UCI-ISR-16-5, December 2016

Contextual Determinants for Users' Acceptance of Personal Data Processing: A Multinational Analysis

Pedro Leon[1], Alfred Kobsa[2], Carolyn Nguyen[3]
[1] Banco de México, pedro.leon@banxico.org.mx
[2] Institute for Software Research, UC Irvine, Kobsa@uci.edu
[3] Microsoft Corporation, Carolyn.Nguyen@microsoft.com

Abstract. The importance of "context" in people's privacy decisions is widely recognized, mostly in the area of inter-personal privacy. A comprehensive multinational analysis of what *users* consider to be the main contextual factors impacting their privacy decisions is still largely missing though, rendering it difficult to integrate context into data processing systems and privacy policy frameworks. We present a qualitative study in 4 countries followed by a large-scale (N=9,625) quantitative study in 8 countries aimed at identifying the contextual, attitudinal and demographic determinants that influence individuals' acceptance of scenarios involving the use of their personal data, and at gauging the relative influence of these determinants globally and country-wise. We develop parsimonious regression models to analyze the relative importance of different factors in different countries. The implications of such models in developing context and privacy aware systems and privacy policy frameworks are discussed.

1 Introduction

Over the past decade, numerous privacy scholars have emphasized the importance of "context" in studying people's privacy preferences and decisions (cf. [12, 15, 22, 23, 30]). They argued, e.g., that "privacy only exists in context" [12] and that "privacy should be conceptualized from the bottom up rather than the top down, from particular contexts rather than in the abstract" [30].

However, a comprehensive multinational bottom-up study to determine contextual factors that influence people's acceptance or non-acceptance of personal data processing is largely lacking to date. Such a study should particularly answer the following questions:

a) What do users see as the main context factors for their privacy decisions?
b) To what extent do these factors influence people's privacy decisions (and hence, which factors are more important or less important)?
c) Is the influence of context factors on privacy decisions the same across different countries (i.e., is it universal), or rather not?

The research described in this paper addresses these questions. We will first survey the existing literature on determinants of users' privacy decisions. Our overview shows that comprehensive multinational empirical analyses from the users' point of view are still largely missing, specifically in the area of information privacy. We then

present a qualitative study with 76 participants in four different countries on three continents, in which 19 potential determinants were identified (7 of them relating to the current "context"). We also describe a subsequent quantitative study with 9,625 participants in eight countries on four continents that aimed at validating and weighting these determinants. Logistic regression modeling was employed to determine a parsimonious set of the most influential determinants for participants' acceptance of scenarios involving the use of their personal data, both globally and country-wise. In the last section, we discuss the implications of our results on developing context-aware systems and privacy policy frameworks.

2 Related Work

In the area of *inter-personal* (or *"social"*) *privacy*, it has been recognized for long that privacy attitudes and behaviors are dependent not only on the type of personal information and its perceived sensitivity, but also on who the recipient is. In the words of Schoemann [28], "People have […] different relationships with different people. Information appropriate in the context of one relationship may not be appropriate in another". Altman defines privacy as "*selective* control of access to the self or to one's group" [3, emphasis ours].

Over the years, more and more determinants of inter-personal privacy have been identified mostly through user studies, such as the cultural background against which privacy decisions are made [2], the perceived usage purpose of personal information [1], the situation in which information is requested [19], and the form of access [20]. Increasingly, these determinants have become summarized under the notion of "context" of a privacy decision. In the words of Niessenbaum, "privacy only exists in context, meaning privacy is a relative, contextual concept," and going even further, "privacy is *defined* by its context" [23, emphasis ours].

The field of *information privacy* investigates users' privacy attitudes and behaviors in their interactions with organizations and websites. In contrast to inter-personal privacy, the notion of "context" still remains rather unspecified in this field. The concept typically relates to the presence of different types of service/website/organization in a user study [4, 14, 32]. Hine & Eve [13] identify situational factors that influence the impression of privacy infringement, such as the legitimacy of motives for information requests as well as their situational intrusiveness. John et al. [15] use the notion of context to refer to interface cues that sway users to disclose more information about themselves than they do without the presence of such cues.

This paper seeks to partially filling this gap. We present the results of a large-scale multi-national study that aimed to determine empirically

- what Internet users today see as the principal contextual determinants of the acceptability of personal data collection by organizations and websites, and
- what the relative influence of those determinants is and whether is the same in different countries.

Our approach thus follows Solove's recommendation that "privacy should be conceptualized from the bottom up rather than the top down" [30].

3 Qualitative Study

3.1 Method

In the first phase of the study, qualitative research was carried out in four countries: Canada, China, Germany and the US. These countries were chosen due to their different cultural norms and approaches to privacy regulation. In each country, 4 to 8 one-hour individual in-depth interviews were conducted, followed by 2 separate two-hour focus groups, each with 6 to 8 different individuals. In total, there were 26 interviews and 8 focus groups, with 76 participants. Participants ranged in age from 21 to 60, with an even mix of male and female, single and married. They included a variety of professions as well as students. Participants were recruited through established consumer panels maintained by the market research firm Ipsos in the respective countries, and were screened through preliminary surveys.

At the start of each session, to get participants thinking about issues related to personal data processing, they were asked about the online services they regularly access, the identities used for each service, the personal information provided, and the devices used for access. This was followed by a general discussion about their attitudes towards technology today and how this may evolve in the year 2030, to unveil their relationship with technology in general. Participants were then presented with different scenarios, to elicit their reasoning with regards to the different uses to which data about them may be put. The questions were intended to reveal their considerations on a number of issues, including the following: sensitivities to how the data was collected, the types of data collected, and uses of different types of data; harms that may arise from the use of data; motivations for managing their personal data or why they find certain data use more acceptable than others; and the role of trust in their decisions on whether the use of their data is acceptable.

