disclosure Profiles for Segmentation and Recommendation. In *SOUPS2014 Workshop on Privacy Personas and Segmentation*, Menlo Park, CA, 2014.

[27] B. P. Knijnenburg. *A user-tailored approach to privacy decision support*. Ph.D. Thesis, University of California, Irvine, Irvine, CA, 2015.

[28] B. P. Knijnenburg, A. Kobsa, and H. Jin. Dimensionality of information disclosure behavior. *International Journal of Human-Computer Studies*, 71(12):1144–1162, 2013.

[29] S. Lederer, J. Mankoff, and A. K. Dey. Who wants to know what when? privacy preference determinants in ubiquitous computing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 724–725, Ft. Lauderdale, FL, 2003. ACM Press.

[30] P. G. Leon, A. Kobsa, and C. Nguyen. Contextual Determinants for Users' Acceptance of Personal Data Processing: A Multinational Analysis (unpublished).

[31] S. J. Milberg, S. J. Burke, H. J. Smith, and E. A. Kallman. Values, Personal Information Privacy, and Regulatory Approaches. *Commun. ACM*, 38(12):65–74, Dec. 1995.

[32] S. J. Milberg, H. J. Smith, and S. J. Burke. Information Privacy: Corporate Management and National Regulation. *Organization Science*, 11(1):35–57, Feb. 2000.

[33] C. L. Miltgen and D. Peyrat-Guillard. Cultural and generational influences on privacy concerns: a qualitative study in seven European countries. *Eur J Inf Syst*, 23(2):103–125, Mar. 2014.

[34] H. Nissenbaum. *Privacy in Context: Technology, Policy, and the Integrity of Social Life*. Stanford University Press, Nov. 2009.

[35] P. A. Norberg and D. R. Horne. Privacy attitudes and privacy-related behavior. *Psychology and Marketing*, 24(10):829–847, 2007.

[36] J. Phelps, G. Nowak, and E. Ferrell. Privacy Concerns and Consumer Willingness to Provide Personal Information. *Journal of Public Policy & Marketing*, 19(1):27–41, Mar. 2000.

[37] S. Piramuthu. Input data for decision trees. *Expert Systems with Applications*, 34(2):1220–1226, Feb. 2008.

[38] C. Posey, P. B. Lowry, T. L. Roberts, and T. S. Ellis. Proposing the online community self-disclosure model: the case of working professionals in France and the U.K. who use online communities. *Eur J Inf Syst*, 19(2):181–195, Apr. 2010.

[39] J. R. Quinlan. *C4.5: Programs for Machine Learning*. Elsevier, June 2014.

[40] S. Rodgers and M. A. Harris. Gender and E-Commerce: An Exploratory Study. *Journal of Advertising Research*, 43(03):322–329, Sept. 2003.

[41] S. H. Schwartz. Are There Universal Aspects in the Structure and Contents of Human Values? *Journal of Social Issues*, 50(4):19–45, Jan. 1994.

[42] K. B. Sheehan and M. G. Hoy. Dimensions of Privacy Concern Among Online Consumers. *Journal of Public Policy & Marketing*, 19(1):62–73, Mar. 2000.

[43] H. J. Smith, S. J. Milberg, and S. J. Burke. Information Privacy: Measuring Individuals' Concerns about Organizational Practices. *MIS Quarterly*, 20(2):167–196, June 1996.

[44] P. B. Smith, S. Dugan, and F. Trompenaars. National Culture and the Values of Organizational Employees A Dimensional Analysis Across 43 Nations. *Journal of Cross-Cultural Psychology*, 27(2):231–264, Mar. 1996.

[45] J.-B. E. Steenkamp and I. Geyskens. How Country Characteristics Affect the Perceived Value of Web Sites. *Journal of Marketing*, 70(3):136–150, July 2006.

[46] E. Toch, J. Cranshaw, P. H. Drielsma, J. Y. Tsai, P. G. Kelley, J. Springfield, L. Cranor, J. Hong, and N. Sadeh. Empirical models of privacy in location sharing. In *Proceedings of the 12th ACM international conference on Ubiquitous computing*, pages 129–138, Copenhagen, Denmark, 2010. ACM Press.

[47] H. C. Triandis. *Individualism & collectivism*, volume xv of *New directions in social psychology.* Westview Press, Boulder, CO, US, 1995.

[48] J. A. Vandello and D. Cohen. Patterns of individualism and collectivism across the United States. *Journal of Personality and Social Psychology*, 77(2):279–292, Aug. 1999.

