Let's Do It at My Place Instead? Attitudinal and Behavioral Study of Privacy in Client-Side Personalization

Alfred Kobsa, Bart P. Knijnenburg
Department of Informatics
University of California, Irvine
{kobsa, bart.k}@uci.edu

Benjamin Livshits
Microsoft Research
Redmond, WA
livshits@microsoft.com

ABSTRACT
Many users welcome personalized services, but are reluctant to provide the information about themselves that personalization requires. Performing personalization exclusively at the client side (e.g., on one's smartphone) may conceptually increase privacy, because no data is sent to a remote provider. But does client-side personalization (CSP) also increase users' *perception* of privacy?

We developed a causal model of privacy attitudes and behavior in personalization, and validated it in an experiment that contrasted CSP with personalization at three remote providers: Amazon, a fictitious company, and the "Cloud". Participants gave roughly the same amount of personal data and tracking permissions in all four conditions. A structural equation modeling analysis reveals the reasons: CSP raises the fewest privacy concerns, but does not lead in terms of perceived protection nor in resulting self-anticipated satisfaction and thus privacy-related behavior. Encouragingly, we found that adding certain security features to CSP is likely to raise its perceived protection significantly. Our model predicts that CSP will then also sharply improve on all other privacy measures.

Author Keywords
Privacy; personalization; client-side; structural equation modeling (SEM); attitudes; behaviors.

ACM Classification Keywords
H.5.2. Information interfaces and presentation: User interfaces–evaluation/methodology, theory and methods, K.4.1. Computers and society: Public policy issues–privacy.

General Terms
Human Factors; Design; Measurement; Verification.

INTRODUCTION
Personalized services are widely used. For instance, 20–30% of Amazon purchases and 60% of Netflix views are a result of personalized recommendations [42]. At the same time though, users are reluctant to disclose personal data or allow their system usage to be tracked, which is a requisite for personalization [7,48]. For instance, people often loathe location tracking which enables location-based services [1], recommendations and advertisements [51], or they are unwilling to disclose their music preferences and personality traits to a music recommender [14]. The notions of personalization-privacy "paradox"[1] [3,51], "trade-off" [29] and "dilemma" [7] have been used to refer to these seemingly conflicting user desires for both personalization and privacy.

Several proposals have been made to allow users to enjoy both a reasonable personalization quality and a higher degree of privacy (see [28,49] for overviews). One technical solution that became popular recently abandons the assumption that personal data collected on users' local devices must be sent to a remote site for personalization to take place. In the paradigm of "client-side personalization" (CSP) [6,18,35,37], users' personal data remains instead on the user's device, where all personalization is carried out. Various forms of CSP have been explored so far, which we will survey in the next section.

From a conceptual and technical point of view, preventing others from accessing personal data enhances the privacy of the data subjects [46]. It has, however, been argued that CSP will also increase users' *perception* of privacy. For instance, the following claims can be found in the literature:

- "the user does not need to worry about [...] privacy infringement" [43],
- "it is easy for a consumer to understand that their personal information will stay under their control at all times" [35],
- "the user may also find it desirable when they prefer that their user model be kept only on their phone and under their own control" [15],
- "client-side solutions […] instill a greater sense of user trust" [35].

[1] The notion of "privacy-personalization paradox" is different from the "privacy paradox" [38], a conflict between stated privacy attitudes and/or intended behavior and actual behavior observed later on (see e.g. [47]).

CHI 2014, April 26 - May 01 2014, Toronto, ON, Canada

ACM 978-1-4503-2473-1/14/04…$15.00.
http://dx.doi.org/10.1145/2556288.2557102

Moreover, confidence-inducing terms are sometimes used in references to the client side, such as "local" [6], "home" [17] and "my place" [35, this paper].

This portrayal of the client side as being "obviously more privacy-friendly" also from a user's point of view has however never been empirically verified. In this paper, we contrast CSP with personalization performed by three different remote providers. We compare personalization at

- the user's smartphone (this is the CSP condition),
- American Personalization, a fictitious company with which no study participant would have prior negative or positive experience,
- Amazon, a company that generally enjoys a high reputation [27,41], and
- the "Cloud", which many users claim they do not entrust with sensitive data [17].

Our experiment includes a strong behavioral component, since prior studies found considerable discrepancies between users' stated privacy attitudes and observed behaviors (e.g., [47,38]). To make the differences between client-side and remote personalization as tangible as possible, we also took care to create an experimental setting that includes a local software client, namely an Android app named Check-it-Out (CiO) that runs on participants' smartphones. CiO purportedly gives personalized recommendations based on users' demographic and context data[2]. In the client-side condition, CiO claims to carry out personalization locally, and in the "remote" conditions, to interact with one of the three remote personalization providers. Finally, we took great care to ascertain users' comprehension of their personalization scenario and its privacy implications.

In the remainder of this paper, we will first review recent work on CSP, to characterize the class of applications that our study addresses. Based on prior privacy research, we then postulate a *model* of users' privacy attitudes and behaviors when personalization is carried out by different providers, both locally and remotely. The model not only allows us to describe and analyze the status quo of personalization performed by the different providers, but also to make *predictions* about the effects of changes to the current situation. We describe our experiment with the CiO prototype, and present the results of a structural equation modeling analysis of the data of 390 participants to validate our model. Finally, we discuss the results with regard to our research question and their implication for CSP.

PRIOR WORK ON CSP FOR PRIVACY

Client-side personalization as a means for privacy protection was proposed over a decade ago [6,18,35]. Broader technical developments in this area however only started in recent years. To the best of our knowledge, only research prototypes have been developed so far.