For consistency, the same moderator and one of the authors were involved and present in all interviews and focus groups. The moderator conducted all the sessions that were in English, and worked closely with local moderators to ensure accurate translations when the sessions that were not.

While conducting the interviews and focus groups, we performed a grounded-theory analysis focused on collecting themes relating to participants' decisions on whether or not to agree to data processing. We used open coding, purposeful sampling, and constant comparison [11]. Below we describe some of the themes that emerged.

3.2 Results

Context Factors. We found that the contextual factors impacting user sensitivities to the access and use of data related to them can be grouped into 7 key categories – we define these as the *data context*.

1) *Type of personal data:* Banking data is most sensitive for all participants (who are therefore reluctant to give it out, unless it is considered a vital part of the service

offered). This is followed by government identification, health information and peer contact information.

2) *Type of entity that the user is interacting with:* Discussions included interaction with commercial service providers, employers, and government. Most participants do not want the government to access any private data. They are also concerned about unknown vendors accessing their personal data without their permission. Participants consistently have least trust in social network companies and most trust in banks.
3) *Trust in service provider*: We found that 3 elements impact users' trust levels: (i) reputation: brand familiarity, word of mouth recommendations, a personal relationship, company size, (ii) location: organizations with a local/national presence are more likely to abide regulations, and, (iii) free vs. paid services: trade-off between user risk and free services, understanding that free services generate income by selling user data.
4) *Collection method:* Participants prefer to give out personal data themselves rather than the service provider collecting or inferring data about them without their knowledge.
5) *Device used:* Participants differ in their views of which device is safest to access data from - no one device is universally deemed as being "safest", but the type of device is a relevant factor.
6) *Data usage purpose:* Participants do not want their data to be used without their knowledge nor by an unknown vendor. They also have negative reactions to automated uses of data as they feel it lacks flexibility and control for the user.
7) *Value exchange*: Participants are willing to exchange personal data for an immediate personal benefit that reflects the value of their data. The most appealing benefits include discounts, better service / improved product, and convenience / time savings. A benefit to the social good is not as motivating as a personal benefit, but can be a factor.

These finding corroborated the hypothesis that context is relevant to personal data processing, and that what is considered acceptable data use is a nuanced, personal decision, influenced by social norms and other cultural factors. Like Martin [20], we found that "all respondents held different privacy norms across hypothetical contexts, thereby suggesting privacy norms are contextually understood within a particular community of individuals." A participant from China stated this elegantly as *"I have a lot of concerns. I have a safety boundary that I cannot talk about [articulate]. A person is not a machine. They have complicated emotions."*

Attitudinal Factors. Besides the above-mentioned context factors, we also found that the following privacy-related attitudes of participants seem to play a role in their agreement with the processing of their personal data:

- Level of concern about how companies use personal information,
- Effort taken to protect personal data,
- Claimed knowledge and awareness of company practices,
- Perceptions of government regulation to protect privacy,

- Trust in companies overseen by a third-party (self-regulation model) to protect personal information,
- Trust in government to protect personal information.

Attitudinal Factors. We also decided to include a number of demographic factors into our subsequent quantitative studies. While they were not specifically discussed during our qualitative study, research has shown that they do exhibit some effect on privacy decisions and behaviors. These added variables are:

- *Age* (see e.g. [10, 21, 24])
- *Gender* (see e.g. [8, 26])
- *Education* (see e.g. [25, 26])
- *Have kids younger than 18* (see e.g. [9])
- *Technology Adoption* (see e.g. [34, 36]).

4 Quantitative Study

We built upon the qualitative results to develop a survey for our quantitative research. Its primary objectives are

- to validate and quantify the data context variables,
- to determine the relative importance of the elements within each of the data context variables,
- to understand the relative impact of the context variables and how it varies across individuals and cultural boundaries, and
- to develop insights about how other factors, namely attitudinal and demographics variables, may affect the findings.

We conducted 9,625 online surveys uniformly distributed across eight countries: the original four, plus the United Kingdom, Sweden, Australia and India. Again, the countries were chosen to be representative of the existing privacy / data protection landscape and social/cultural attitudes.

4.1 Subject Recruitment and Screening

In each country, respondents were screened and selected from an existing online panel maintained by Ipsos. The panel has over 4.5 million online participants in 46 countries, and complies with the quality management standards ISO 9001, 20252 and 27001. Ipsos uses various quotas, such as age, gender and ethnicity, and "nests" them together (i.e., use of age, gender, and ethnicity together, instead of age, gender, and ethnicity quotas separately), to ensure comparable profile to the population at large for each of the countries represented by the panel.

The respondents ranged in age from 18 to 65, with an equal mix of genders and occupation (see Table 1 for a more detailed breakdown). The 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 1. Participants' demographics (N = about 1200 per country.

	Overall	US	China	Australia	UK	India	Canada	Germany	Sweden
Mean age	42	43	37	45	42	36	42	43	46
% of male	50.1	49.7	52	48.9	50	51.5	49.2	50.5	49.5

4.2 Survey Design

The survey took 20-25 minutes to complete and was conducted online. In non-English speaking countries, the surveys were translated into the local language. 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 survey were explained through mouse-over text.