[49] P. Wang and L. A. Petrison. Direct marketing activities and personal privacy. A consumer survey. *J. Direct Mark.*, 7(1):7–19, Dec. 1993.

[50] J. Xie, B. P. Knijnenburg, and H. Jin. Location Sharing Privacy Preference: Analysis and Personalized Recommendation. In *Proceedings of the 19th International Conference on Intelligent User Interfaces*, IUI '14, pages 189–198, New York, NY, USA, 2014. ACM.

[51] Z. Yang, R. T. Peterson, and S. Cai. Services quality dimensions of Internet retailing: an exploratory analysis. *Journal of Services Marketing*, 17(7):685–700, Dec. 2003.

[52] L. Zhang, J. Zhu, and Q. Liu. A meta-analysis of mobile commerce adoption and the moderating effect of culture. *Computers in Human Behavior*, 28(5):1902–1911, Sept. 2012.

A Appendix

Role	Name (Type)	Definition
Demographic Features	Age (Numeric)	Age in years
	Gender (Binary)	Indicator of the participant's gender
	Work (Nominal)	Employment status (self-employed, retired, student, employed full-time or part-time)
	Education (Nominal)	Highest level of formal education
	Income (Nominal)	Grouped household income before taxes
	Job (Nominal)	Types of industries (accounting/tax, computer, education, government, etc.)
	Children (Numeric)	Number of children under age of 18 currently living in your household
	Race (Nominal)	Ethnic or racial background or heritage
Attitudinal Features	Technology dependence (Numeric)	Standardized factor score based on participants' agreement on a 7-point scale (strongly disagree – strongly agree) to the following items: • Technology is important to me for staying organized (e.g. scheduling, planning) • Technology is important to me for having fun / being entertained • I embrace technology and social media (e.g. blogs, Twitter, Facebook) to connect with others and express myself in new, interesting, and fun ways • It's important for me to have Internet access when I am "on-the-go" - away from home or work • I am constantly on the go, and my computer has to keep up so things like quick connectivity, fast start up / resume, and battery life are important • It is important for me to be able to keep my files and information (e.g. calendar, pictures, music, contacts) in sync across all my devices (e.g. computer, phone) • My mobile phone / smartphone should help me get work done when and where I want
	Perception of third-party accountability (Numeric)	Standardized factor score based on participants' selection of the following answers to the following check-all-that-apply question: "Which of the following do you believe companies (without *) / the government (with *) should be accountable for regarding the collection, access and use of your personal information?" • Being transparent as to how your personal information will be used • Using your personal information in a manner consistent with your expectations • Ensuring that employees handle your personal information appropriately • Not using your personal information without your consent • Notifying you if your personal information has been compromised • Giving you the ability to view, change and delete your personal information • Providing appropriate records showing how your personal information is used by regulatory agencies when requested • Deleting your personal information after the relationship is severed • Establishing regulations to protect you and ensure the responsible use of your personal information* • Monitoring to ensure responsible use of your personal information* • Enforcing regulations with appropriate punishment (e.g. sanctions or fines)*
	Perception of personal information – narrow (Numeric)	Standardized factor score based on participants' selection of the following answers to the following check-all-that-apply question: "Which of the following do you consider to be personal information that could be used to identify you by service providers such as online retailers, social networking sites, government agencies and businesses, including their mobile applications?" • Full name • Home address • Mobile phone number • Date of birth

	Perception of personal information – broad (Numeric)	Standardized factor score based on participants' selection of the following answers to the following check-all-that-apply question: "Which of the following do you consider to be personal information that could be used to identify you by service providers such as online retailers, social networking sites, government agencies and businesses, including their mobile applications?" • Photos or videos you have captured • Employment history • Friend or family member's contact information (email address, phone) • Current physical location • Your family members' locations • Bank or credit card number(s) • Medical history or records • Device identifier • IP address • Photos or videos of you captured by others in public places • Government-issued ID number
	Perception of threat to personal information (Numeric)	Standardized factor score based on participants' answer on a 7-point scale (extremely unlikely – extremely likely) to the following statements: "When you provide your personal information to a service provider, how likely is each of the following to happen to you?" • Your financial information (e.g. credit card number, bank account number) will be stolen • Your online identity will be compromised • Your account will be compromised
	Perception of social threat (Numeric)	Standardized factor score based on participants' answer on a 7-point scale (extremely unlikely – extremely likely) to the following statements: "When you provide your personal information to a service provider, how likely is each of the following to happen to you?" • Your reputation with your peers and/or family will be damaged • You will be discriminated against or penalized • You or your family will incur physical harm
	Importance of notice and control (Numeric)	Standardized factor score based on participants' answer on a 5-point scale (not at all important – extremely important) to the following statements: • Ability to control how my personal information is used when it is collected • Ability to control how my personal information will be used after it is collected • Knowing how my personal information will be used • Knowing who will use my personal information • Receiving benefits that are valuable to me (e.g. a financial reward, enhanced service, convenience) • Knowing that service providers will be held accountable for how my information is handled
Contextual Features	Collection method (Nominal)	"From my computer, I will provide my bank account number (to entity)" "From my mobile device, I will provide my bank account number (to entity)" "From my computer, I will provide my government issued ID (to entity)" "From my mobile device, I will provide my government issued ID (to entity)" "(Entity) can collect my current location when I am using my mobile device" "From my computer, I will provide my current location and contact list (to entity)" "From my mobile device, I will provide my current location and contact list (to entity)" "From my computer, I will provide my medical history (to entity)" "From my mobile device, I will provide my medical history (to entity)" "(Entity) can take a photo or video image of me at home" "(Entity) can take a photo or video image of me while I'm in a public space" "(Entity) can take a photo or video image of me at work"