In client-side personalization, the user profile acquisition methods are largely the same as in remote personalization. However, the inference methods become quite limited, since the users' personal data never leaves the client. Typical methods that can still be used are if-then rules, and classifying users under group profiles/stereotypes/personae with associated personalization rules [26]. Those rules and profiles could stem from prior market or user research, or be based on data of users who did not opt for CSP. Personalization methods that require data of many users (such as collaborative filtering) can still be *added* to the CSP paradigm though if these methods can be carried out in a different privacy-friendly manner (e.g., anonymously or using homomorphic encryption; see [28] for a survey).

We can divide CSP into two categories based on whether the client actually performs the personalization itself or serves as a platform for delivering remote personalization:

- *Client performs personalization*
  The CSP functionality tracks the user locally and performs all personalization without contacting a remote site. This type of CSP has currently been mostly implemented in online behavioral advertising [16,18,50].
- *Personalization code runs on the client side*
  The client stores the user's information and allows personalization code from a personalization provider to be executed client-side (e.g., via a browser extension) and to access the user data in this process [2,11,11,13]. The client itself has no personalization capabilities. Businesses may prefer this type of CSP since they can maintain control over the personalization logic and update it anytime. The client can also provide a trusted computing platform to ensure the confidentiality of the (possibly proprietary) personalization code.

Another differentiation of CSP is by generality:

- *Single-application CSP*
  CSP is restricted to a single application only, like a standalone insurance pricing app in a car [10,11]. If multiple apps run on the same platform, each of them would carry out its own CSP.
- *Application-independent CSP*
  CSP can also be made available as a central service for all applications on a client. Examples are MoRePriv [12] and PersonisJ [15]. Both provide client-side personalization services at the OS level for smartphones (Windows Phone 7 and Android OS). They provide basic personalization functions such as interest modeling, and offer APIs for application developers.

In all these cases, care must be taken to ensure that the result of the personalization process does not reveal too much about the user. For instance, one can pre-load *all* available ads locally rather than fetching them individually

[2] "Context data" is data about a user's smartphone usage, such as his/her web browsing, app usage, or location.

from an ad server when needed, as the latter would possibly allow the server to construct a user model based on the known tags and categories of each ad. If this is not possible (e.g., since a client cannot possibly preload tens of thousands of Netflix movies that might become recommended by CSP), care must be taken to ensure that users remain anonymous when personalization leads to a remote action.

A MODEL OF PERCEIVED PRIVACY IN CSP

From a technical and conceptual point of view, denying remote parties access to local data increases the privacy of the data. Since denial of access is inherent in CSP, it can potentially prevent numerous types of privacy breaches [46]. However, there exists no empirical research as yet whether users also *perceive* CSP as more privacy-friendly than remote personalization. The claim that people prefer to have their data kept locally under their physical control [6,15,35] *might* be correct. However, people are also quite concerned about data on their smartphones being lost or stolen, or being accessed by wireless network hackers [9]. Several study participants in [36] "thus decided not to store any sensitive or valuable data on such devices." Also, synchronization of data between different devices is often cumbersome to set up and manage [39], which would speak against CSP across multiple devices. Our research will show that both attitudinal predictions from the literature are correct to some degree and that, as a result, people steer a middle course behaviorally.

To study and evaluate privacy in CSP from the user's point of view, we developed a model that includes both attitudinal and behavioral constructs. Regarding *attitudes*, Perceived Privacy, Perceived Protection from harm, and self-anticipated Satisfaction with a system are central constructs in explanatory models of people's privacy reactions when using personalized services [3,7,25,51]. Perceived Protection and Perceived Privacy are both antecedents of trust [5,8,44][3], but the two are distinct from each other [8,33,40]. Since users of a personalized system must typically rely on self-anticipated rather than post-usage satisfaction when making disclosure decisions, we also measure Satisfaction in this way.

Our first hypotheses H1-H3 operationalize the above claims about users' superior privacy perception of CSP [6,15,35]:

H1. CSP users (i.e. participants in the CSP condition) experience the highest level of Perceived Protection.
H2. CSP users perceive the lowest level of Privacy Concerns regarding CiO.
H3. CSP users perceive the highest level of Satisfaction regarding CiO.

[3] Our Perceived Protection construct is also related to the "benevolence" sub-construct of Mayer et al.'s [32] classical tri-partite trust construct.

Privacy *behavior* is generally measured by the amount of personal data that users disclose and tracking permissions they give [14,22,27,47]. Following the "standard model" that was compiled from several hundred privacy studies [45], we posit a negative effect of Privacy Concerns on Disclosure (H4). Based on [22], we also hypothesize that satisfaction has a positive effect on disclosure (H5). Knijnenburg et al. [21,22,23] found that the type of the item (specifically Demographic vs. Context data) also has an effect on the amount of disclosure. We thus postulate H6, and also argue that the effects of H4 and H5 may be moderated by Type of item:

H4. Users' Disclosure decreases with Privacy Concerns, but the effect may differ per Type of item.
H5. Users' Disclosure increases with Satisfaction, but the effect may differ per Type of item.
H6. Users' Disclosure is higher for Demographics items than for Context items.

In the privacy literature, an inverse correlation has been found between Privacy Concern and Satisfaction [30,34]. Since we did not want to commit to a causal direction, we included a correlation in our model as well [20]. We also hypothesize that Perceived Protection decreases Privacy Concern and increases Satisfaction. Note that the alternative causal direction is possible; in fact, previous work has modeled Perceived Protection and Privacy Concern with bi-directional effects [44] or as second-level constructs [5]. We decided to rather hypothesize a unidirectional effect of Perceived Protection on Privacy Concern, since the questions about Perceived Protection were specific to each Provider (condition), putting it closer in the causal chain to the provider manipulation than Privacy Concern and Satisfaction. Perceived Protection then becomes a manipulation check, and literature on mediation analysis [4] recommends this direction. The effect of Perceived Protection on Satisfaction was also found by [44], and this direction also showed the best fit compared with alternative models.