A core part of the survey was a data context conjoint exercise, in which respondents were asked to view 6 different screens, each containing 4 different randomized scenarios regarding the use of their personal data. They were then asked to indicate whether the scenario as described is 'acceptable' or 'not acceptable' to them.

A scenario constitutes one value assignment for each of the 7 context variables identified in the qualitative research (see Section 3.2). The possible value assignments for the context variables Collection Method, Trust in Service Provider, Usage Application and Value Exchange are shown in Table A1 in the Appendix. Examples of the tested scenarios include:

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

A full combinatorial design would have resulted in too many different possible scenarios and would have made the surveys unreasonably long. We reduced the number of possible scenarios by removing combinations of elements that do not make sense (e.g., when the Collection Method is *passive*, the Usage Application cannot be *as I agreed to*), and by restricting the number of choices for the variables. This resulted in a total of 192 scenarios that were tested. In the analysis below, we focused solely on the results for the scenarios that involved interactions with service providers, and not on those that involved interactions with either employers or government agencies.

The remainder of the survey included a number of other potential factors that may impact users' sensitivity towards use of data related to them, including demographics, technology adoption, and attitudes on a number of the topics uncovered in the qualitative research. Refer to Table A1 in Appendix A for a listing of some of these factors.

4.3 Analysis

We built Generalized Linear Models with Random Effects (GLMRE) and used the Bayesian Information Criterion (BIC) and deviance metric to select the most parsimonious models that would fit our data sets. We used the lme4 package for the R statistics language. Since our outcome variable was binary, we specified a binomial family for the regression models. We built one global model considering data from all countries, as well as individual models for each country whose fit turned out to be considerably better.

We built additive models, adding one variable at a time and comparing the fit of consecutive models. We continued adding independent variables until the values for the BIC and deviance stopped improving. We first selected contextual variables, followed by demographic and finally attitudinal variables.

Since each participant was asked to assess the acceptability of 24 scenarios in a within-subjects design, each of the models included a subject random effect. Due to the lack of independence between *types of personal data* and *collection methods*, *collection methods* and *devices*, and *collection entity* and *trust*, our models could not include all 7 contextual variables simultaneously. In addition, in preliminary analysis we found that the trust contextual variable was a significant predictor of acceptability; however, as explained above, the trust contextual variable only existed for the service provider entity, but not for Government or Employer. Furthermore, 152 out of 192 scenarios had service provider as the collection entity. Therefore, we focused our analysis on the subset of scenarios that had a service provider as the collection entity. This represented 73% of the total 231,000 assessed scenarios (24 scenarios for each of 9,625 respondents).

Table A1 in the Appendix shows all variables that were considered during the model selection process. The general structure of the regression models is as follows:

$$Y_{nk} = B_0 + \sum B_j V_{jk} + \sum C_j D_{jn} + U_n + e_k$$

where,

Y_{nk} = Response of participant n to scenario k
B_0 = Intercept
B_j = Regression coefficient of the contextual factor V_j in scenario k.
V_{jk} = Contextual factor j in scenario k.
C_j = Regression coefficient of the demographic/attitudinal variable D_j.
D_{jn} = Demographic/attitudinal variable j of respondent n.
U_n = Respondent-level residual (a.k.a. subject random effect)
e_k = Scenario-level residual (a.k.a. residual error)

4.4 Overall acceptability

We first discuss descriptive statistics showing the acceptability of some of the relevant factors. From our qualitative study, we had reason to believe that contextual factors would have an effect on acceptability. As an example, Figure 1 shows the fraction of respondents who accepted a given scenario via different collection methods in each country. We observe that scenarios in which the collection method was *active* (users divulge personal information themselves) were the most readily accepted, followed by those scenarios where collection method was *as by-product* of the user activities or *inferred* based on analytics of data collected about the user. This observation applies to all countries, but the exact numbers per country are somewhat different.

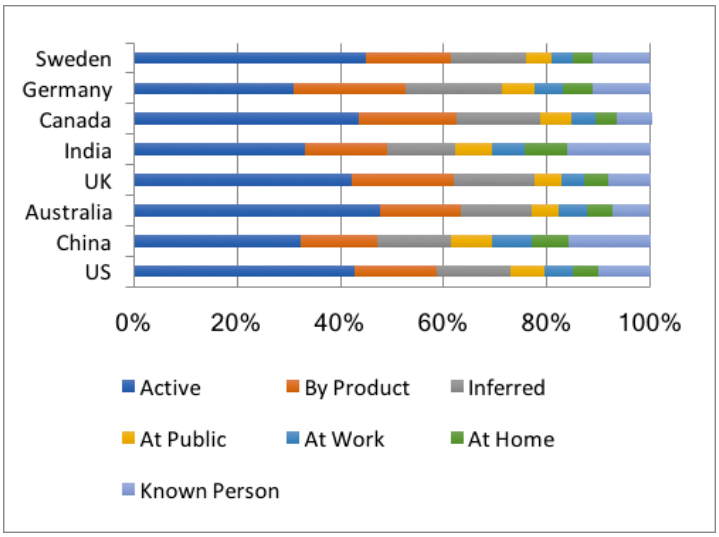

Figure 1. Relative Effect of Collection Methods Per Country.