		"A person that I know can provide my photo or video image (to entity)" "(Entity) can collect my purchase history when I am using a computer" "(Entity) can use my purchase history to generate other information about me"
	Entity(Nominal)	"a service provider of free services" "a service provider in my country" "my employer" "a service provider not in my country" "a service provider that I pay for services" "a service provider I have an existing relationship with" "an unfamiliar service provider" "a well-known service provider" "the government"
	Usage purpose of the data (Nominal)	"as I agreed" "to autonomously make decisions for me" "to customize the options presented to me"
	Value exchange from data collection (Nominal)	"benefits the community" "no other benefit" "saves me time or money" "a unique or compelling value"
Culture-Related Features	Country (Nominal)	8 countries where the survey was distributed
	Language (Nominal)	Languages spoken within household
	Power Distance (Numeric)	The extent to which the less powerful members of institutions and organizations within a country expect and accept that power is distributed unequally. The higher the score, the more acceptable to inequalities.
	Individualism (Numeric)	The degree of interdependence a society maintains among its members. The higher the score, the more individualistic its people are.
	Masculinity (Numeric)	What motivates people, wanting to be the best (masculine) or liking what you do (feminine). The higher the score, the more the society is driven by competition, achievement, and success.
	Uncertainty Avoidance (Numeric)	The extent to which the members of a culture feel threatened by ambiguous or unknown situations and have created beliefs and institutions that try to avoid these. The lower the score, the more comfortable people feel with ambiguity and unknown situation.
	Pragmatism (Numeric)	How every society has to maintain some links with its own past while dealing with the challenges of the present and future? The higher the score, the more likely they depend on past to prepare for the future.
	Indulgence (Numeric)	The extent to which people try to control their desires and impulses, based on the way they were raised. Relatively weak control is called "indulgence" and relatively strong control is called "restraint". The higher the score, the more indulgent a society is.
Outcome variable	Acceptable or not (Binary)	Whether the scenario as described would be acceptable or not acceptable.

Table A1. Variables used in the study.

Demographics	Overall	US	China	Australia	UK	India	Canada	Germany	Sweden
Language	100% English	100% Chinese	100% English	100% English	100% English	100% English	23.9% French, 76.1% English	100% German	100% Swedish
Mean age (years)/SD.	42	43	37	45	42	36	42	43	46
Gender (% of Male)	50.1	49.7	52.0	48.9	50.0	51.5	49.2	50.5	49.5
Hofstede's Cultural Dimensions									
Power Distance		40	80	36	35	77	39	35	31
Individualism		91	20	90	89	48	80	67	71
Masculinity		62	66	61	66	56	52	66	5
Uncertainty Avoidance		46	30	51	35	40	48	65	29
Pragmatism		26	87	21	51	51	36	83	53
Indulgence		68	24	71	69	26	68	40	78

Table A2. Participants' demographics, and cultural values of surveyed countries.

Algorithm	Feature	Acc. (%)	F
Naive Bayes	baseline	75.3558	0.742
	+country	74.9961	0.746
	+language	**75.2831**	**0.746**
	+Hofstede	72.2377	0.723
Logistic Regression	baseline	76.1524	0.726
	+country	76.8273	0.738
	+language	76.8290	0.739
	+Hofstede	**76.8307**	**0.739**
k-NN	baseline	**75.3277**	**0.742**
	+country	72.2390	0.720
	+language	72.2065	0.720
	+Hofstede	72.1320	0.719

Table A3. Accuracies and F-measures using different algorithm.