Numerous studies have shown that different providers, including companies and the Cloud, enjoy different privacy perceptions (e.g., [17,27,41]). The importance of Perceived Protection in determining Privacy Concern and Satisfaction may thus also differ per personalization provider. Moreover, the fact that questions about Perceived Protection were specific to each Provider also makes this variable more likely to interact with Provider to influence Privacy Concerns and Satisfaction. In H7 and H8 below, we therefore assume that the postulated effects may differ by Provider:

H7. Users' Privacy Concerns decrease with Perceived Protection [5,8,44], but the effect may differ by Provider.
H8. Users' Satisfaction increases with Perceived Protection, but the effect may differ by Provider.

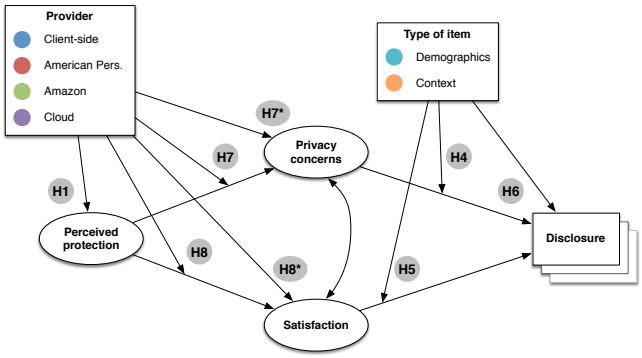

Figure 1: Study hypotheses. Ellipses depict latent factors, and the stacked boxes of Disclosure a repeated measure. Other rectangles depict experimental conditions and controls. H2, H3 and H9 are estimated outside the model.

Fig. 1 shows an overview of these hypothesized effects in our model[4]. The combination of H3-H8 suggests an indirect (mediated) effect of personalization provider on disclosure. Since CSP enjoys superior privacy perceptions according to H1, this leads to:

H9. Overall, CSP users exhibit the highest level of Disclosure.

EXPERIMENTAL SETUP

Experimental procedures

Our study was advertised as an opportunity to learn about an Android app that gives personalized recommendations, to download and work with it, and tell one's opinion about it. No indication of the originator of the user study was given, to avoid biasing subjects (e.g., induce a feel of safety or unease by mentioning a well-known company name).

Instruction and Comprehension Tests

Participants who followed the invitation were randomly assigned to one of four conditions. First, a smartphone app named "Check-it-Out" was described to them. Everyone was told that CiO would analyze what they did on their smartphone, specifically: the webpages that they visit; the email messages that they send and receive; their Facebook posts, and posts of others on their wall; the music to which they listen; and with whom they talk or text on the phone.

Participants in the CSP condition were then told that all these activities on their smartphone *will be kept on their smartphone*, and not be shared with anyone else. Subjects in the three remote conditions were instead told that all these activities on their smartphones *will be sent to American Personalization / Amazon / the Cloud*, and not be shared with anyone else. These entities were not further explained.

After these instructions came a quick comprehension test. Subjects who failed it received the same instructions for a second time, repeated the test and were terminated if they failed again. The survey then presented three examples of the personalized services of Check-it-Out:

- CiO points out an upcoming U2 concert, since the user played their music and chatted with friends about U2.
- CiO points out a Sears promotion for appliances, since the user searched for dishwashers on the Web.
- CiO recommends a friend of a friend who is interested in Salsa dancing, since the user searched for a Salsa class online and bought a book on that topic.

Subjects were then administered a more extensive 15-item comprehension test whose answer correctness depended on the condition they were in.

"Testing the app"

Participants were then told that CiO can give even better recommendations if it has additional information about them. They were asked to download and to install the CiO app on their Android phone. In the client-side condition, the app informed participants that all data that they enter would remain on their phone. Participants were encouraged to turn off their network connection. In the three remote conditions, participants were instead told that all data that they enter will be sent to American Personalization / Amazon / the Cloud, for generating personalized recommendations. If their network connection was disabled, participants were asked to turn it on. The app would not proceed otherwise.

The app then asked participants a sequence of 12 questions about their Demographics (e.g., the size of their household), alternating with 12 permission requests to track various Context data (e.g. "May we track your location?"). See below for details. Participants could individually answer questions and grant requests or decline to do so.

Attitudinal survey

Participants were then asked about their anticipated Satisfaction and Privacy Concerns with CiO, and their Perception of the Protection provided by their smartphone, American Personalization, Amazon, and the Cloud. The questionnaire items will also be explained in detail below.

Pilot testing

The instructions and comprehension tests for the four experimental conditions as well as the attitudinal survey were pilot-tested with 16 participants from the Puget Sound area. We gauged users' comprehension of the different personalization scenarios, as well as the clarity of the survey items and their convergence onto postulated factors. We made some minor adjustments to the experimental materials based on those pilot tests. Since some participants mentioned their concern about loss and theft of smartphone

[4] To correctly estimate the moderated effects H7 and H8, the model has to include intercepts for Provider, which are represented by H7* and H8* in Fig. 1. In contrast, H2 and H3 concern the "overall" effect of Provider on Privacy Concerns and Satisfaction, absent of the Perceived Protection effect. They are thus estimated outside the model. The same holds true for H9.