In agreement with our qualitative study, we also found that some of the non-contextual variables influenced the acceptability of scenarios. An example is the level of technology adoption. We gauged technology adoption by asking respondents a number of questions related to the use of and attitudes towards technology. We then bucketed respondents in two groups, technology adopters and non-adopters. Figure 1 shows that tech-adopters responded more positively to scenarios across all countries. We further suspected that acceptability might vary across countries due to cultural and demographic aspects, as well as due to the different regulatory frameworks for privacy in those countries. Figure 2 shows that the overall acceptability of scenarios for each of the eight countries given the level of technology adoption of the respondents is different per country. We can also observe that respondents from China and India were in general more willing to accept usage scenarios.

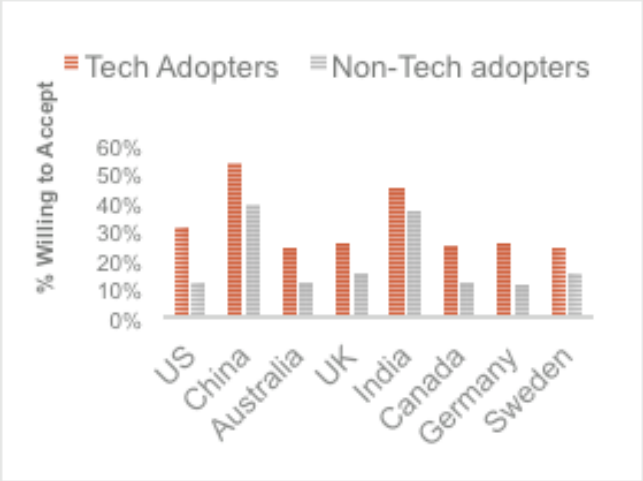

Fig. 2. The impact of tech adoption on acceptability of data use scenarios. In all countries, tech-adopters were more willing to accept data uses than non tech-adopters.

4.5 Regression Models

As mentioned before, some of the contextual factors were not independent from each other. In particular, a subset of *data types* only existed for a given *collection method*, and a subset of *collection methods* only existed for some *devices*. As a result, these contextual factors could not coexist in a given model. Table A2 in the Appendix shows the regression coefficients of all variables in our final models. We found that both the global and per-country models that included the context variables *Collection Method, Trust, Use* and *Value Exchange* explained the most variance of the data (i.e., had the lowest BIC and deviance).

While we suspected that a global model (i.e., a model that considers all countries together) might not have a good fit given the number of differences in demographics and attitudes across all countries, we did build such a model and compared its fit with the fit of each country-specific model. In particular, we compared two models for each country, namely the one with the best fit after following the process described above using that country data subset, and the one that had all the variables that fitted a global model best. We found that the residual errors for each of the models built on a country-level basis were smaller compared with the corresponding model when the globally found significant predictors were used. In addition, the country-level models were more parsimonious than the global ones. We concluded that country-level models were more appropriate to predict respondents' acceptance of scenarios. They also allow for a comparison of the relative influence of the various factors on the acceptability of scenarios in each individual country. Therefore, we focus our discussion on the results of each per-country model.

4.6 Impact of Context

We found that *Collection Method, Trust, Use,* and *Value Exchange* context variables significantly impacted participants' opinions about the acceptability of the different scenarios they saw (see Table A2). While the general direction of the impact of these factors was similar for all the countries, their relative importance was different in some cases. We now discuss the particular impact of contextual variables on participants of each country.

Collection Method. Collection method was the contextual variable with the highest relative importance in all countries, except for Sweden (there, usage purpose had the largest impact, followed by collection method).

Unsurprisingly, with the exception of Germany, active collection was preferred over scenarios where a third-party (i.e., known person) would disclose the respondent's personal information. Active collection and passive collection methods where the information was collected as a by-product of the user's activities or inferred based on analytics of data collected about the user (*Passive Collection: By Product of Activity* and *Passive Collection: Independent of User Activity* in Table A2) had a similar impact on respondents' acceptability of scenarios.

While other passive collection methods were also preferred over a third-party disclosure, these methods had a much milder effect on acceptability (i.e., the regression coefficients were smaller). This suggests that participants do not welcome silent, passive collection that is not associated with present or past user actions. The coefficients of the passive collection factors, in turn, reveal participants' sensitivity to silent data collection at different locations. In particular, with the exception of India, regression coefficients are smaller for passive collection at home compared with passive collection at public locations or at work.

Participants from Germany had a slightly different reaction to collection method. They preferred passive collection methods (where the information was collected as by-product of the user activities or inferred based on analytics of data aggregated about the user) over active collection. We suspect that this was caused by the type of information associated with the collection method. As discussed before, collection methods were matched with specific types of data. In the particular case of the active collection method, the associated data type was *location and contact information*, whereas for the two passive collection methods in discussion, the associated data type was *location* only, suggesting that for participants from Germany, contact information might have had a stronger effect than the active collection method.

Trust in Service Provider. Different levels of trust impacted respondents' stated acceptance of scenarios. We investigated seven levels of trust. With the exception of China and Canada, a *company with an existing relationship* and a *well-known company* had the same high positive effect in comparison to an *unfamiliar company* (i.e., very similar regression coefficients). Similarly, the regression coefficients of *company located outside of country* and *unfamiliar company* are also similar and the smallest when compared with other levels of trust.

Interestingly, while small in magnitude, the coefficients for a *company providing paid services* are consistently higher than those for a *company providing free services*, in all countries but China. This suggests that in those countries, additional trust might be gained when users pay for a service than when a company provides free service. In the case of China though, the coefficient for a *company providing free services* (0.6) is bigger than the coefficients for a *company inside the country* (0.4), a *company providing paid services* (0.5), a *company outside the country* (0.3), and as big as the coefficient for a *company with an existing relationship* (0.6), and just smaller than the coefficient for a *well-known company* (0.8).

Value Exchange. Value Exchange had a milder impact than Trust in Service Provider, in all countries except China. In general, trust coefficients were almost twice as large as the value exchange coefficients, suggesting that for those countries, trust is more important than value exchange.

In addition, in all western countries (US, UK, Canada, Germany and Sweden) and in Australia, the coefficients for *saving time and money* (0.4 – 0.5) and *something of unique value* (0.4 – 0.6) were consistently higher than those for *benefiting the community* (0.2 – 0.3). In China and India, in contrast, the coefficients for all value exchange levels were similar within each country, but much larger for China (0.9-1.1) than for India (0.3-0.4).

Usage Purpose. Usage purpose was the contextual variable with the second largest effect in all countries except Sweden and China. In the case of Sweden, usage purpose had the largest impact followed by collection method. For Chinese participants, value exchange had a much larger impact (coefficients in the range from 0.9 to 1.1) than usage purpose with relatively smaller coefficients for this contextual variable compared with other countries.

Usage *to customize choices* was preferred in all countries over usage *to make autonomous decisions on behalf of users*. The impact of *customize choices* in China (coefficient 0.1) and India (coefficient 0.4) was smaller than for the other countries where coefficients for *customize choices* ranged from 0.7 (Germany) to 1.2 (Sweden).

Unsurprisingly, participants were more likely to accept scenarios for which they were told that the information collected was going to be used *as agreed*. The coefficients for *as agreed* were the lowest in India and China (both 0.6), while in the other countries they ranged from 1.7 (Germany) to 2.7 (Sweden).

4.7 Impact of Participant's Demographics

The demographics variable that had the largest impact on acceptability was technology adoption. It had a positive effect in seven countries, but not Sweden.

Surprisingly, although other research [25] had suggested that education is a good predictor for privacy attitudes, it was not a statistically significant determinant in any of the countries polled in our study. The reason might be the positive correlation of tech adoption with education. We verified this suspicion by building models with only the education variable but not tech adoption. The fits of these models were signifi-

cantly improved; nevertheless, when using the tech adoption variable, the models fits were even better (i.e., they explained more variance in the data).

In all countries except for the US and the UK, age was a significant predictor. It reduced the odds of the probability of accepting a given personal-data processing scenario by 0.98 (exp -0.02) for every additional year. This result is consistent with previous findings that age is a significant negative predictor of privacy attitudes in Germany [31] but that the results for the U.S. are mixed [10].

Interestingly, we found that being a parent of kids younger than 18 positively impacted the acceptability of scenarios in the US. We suspect that this can be a result of parents witnessing kids' adoption of new technologies; which might make parents more comfortable with participating in online exchanges of information.

4.8 Impact of Participants' Attitudes

Overall, respondents' trust that the government and companies certified by a third party (i.e., self-regulation model) will protect their personal information had a significant positive impact on the acceptability of scenarios in all countries. Its effect was strongest in the US and the weakest in Sweden and Australia.

Similarly, positive perception of government regulation (i.e. favorable views on its adequacy and enforcement), positively impacted the acceptability of personal data processing in all countries, except for Canada. Its largest effect was on US participants with a regression coefficient of 1.4, and its smallest effect in Australia with a coefficient of 0.4. In China and India, the effects were 0.7 and 0.9, respectively.

Participants' stated concerns about service providers' uses of their personal information also considerably impacted the acceptability of scenarios in all countries but Canada. For example, for a US respondent who was concerned about service providers' data uses, the odds of the probability of accepting a given scenario would be reduced by half (exp -0.7).

In general, we found that while the contextual variables had the largest effect on acceptability, demographic and attitudinal factors also played an important role. In addition, both attitudinal and demographic factors have different effects in different countries, and some of these may not even be significant for some countries while being significant for others.

5 Discussion

5.1 Implications for system designs that respects users' privacy preferences

Our study gives three important insights into determinants for people's acceptance of the processing of their personal data, which have considerable impacts on the design of privacy-sensitive interactive systems, specifically when they are to be used by an international audience:

First, our results show that commonalities exist in the investigated countries between the determinants of users' acceptance of personal-data processing. However, those commonalities are too small to allow an analysis of the determinants at a global rather than a country-specific level. This is bad news for the design of privacy-aware interactive systems since far higher efforts will be needed for gathering sufficient country-specific data on the population's privacy preferences.

As a "compromise", one may wonder whether models for world regions like "Asian countries" or "countries with Anglo-Saxon tradition" would be amenable to joint "umbrella" models. While we did not survey sufficiently many countries to be able to offer a definitive opinion, it seems that the differences between the European countries, between the U.S. and Canada, and between India and China are too large to make this endeavor promising. Countries will likely need to be surveyed individually for international coverage.

Finally, our results allow researchers to prioritize their research efforts if they should plan to survey privacy determinants in countries that we did not cover. We identified 14 factors via the qualitative study and five more demographic factors based on the literature. Two of them turned out to be very important (namely the context variables "collection method" and "usage purpose") and should therefore be studied with highest priority. Next come the remaining context factors "trust in service provider" and "value exchange". Those four variables might even suffice for a "quick-and-dirty" gauge of the acceptability of a personal data-processing scenario in the respective country. However, our BIC and deviance metric also showed that many of the other factors do play a minor role, so that those will probably have to also be measured to obtain a better-quality prediction for additional countries.