Subjective construct	Items	Factor loading
Perceived Protection Alpha: 0.95 AVE: 0.886 $r_{\text{Privacy Concerns}}$: −0.510 $r_{\text{Satisfaction}}$: 0.490	I feel my personal data is safe [on my smartphone / at American Personalization / at Amazon / in the Cloud]	0.917
	I feel [my smartphone / American Personalization / Amazon / the Cloud] will not share my personal data with anyone	0.954
	I feel my interests will be protected when my personal data is [on my smartphone / with American Personalization / with Amazon / in the Cloud]	0.953
Privacy Concerns Alpha: 0.77 AVE: 0.593 $r_{\text{Perceived Protection}}$: −0.510 $r_{\text{Satisfaction}}$: −0.720	Check-it-Out has too much information about me	0.752
	Check-it-Out does not know anything I would be uncomfortable sharing with it	
	I felt tricked into disclosing more information than I wanted	
	I find the questions intrusive that Check-it-Out asks me	0.854
	I'm afraid Check-it-Out discloses information about me to third parties	0.696
Satisfaction Alpha: 0.92 AVE: 0.725 $r_{\text{Perceived Protection}}$: 0.490 $r_{\text{Privacy Concerns}}$: −0.720	Check-it-Out is useful	0.887
	Using Check-it-Out makes me happy	0.882
	Using Check-it-Out is annoying	−0.730
	Overall, I am satisfied with Check-it-Out	0.920
	I would recommend Check-it-Out to others	0.905
	I would quickly abandon using this system	−0.764

Table 1: Items measuring the subjective constructs, with their CFA loadings and validity statistics. Items without loadings were removed from the CFA.

data (cp. [9,36]), we added survey items in the client-side condition that gauged participants' Perceived Protection if remote locking and/or automatic backup were available.

Participant recruitment and screening

Study participants were recruited through the crowdsourcing platform Mechanical Turk and a similar corporate service. They were restricted to U.S. residents and received a reward of US$ 2.50 for the valid completion of the study. Announcements were also posted in eight metro areas across the U.S. via Craigslist.com. The first 100 Craigslist participants received a $10.00 Amazon coupon. We verified that removing any of the subsamples does not change our results described below, thus increasing their robustness.

Since MTurk's General Policies do not allow HITs that require workers to download software, we followed [19] and gave them the choice between the full study including app download, or merely completing the survey part for a reward of US$ 0.25. 63.5% of those who selected this latter option indicated not owning an Android phone. We found no significant differences in collection- and control-related privacy concerns between those who chose to download the CiO app and the "control group" who did not.

Participants' results were filtered for completeness, uniqueness of IP address, other signs of multiple submission, duration of survey completion, and correct answers to attention tests. The data of 390 subjects passed this screening and was used in our statistical analysis. Their average score on the 15-item comprehension test was 13.4. Just 8 subjects scored lower than 10, which we deemed quite satisfactory. As an extra precaution, we also performed the complete statistical analysis discussed below on subjects who scored 10+. The difference in results was minimal.

Manipulations and measurement

Personalization provider

Provider was manipulated between subjects, to obtain more insights on established and future personalization providers:

- *Client-side* ("all data remains on your smartphone"),
- *American Personalization* ("all data is sent to A.P."),
- *Amazon* ("all data is sent to Amazon"),
- *Cloud* ("all data is sent to the Cloud").

Amazon and the name "American Personalization" were chosen based on a pre-study (N=99) on trust perception (single-item) of various existing and fictitious company names. Amazon turned out to be a positive extreme while American Personalization was a neutral anchor point.

Perceived Protection, Privacy Concerns and Satisfaction

The attitudinal constructs Perceived Protection, Privacy Concerns and Satisfaction were each measured with multiple items on a 7-point scale, ranging from "strongly agree" to "strongly disagree". The items are based on or derived from [22,24,51]. A total of 14 items were subjected to a Confirmatory Factor Analysis (CFA). We used a weighted least squares estimator that treats the items as ordered-categorical, thereby not assuming normality. Table 2 lists these items and their factor scores. Two items were removed due to low communality, high cross-loadings, or residual correlations. The resulting latent factor structure shows good convergent and discriminant validity[5].

[5] Commonly accepted cutoff values for convergent validity are AVE > 0.5 and Cronbach's alpha > 0.7. Discriminant validity is attained if the square root of AVE is higher than the highest correlation between factors.

Seq. #	Item	Disclosure
	Demographics data	
1	Phone data plan	94.9%
3	Household composition	87.4%
5	Field of work	91.5%
7	Housing situation	85.9%
9	Relationship status	93.6%
11	Children	90.0%
13	Household income	80.8%
15	Household savings	66.7%
17	Household debt	68.5%
19	Race	93.1%
21	Political preferences	82.8%
23	Workout routine	85.1%
	Context data	
2	Recommendation browsing	79.5%
4	Location	50.5%
6	App usage	72.8%
8	App usage location	56.2%
10	App usage time	70.5%
12	Web browsing	56.9%
14	Calendar data	49.7%
16	E-mail messages	16.4%
18	Phone model	83.3%
20	Accelerometer data	58.2%
22	Microphone	17.9%
24	Credit card purchases	0.0%

Table 2: Items requested by Check-it-Out

RESULTS

The model was tested using repeated-measures Structural Equation Modeling (SEM) with a weighted least squares estimator. This model has an excellent overall fit ($\chi^2(680) = 638.076$, $p = .874$, $RMSEA < .001$, 90% CI: [.000, .005], $CFI = 1.00$, $TLI = 1.00$)[6]. Conceptually, the SEM can be seen as a series of regressions: items with incoming arrows are dependent variables and items with outgoing arrows are independent variables. The results for each independent variable are discussed below and summarized in Fig. 2.