5.2 Implications for privacy policy

One of the first questions we investigated in our qualitative study regarded people's relationship with technology in general. Overall, people recognize that technology enables them to do more, but it also makes them feel powerless since they are increasingly reliant on something they don't fully understand. The quote "*When it works, it works great. But everything is now dependent on technology*" (Canada) typifies this sentiment. Users' ambivalence between feelings of empowerment and helplessness is a well-known observation. Given this ambivalence, the desire for control and transparency is natural (and independent of the type of technology involved or how it is used). This ambivalence is often overlooked by technology companies and policymakers, and the expressed desire for control often misinterpreted as completely rational or absolute. However, this is not necessarily the case. People often don't want to understand or control everything, but only to be reassured that the technology is capable of working for them, on their behalf, and does not pose any risks or harms – and if the technology does malfunction, that there are appropriate enforcement and user control points to rectify the situation. This is different from wanting control at every point in the system or perfect information on how the system works. Current policy approaches are too often based on a literal translation of this perceived desire, leading to regulations or implementations that are ineffective and cumbersome.

The research findings and resulting models suggest that a simplistic, binary approach to data use policies that applies universally or country-wide is neither appropriate nor flexible enough to accommodate the nuances of what is considered acceptable or contextually appropriate data use to each individual. Defining these nuances in regulations will be very difficult, specifically since relying on users' explicit consent to personal data processing will be less and less meaningful in the future [16, 5, 6, 29, 35]. However, the models derived suggest that it may be possible to develop context-aware systems that can predict what may be acceptable and appropriate data uses (cf. [17] for independent corroborating evidence at the individual-user level). Technology companies and policymakers should consider these and similar technologies in developing data governance policy frameworks that are more nuanced and context-aware.

The 7 context variables can be divided into *objective* factors (that are the same for all individuals participating in the same transactions, namely the Type of Data, Type of Entity, Collection Method, Data Usage Purpose, and Device Context), and *subjective* factors (that are different for each individual participating in the transaction, namely Trust and Value Exchange). The subjective variables, along with the attitudinal factors and demographics, are likely to evolve over time, as a result of changes in personal preferences, and natural evolution in social and cultural norms. This should challenge technology companies and policymakers into considering how to develop policy frameworks that are flexible enough to accommodate these changes over time. One way to achieve this is to consider policies that are principle- and outcome-driven, rather than process- or technology-driven. For example, a principle can be that the use of data should be contextually consistent with user preferences [22], instead of specifying the process for how users would state their preferences.

The context variables can also be leveraged to develop a more nuanced data policy framework and resultant systems. From the analytics, the 2 variables with the highest impact are Collection Method and Data Usage. The highest acceptability rates are found in scenarios with *active data collection* and data usage is *as users agreed*. This is the equivalent of the notice/consent model of today. However, as passive data collection becomes more prevalent in the future, the remaining context variables can be leveraged to increase acceptability of these scenarios. For example, principles that can enable trustworthy data practices become more important as part of an overall data policy framework. This is a more foresightful approach to data policy than trying to impose the existing approach of notice/consent on a world that will be vastly different than the existing world [6, 29, 35].

The analytical results also point out that although the notion of context exists in all the countries studied, models needed to be considered on a country-by-country basis, and not at a global scale. This poses problems for the global nature of data flow, and challenges policy makers to coordinate globally to enable this data flow while respecting the preferences of individuals in each country. However, the evidence that context exists in all the countries point to a possible approach where perhaps a global framework can be established, with allowances for local preferences.

It is important to understand how users think about the processing of their personal data, and what contextual, attitudinal and demographic determinants drive their acceptance or non-acceptance of personal data processing. As we move into a world with increasingly ubiquitous personal data, users' opinions on passive data collection

become particularly important. Moreover, since companies can now easily serve users in many different countries, a multi-national understanding is required.

Our findings are "inconvenient" in some sense. Country differences turned out to be larger than expected, so that a single universal model to describe the size of the influence of the different determinants on users' acceptance of personal-data processing turned out to be too inexact. The feasibility of umbrella models for "world regions" is also put into question based on the data of the countries that we sampled. Judging by our BIC and deviance metric, *many* contextual, attitudinal and demographic factors will need to be taken into account for accurate predictions (about 15 variables or 25 variable values). However, we also found that the number of very important factors is quite small, and *all of them are context factors*.

Our study obviously needs to be extended to additional countries beyond the eight that we selected. Moreover, our study is only a snapshot in time. Technology and people's perception of privacy changes, and it is therefore advisable to repeat this study in regular intervals.

Our results have both implications on the design of privacy-aware systems and on privacy policy. In the light of upcoming developments in personal data processing, this study points out the need for an interdisciplinary dialogue that include technologists, social scientists, economists, and policymakers, if a privacy policy framework is to be effective and implementable.

Acknowledgments. We thank Paul Evans and Jessica Leask from Ipsos who collaborated closely with us on the qualitative and quantitative research, and William Hoffman for valuable comments on earlier versions of this paper.

References

[1] Adams, A. and Sasse, M.A. 2001. Privacy in multimedia communications: Protecting users, not just data. *People and computers XV: interactions without frontiers: Joint proceedings of HCI 2001 and IHM 2001* (2001), 49.

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

[3] Altman, I. 1976. Privacy: A Conceptual Analysis. *Environment and Behavior*. 8, 1 (Mar. 1976), 7–29.

[4] Bansal, G., Zahedi, F. and Gefen, D. 2008. The Moderating Influence of Privacy Concern on the Efficacy of Privacy Assurance Mechanisms for Building Trust: A Multiple-Context Investigation. *Proceedings of the 29th International Conference on Information Systems* (Paris, France, Dec. 2008), 1528–1546.

[5] Barocas, S. and Nissenbaum, H. 2009. On notice: The trouble with Notice and Consent. *Proceedings of the Engaging Data Forum: The First International Forum on the Application and Management of Personal Electronic Information* (2009).

[6] Cate, F.H. and Mayer-Schönberger, V. 2013. Notice and consent in a world of Big Data. *International Data Privacy Law*. 3, 2 (2013), 67–73.

[7] Cate, F.H. and Mayer-Schönberger, V. 2012. *Notice and Consent in a World of Big Data*.

[8] Caverlee, J. and Webb, S. 2008. A Large-Scale Study of MySpace: Observations and Implications for Online Social Networks. *ICWSM* (2008).

[9] Cranor, L.F., Reagle, J. and Ackerman, M.S. 1999. *Beyond Concern: Understanding Net Users' Attitudes About Online Privacy*. Technical Report #TR 99.4.3. AT&T Labs - Research.
[10] Dommeyer, C.J. and Gross, B.L. 2003. What Consumers Know and What They Do: An Investigation of Consumer Knowledge, Awareness, and Use of Privacy Protection Strategies. *Journal of Interactive Marketing*. 17, 2 (2003), 34–51.
[11] Glaser, B.G. 1998. *Doing grounded theory: issues and discussions*. Sociology Press.
[12] Gutwirth, S. 2002. *Privacy and the information age*. Rowman & Littlefield Publishers.
[13] Hine, C. and Eve, J. 1998. Privacy in the Marketplace. *The Information Society*. 14, 4 (1998), 253–262.
[14] Hsu, C. (Julia) 2006. Privacy concerns, privacy practices and web site categories: Toward a situational paradigm. *Online Information Review*. 30, 5 (Sep. 2006), 569–586.
[15] John, L.K., Acquisti, A. and Loewenstein, G. 2011. Strangers on a Plane: Context-Dependent Willingness to Divulge Sensitive Information. *The Journal of Consumer Research*. 37, 5 (Feb. 2011), 858–873.
[16] Kamp, J. and Connelly, K. 2007. Beyond Consent: Privacy in Ubicomp. *Digital Privacy: Theory, Technologies, and Practices*. A. Acquisti, S. Gritzalis, C. Lambrinoudakis, and S. De Capitani di Vimercati, eds. Auerbach Publications. 327–366.
[17] Knijnenburg, B.P., Kobsa, A. and Jin, H. 2013. Dimensionality of information disclosure behavior. *International Journal of Human-Computer Studies*. 71, 12 (Dec. 2013), 1144–1162.
[18] Kuner, C., Cate, F.H., Millard, C. and Svantesson, D.J.B. 2012. The challenge of "big data" for data protection. *International Data Privacy Law*. 2, 2 (May 2012), 47–49.
[19] Lederer, S., Mankoff, J. and Dey, A.K. 2003. Who wants to know what when? privacy preference determinants in ubiquitous computing. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Ft. Lauderdale, FL, Apr. 2003), 724–725.
[20] Martin, K.E. 2012. Diminished or Just Different? A Factorial Vignette Study of Privacy as a Social Contract. *Journal of Business Ethics*. 111, 4 (Dec. 2012), 519–539.
[21] Milne, G.R. and Boza, M.-E. 1999. Trust and Concern in Consumers' Perceptions of Marketing Information Management Practices. *Journal of Interactive Marketing*. 13, 1 (1999), 5–24.
[22] Nissenbaum, H. 2004. Privacy as Contextual Integrity. *Washington Law Review*. 79, (2004), 119–157.
[23] Nissenbaum, H.F. 2009. *Privacy in context : technology, policy, and the integrity of social life*. Stanford Law Books.
[24] Page, X., Knijnenburg, B.P. and Kobsa, A. 2013. What a Tangled Web We Weave: Lying in Location-Sharing Social Media. *Proc. CSCW 2013* (2013).
[25] Phelps, J., Nowak, G. and Ferrell, E. 2000. Privacy Concerns and Consumer Willingness to Provide Personal Information. *Journal of Public Policy & Marketing*. 19, 1 (2000), 27–41.
[26] Riquelme, I.P. and Román, S. Is the influence of privacy and security on online trust the same for all type of consumers? *Electronic Markets*. 1–15.
[27] Rubinstein, I.S. 2013. Big Data: The End of Privacy or a New Beginning? *International Data Privacy Law*. 3, 2 (May 2013), 74–87.
[28] Schoeman, F. 1984. Privacy and intimate information. *Philosophical Dimensions of Privacy: An Anthology*. F.D. Schoeman, ed. Cambridge University Press. 403–418.
[29] Solove, D.J. 2013. Privacy Self-Management and the Consent Dilemma. *Harvard Law Review*. 126, (2013), 1880–1903.
[30] Solove, D.J. 2008. *Understanding Privacy*. Harvard University Press.

[31] Stone, E.F., Gueutal, H.G., Gardner, D.G. and McClure, S. 1983. A field experiment comparing information-privacy values, beliefs, and attitudes across several types of organizations. *Journal of Applied Psychology*. 68, 3 (1983), 459–468.

[32] Telefónica Goes a Little Bit "Big Brother:" 2012. *http://on.wsj.com/RaEa6u*. Accessed: 2013-06-03.

[33] Xu, H. and Teo, H.-H. 2004. Alleviating Consumers' Privacy Concerns in Location-Based Services: A Psychological Control Perspective. *ICIS 2004 Proceedings*. (Dec. 2004).