Perceived Protection (H1)

Table 3 shows the leftmost part of the model, i.e., the regression of Perceived Protection on the four Providers. As Perceived Protection is an intercept-free scale, its sample standard deviation is set to one, and its value is fixed at zero for the Client-side condition. The first line of Table 3 presents an omnibus test of the differences between the conditions; the remaining lines compare each alternative version to the client-side baseline.

The results show that the four versions differ in Perceived Protection. Client-side has a significantly higher Perceived Protection than American Personalization and Cloud, but Amazon enjoys the highest level; H1 is thus partially supported.

[6] A non-significant chi-square indicates that there is no significant difference between the presented model and a saturated model, which means that the model accounts for nearly all variance. The alternative fit indices have the following accepted cut-off values: CFI > 0.96, RMSEA < 0.05 (within [0.00, 0.10]), CFI > .96, TLI > .95.

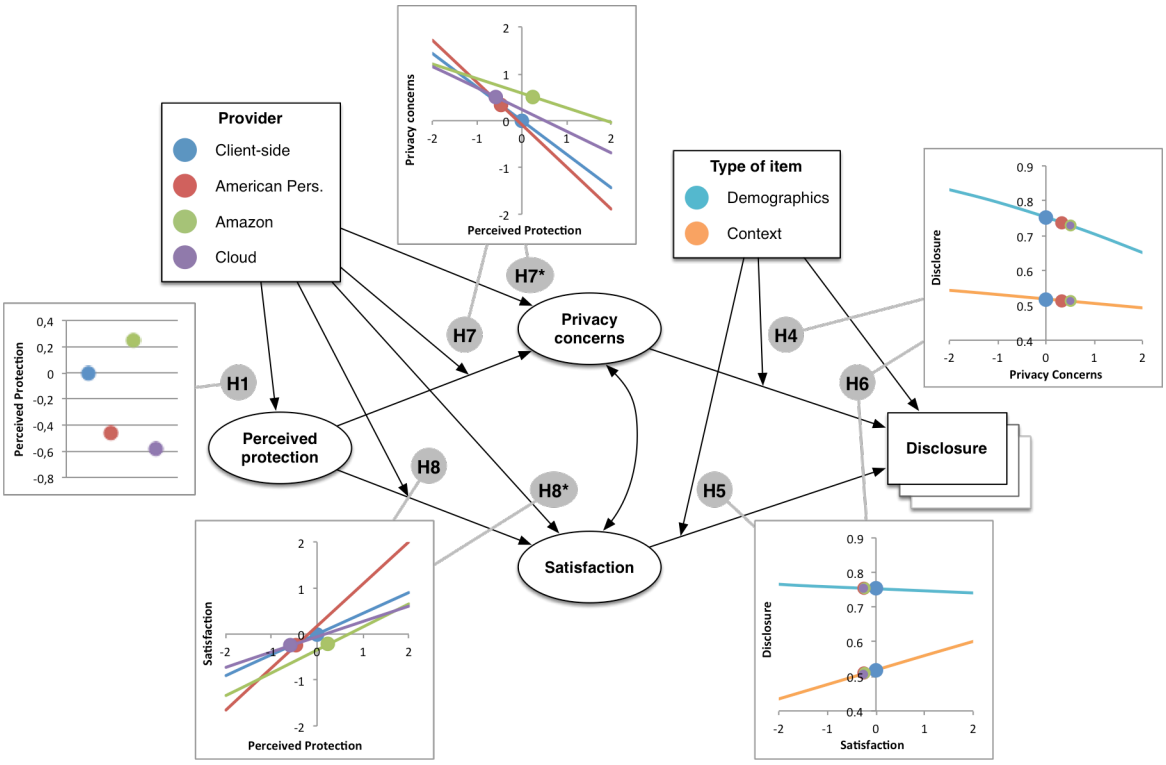

Figure 2: Study Results. Colored dots depict mediated effects of condition on Privacy Concerns, Satisfaction, and Disclosure

Independent var.	Coef.	(95% CI)	*p*-value
H1. Provider	$\chi^2(3) = 78.537$		< .001
Client-side	0		
American Pers.	−0.460	(−0.664,−0.256)	< .001
Amazon	0.244	(0.034, 0.454)	.023
Cloud	−0.583	(−0.797,−0.369)	< .001

Table 3: Linear regression of Perceived Protection

Privacy Concerns, Satisfaction: Unmediated Effects (H2-H3)
Table 4 shows the overall effect of the four Providers on Privacy Concerns. This effect is presented as an *unmediated* (marginal) effect; it is *not* part of the model of Fig. 1. The moderated mediation via Perceived Protection that is estimated in the model is discussed below (H7). Again, the value of the dependent variable is fixed at zero for the Client-side condition.

The differences in Privacy Concerns between conditions are small overall (p=.107), but the Privacy Concerns in the Client-side condition seem to be significantly lower than in the Amazon (p=.032) and Cloud conditions (p=.027), and somewhat lower than in the American Personalization condition (p=.12). H2 is thus supported. The next section will show that this effect is *mediated* by Perceived Protection, thus increasing its robustness.