[34] Zanfir, G. 2014. Forgetting About Consent. Why The Focus Should Be On "Suitable Safeguards" in Data Protection Law. *Reloading Data Protection*. S. Gutwirth, R. Leenes, and P.D. Hert, eds. Springer Netherlands. 237–257.

[35] Zhao, L., Lu, Y. and Gupta, S. 2012. Disclosure Intention of Location-Related Information in Location-Based Social Network Services. *International Journal of Electronic Commerce*. 16, 4 (Jul. 2012), 53–9

Appendix

Table A1. Variables considered in model building process. See Table A2 for final models.

	Variable	Values
Contextual	Collection method	(1) Active: user explicitly provides the information; (2) Passive: a person the user knows provides the information; (3) Passive: Service provider (SP) silently collects the information while the user is at home; (4) Passive: SP silently collects the information while the user is at work; (5) Passive: SP silently collects the information while the user is in a public location; (6) Passive: SP silently collects the information as part of an active transaction; (7) Passive: SP infers the information based on analytics of data aggregated about the user (i.e., independent of user current activity).
	Trust in service provider	(1) Well known; (2) Unfamiliar; (3) Existing relationship; (4) Locations in my country; (5) Location outside of my country; (6) Provides free services; (7) Provides paid services
	Usage purpose	(1) As I agreed; (2) To customize the choices offered to me; (3) to automatically make decisions for me.
	Value exchange	(1) Save time or money; (2) Something of unique value; (3) Benefit to the community; (4) No other benefit.
Attitudes	Concern about how companies use PI	(1) Concerned; (2) Unconcerned
	Effort taken to protect PI	(1) Take measure to protect privacy; (2) Don't take measures to protect privacy
	Claimed knowledge of company practice	(1) Knowledgeable; (2) Not knowledgeable
	Perceptions of gov. regulation to protect privacy	(1) Good regulations exists and are enforced; (2) Good regulations don't exist or are not enforced
	Trust in self-regulation to protect PI	(1) Trust in self-regulation; (2) No trust in self-regulation
	Trust in government to protect PI	(1) Trust in Government to protect personal information; (2) Not trust in Government to protect personal information
Demographics	Age	<numeric>
	Gender	(1) Male; (2) Female
	Education	(1) Less than High School; (2) Some High School; (3) High School Degree; (4) College Degree; (5) Graduated Degree
	Kids < 18?	(1) Have; (2) Don't have
	Tech adoption	(1) Adopter; (2) Non-adopter

Table A2. Regression coefficients for each country. A blank cell indicates that the factor was not included in this model since it did not significantly improve its fit.

	Independent Variable	Reference Category	Australia	Canada	UK	US	Germany	Sweden	China	India
Contextual	Active collection	Passive: A known person provides the information	2.5	2.2	2.2	2.2	1.0	1.7	1.8	1.5
	Passive coll.: by-product of activity		2.1	2.2	2.5	2.1	2.1	1.7	1.6	1.3
	Passive collection: independent of user activity		2.1	2.4	2.3	2.1	2.1	1.7	1.7	1.2
	Passive: collection at public loc.		0.9	1.1	0.9	1.0	0.7	0.4	1.0	0.4
	Passive: collection at work		0.9	1.0	0.7	0.7	0.7	0.3	1.0	0.6
	Passive collection: at home		0.7	0.6	0.6	0.4	0.6		0.5	0.8
	Trust: existing relationship	Trust: unfamiliar	0.9	1.1	0.9	0.7	1.0	0.7	0.6	0.6
	Trust: well known		0.9	0.9	0.9	0.7	0.9	0.7	0.8	0.6
	Trust: provides paid services		0.9	0.9	0.8	0.6	0.8	0.6	0.5	0.5
	Trust: provides free services		0.7	0.8	0.7	0.5	0.6	0.5	0.6	0.4
	Trust: location in my country		0.5	0.7	0.5	0.4	0.6	0.4	0.4	0.3
	Trust: location outside my country		0.1	0.2	0.1	0.1	0.3	0.2	0.3	0.1
	Usage purpose: as previously agreed	Usage purpose: to make autonomous decisions	1.9	2.0	2.0	2.0	1.7	2.7	0.6	0.6
	Usage purpose: customize choices		0.8	0.9	0.8	0.9	0.7	1.2	0.1	0.4
	Value exchange: save time and money	Value exchange: no additional benefit	0.5	0.4	0.6	0.4	0.4	0.5	1.1	0.3
	Value exchange: get something of unique value		0.4	0.3	0.5	0.4	0.5	0.6	1.0	0.4
	Value exchange: benefits community		0.2	0.2	0.2	0.1	0.3	0.3	0.9	0.3
Attitudes	Government regulation: good	Gov. regulation not good	0.4		0.5	1.4	0.6	0.3	0.7	0.9
	Trust in self-regulation: yes	Trust in self-reg.: no	0.6	0.9	0.7	1.0		0.4	0.7	0.5
	Trust in government to protect PI: Yes	Trust in government to protect PI: no	0.3		0.5	0.7	0.6	0.4	0.4	
	Concern about data use: concerned	Data use: unconcerned	-0.4		-0.6	-0.7	-0.5	-0.6	-0.6	-0.4
Demo-graphics	Tech adopter: yes	Tech adopter: no	0.2	0.3	0.4	0.6	0.8		0.4	0.3
	Parenthood: yes	Parenthood: no				0.3				
	Age	<numeric>	-.02	-.03			-.02	-.02	-.02	-.02