Independent var.	Coef.	(95% CI)	*p*-value
H2. Provider	$\chi^2(3) = 6.095$		.107
Client-side	0		
American Pers.	0.258	(−0.067, 0.583)	.120
Amazon	0.367	(0.032, 0.702)	.032
Cloud	0.360	(−0.041, 0.679)	.027

Table 4. Unmediated effects of Provider on Privacy Concerns

Table 5 shows the overall effect of the four Providers on Satisfaction. Again, the moderated mediation via Perceived Protection that is estimated in the model is discussed below (H8). Satisfaction is somewhat higher in the Client-side condition, but statistically speaking there are no significant differences. H3 is thus not supported. The next section will show that despite the lack of an unmediated effect, Provider does has an effect on Satisfaction that is mediated by Perceived Protection.

Independent var.	Coef.	(95% CI)	*p*-value
H3. Provider	$\chi^2(3) = 3.852$		.278
Client-side	0		
American Pers.	−0.236	(−0.536, 0.064)	.124
Amazon	−0.182	(−0.456, 0.092)	.193
Cloud	−0.224	(−0.491, 0.043)	.098

Table 5. Unmediated effects of Provider on Satisfaction

Disclosure (H4-H6)
Table 6 shows the regression of Disclosure on Privacy Concerns, Satisfaction and item Type. The Disclosure variable is a repeated measure of the odds ratios of disclosure, with 24 measurements (one for each item) per participant.

Privacy Concerns have a significant negative effect on Disclosure, but only for Demographics items (this supports H4). Users with above-average Privacy Concerns (i.e. one standard deviation higher than average) are predicted to be 21.7% less likely to disclose. In contrast, Satisfaction has a significant positive effect on Disclosure, but only for Context items (this supports H5). Users with above-average Satisfaction are predicted to be 18.3% more likely to disclose. Privacy Concerns and Satisfaction thus have complementing effects on the Disclosure of Demographics and Context items.

Independent Variable	Odds Ratio	(95% CI)	*p*-value
H4. Priv. Concerns × Type of item	$\chi^2(1) = 4.485$		.034
Demographics	0.783	(0.681, 0.900)	.001
Context	0.950	(0.857, 1.054)	.335
H5. Satisfaction × Type of item	$\chi^2(1) = 5.359$		.021
Demographics	0.969	(0.849, 1.104)	.631
Context	1.183	(1.068, 1.310)	.001
H6. Type of item	$\chi^2(1) = 303.962$		< .001
Demographics	3.050	(2.674, 3.478)	
Context	1.076	(0.995, 1.212)	

Table 6: Repeated logistic regression of Disclosure

At average levels of Privacy Concerns and Satisfaction, there is a significant difference in Disclosure between Context and Demographics items; H6 is thus supported. The odds of disclosure for Demographics items are predicted to be 75.3%, and the odds for Context items 51.8%.

Privacy Concerns and Satisfaction: The Mediating Role of Perceived Protection (H7-H8)
Table 7 shows the regression of Privacy Concerns on the four Providers and on Perceived Protection per Provider. We have already addressed the effect of Provider on Privacy Concerns as an unmediated effect; here we rather focus on the mediating role of Perceived Protection.

Independent var.	Coef.	(95% CI)	*p*-value
H7. Provider*	$\chi^2(3) = 25.055$		< .001
Client-side	0		
American Pers.	−0.086	(−0.329, 0.157)	.486
Amazon	0.588	(0.292, 0.884)	< .001
Cloud	0.238	(−0.038, 0.514)	.092
H7. Perc. Protection × Provider	$\chi^2(3) = 11.960$		.008
Client-side	−0.717	(−0.991,−0.443)	< .001
American Pers.	−0.902	(−1.176,−0.628)	< .001
Amazon	−0.312	(−0.512,−0.112)	.002
Cloud	−0.463	(−0.683,−0.243)	< .001

Table 7: Linear regression of Privacy Concerns

Privacy Concerns decrease with Perceived Protection, but the effect differs per Provider; this supports H7. Specifically, an increase or decrease in Perceived Protection for the Client-side and American Personalization conditions have a much larger effect on Privacy Concerns than an increase or decrease in Perceived Protection for the Amazon and Cloud versions.

Independent var.	Coef.	(95% CI)	*p*-value
H8. Provider*	$\chi^2(3) = 15.483$		*.001*
Client-side	0		
American Pers.	0.174	(−0.087, 0.437)	.195
Amazon	−0.347	(−0.570,−0.124)	.002
Cloud	−0.058	(−0.301, 0.185)	.638
H8. Perc. Protection × Provider	$\chi^2(3) = 12.992$		*.005*
Client-side	0.452	(0.260, 0.644)	< .001
American Pers.	0.917	(0.654, 1.180)	< .001
Amazon	0.502	(0.318, 0.686)	< .001
Cloud	0.333	(0.161, 0.505)	< .001

Table 8: Linear regression of Satisfaction

Table 8 shows the regression of Satisfaction on Provider, and on Perceived Protection per Provider. Again, here we focus on the mediating role of Perceived Protection. Satisfaction increases with Perceived Protection, but like for Privacy Concerns, the effect differs per Provider; this supports H8. It has the largest effect for the American Personalization and the least for the Cloud condition.

Combining H7-H8 with H1, we can conclude that Perceived Protection *mediates* the effect of Provider on Privacy Concerns and Satisfaction[7]. The Client-side condition thus enjoys low Privacy Concerns and high Satisfaction because it scores high in terms of Perceived Protection. Amazon is also rated high on Perceived Protection, but it does not do as well in terms of Privacy Concern and Satisfaction. This is reflected in its intercepts (H7* and H8*), which counter the effect of Perceived Protection.

Disclosure rates for different Providers (H9)

The results show that disclosure rates depend on Satisfaction and Privacy Concerns, which in turn depend on the Perceived Protection, which differs per provider. There is thus an indirect (i.e. mediated) effect of Provider on Disclosure. This effect is small though; the dots in the graphs of Privacy Concerns and Satisfaction on Disclosure in Fig. 2, which represent the *total* effects of different Providers, are in an almost identical position. In fact, the differences in unmediated disclosure rates between Client-side (68.7%), American Personalization (66.6%), Amazon (67.8%) and Cloud (68.7%) are very small. H9 is thus not supported.

DISCUSSION

The results of our study give nuanced answers to our research questions. With regard to Privacy Concerns, CSP indeed comes out the best among the four tested conditions. It has no lead though with regards to Perceived Protection (Amazon does better) and consequent Satisfaction (CSP is best, but not significantly). In contrast, users in the Amazon condition perceived the highest level of Protection, but also the highest level of Privacy Concerns[8]. The other Providers show similar pluses and minuses, and hence Disclosure of participants does not differ very much between Providers. Our model allows us to attribute this similarity in behavioral outcomes to attitudinal causes, which in turn creates opportunities for design interventions that may alleviate some of the specific concerns.

For example, our model suggests that one may be able to increase disclosure by increasing the Perceived Protection of the Provider. Doing so will increase Satisfaction and decrease Privacy Concerns, which in turn increases Disclosure. For CSP, this effect seems to be particularly strong: a 1 SD increase in Perceived Protection leads to a 0.72 SD decrease in Privacy Concerns, a 0.45 SD increase in Satisfaction, and eventually a 18.0% increase in Demographics Disclosure and a 12.6% increase in Context Disclosure.

Is it possible to increase the Perceived Protection of CSP? From [9,36] we know that people are concerned about data on their smartphones being lost or stolen, and we also heard this from our pilot participants. We therefore asked subjects in the Client-side condition not only to rate the Perceived Protection of their smartphone today, but also after the following data security enhancements are introduced:

a) a feature to periodical back up their data,
b) a feature to remotely lock the phone if lost or stolen,
c) both of the above features.

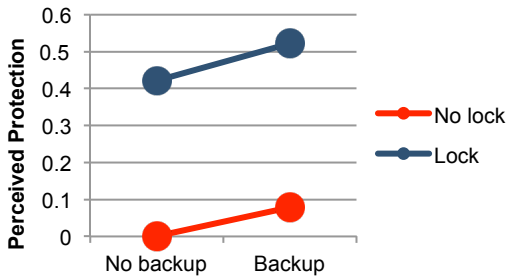

Figure 3: Improving Perceived Protection for CSP through Remote Locking and/or Automatic Backups

Fig. 3 shows how Perceived Protection improves when these two features are introduced in CSP. Allowing users to remotely lock their phones would cause a big increase in Perceived Protection, by 0.43 standard deviations (p <.001). This surge would in turn decrease users' Privacy Concerns by 0.31 SD and slightly increase their Satisfaction by 0.19 SD. Ultimately, users' odds of Demographic and Context data Disclosure would rise by 7.4% and 5.1%, respectively. This would about double the behavioral lead of CSP over the other solutions. Allowing users to create a periodic

[7] Researchers disagree whether a mediated effect has substantive value in the absence of an unmediated effect [31]. We caution not to read too much into this mediated effect on Satisfaction.

[8] A reviewer pointed out that Amazon's high perceived protection might entice users to disclose more data, leading to privacy concerns over the amassed personal data.

backup of their data also increases the Perceived Protection, but to a much smaller extent (0.096 SD, p=.005). The interaction between the two improvements is not significant.

CONCLUSION

We developed a causal model of privacy attitudes and behaviors in client-side personalization and validated it in an experiment that contrasted CSP with personalization performed at three select remote Providers. Many constructs, items and hypotheses of the model have been taken from prior research of others and ourselves. The replicated validation of model elements increases their robustness.

Our model shows that the Provider not only has different unmediated and mediated effects on various privacy-related constructs, but that causal effects between constructs also vary per Provider. These causal constructs, in turn, have complementary effects on the disclosure of different types of data: Privacy Concerns influence the disclosure of Demographics data only, while Satisfaction has an effect merely on Context data. This difference is well worth exploring further.

Our results paint a checkered, yet ultimately promising picture for users' privacy perception of client-side personalization. CSP enjoys lower Privacy Concerns, but it suffers from lower Perceived Protection on smartphones [9,36,39]. As a consequence, users' satisfaction with CSP and their disclosure behavior do not differ significantly from those for other personalization providers.

Encouragingly, our model predicts that increasing the Perceived Protection for CSP yields noticeable improvements in satisfaction and disclosure. More disclosure, in turn, typically leads to better personalization. We found in our study that remote smartphone locking in the case of loss and theft will considerably heighten users' Perceived Protection; periodic backups will also increase this perception, albeit to a lesser extent. These enhancements will need to be considered to raise the adoption of CSP.

ACKNOWLEDGMENTS

We thank Hichang Cho, Xinru Page, Janice Tsai and the anonymous CHI reviewers for their valuable comments. Part of this research was done while Alfred Kobsa was visiting Microsoft Research in Redmond, WA.

REFERENCES

1. Ahmed, R. and Ho, S.Y. Privacy Concerns of Users for Location-Based Mobile Personalization. *CONF-IRM 2011 Proceedings*, (2011), Paper 10.
2. Ankolekar, A. and Vrandečić, D. Kalpana - enabling client-side web personalization. *17th ACM conference on hypertext and hypermedia*, ACM (2008), 21–26.
3. Awad, N.F. and Krishnan, M.S. The Personalization Privacy Paradox: An Empirical Evaluation of Information Transparency and the Willingness to be Profiled Online for Personalization. *MISQ 30*, 1 (2006), 13–28.
4. Baron, R.M. and Kenny, D.A. The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *J Pers Soc Psychology 51*, 6 (1986), 1173–1182.
5. Casaló, L.V., Flavián, C., and Guinalíu, M. The role of security, privacy, usability and reputation in the development of online banking. *Online Information Review 31*, 5 (2007), 583–603.
6. Cassel, L.N. and Wolz, U. Client Side Personalization. *DELOS Workshop: Personalisation and Recommender Systems in Digital Libraries*, (2001), 8–12.
7. Chellappa, R.K. and Sin, R. Personalization versus Privacy: An Empirical Examination of the Online Consumer's Dilemma. *Information Technology and Management 6*, 2-3 (2005), 181–202.
8. Chellappa, R.K. *Consumers' Trust in Electronic Commerce Transactions: The Role of Perceived Privacy and Perceived Security*. Emory Univ., Atlanta, GA, 2008.
9. Chin, E., Felt, A.P., Sekar, V., Wagner, D. Measuring user confidence in smartphone security and privacy. *8th Symp. on Usable Privacy and Security*, (2012), 1:1–16.
10. Coroama, V. and Langheinrich, M. Personalized Vehicle Insurance Rates: A Case for Client-Side Personalization in Ubiquitous Computing. *PEP06, CHI Workshop on Privacy-Enhanced Personalization*, (2006), 56–59.
11. Coroama, V. The Smart Tachograph: Individual Accounting of Traffic Costs and Its Implications. In K.P. Fishkin, B. Schiele, P. Nixon, A. Quigley, eds., *Pervasive Computing: 4th Int'l Conf*. Springer 2006, 135–152.
12. Davidson, D. and Livshits, B. *MoRePriv: Mobile OS-Wide Application Personalization*. Microsoft Research, Redmond, WA, 2012.
13. Fredrikson, M. and Livshits, B. RePriv: Re-imagining Content Personalization and In-browser Privacy. *IEEE Symposium on Security and Privacy,* (2011), 131–146.
14. Van de Garde-Perik, E., Markopoulos, P., de Ruyter, B., Eggen, B., and Ijsselsteijn, W. Investigating Privacy Attitudes and Behavior in Relation to Personalization. *Social Science Computer Review 26*, 1 (2008), 20–43.
15. Gerber, S., Fry, M., Kay, J., Kummerfeld, B., Pink, G., Wasinger, R. PersonisJ: Mobile, Client-Side User Modelling. In P. Bra, A. Kobsa, D. Chin, eds *User Modeling, Adaptation and Personalization*. Springer 2010, 111–22.
16. Guha, S., Cheng, B., and Francis, P. Privad: practical privacy in online advertising. 2011 *USENIX conf. on networked systems design and implementation*, paper 13.
17. Ion, I., Sachdeva, N., Kumaraguru, P., and Čapkun, S. Home is safer than the cloud!: privacy concerns for consumer cloud storage. *7th Symposium on Usable Privacy and Security*, ACM (2011), 13:1–13:20.
18. Juels, A. Targeted Advertising ... and Privacy Too. In D. Naccache, ed., *Topics in Cryptology — CT-RSA 2001*. Springer, Berlin/Heidelberg, 2001, 408–424.
19. Kanich, C., Checkoway, S., and Mowery, K. Putting out a HIT: crowdsourcing malware installs. *5th USENIX conference on offensive technologies*, (2011), 9:1–9:10.
20. Kline, R.B. *Principles and Practice of Structural Equation Modeling*. Guilford Press, 2011.

21. Knijnenburg, B.P., Kobsa, A., and Jin, H. Dimensionality of information disclosure behavior. *Int J Human-Computer Studies 71*, 12 (2013), 1144–1162.
22. Knijnenburg, B.P. and Kobsa, A. Making Decisions about Privacy: Information Disclosure in Context-Aware Recommender Systems. *ACM Transactions on Interactive Intelligent Systems 3*, 3 (2013), 20:1–20:23.
23. Knijnenburg, B.P. and Kobsa, A. Helping users with information disclosure decisions: potential for adaptation. *ACM IUI conference* (2013), 407–416.
24. Knijnenburg, B.P., Willemsen, M.C., Gantner, Z., Soncu, H., and Newell, C. Explaining the user experience of recommender systems. *User Modeling and User-Adapted Interaction 22*, 4-5 (2012), 441–504.
25. Knijnenburg, B.P., Willemsen, M.C., and Hirtbach, S. Receiving Recommendations and Providing Feedback: The User-Experience of a Recommender System. In F. Buccafurri and G. Semeraro, eds., *E-Commerce and Web Technologies*. Springer, 2010, 207–216.
26. Kobsa, A., Koenemann, J., Pohl, W. Personalized Hypermedia Presentation Techniques for Improving Customer Relationships. *Knowl Eng Rev 16*, 2 2001, 111–55.
27. Kobsa, A. and Teltzrow, M. Contextualized Communication of Privacy Practices and Personalization Benefits: Impacts on Users' Data Sharing Behavior. In D. Martin and A. Serjantov, eds., *Privacy Enhancing Technologies:* Springer, 2005, 329–343.
28. Kobsa, A. Privacy-Enhanced Web Personalization. In P. Brusilovsky, A. Kobsa and W. Nejdl, eds., *The Adaptive Web: Methods and Strategies of Web Personalization*. Springer, 2007, 628–670.
29. Li, T. and Unger, T. Willing to pay for quality personalization? Trade-off between quality and privacy. *Eur J of Information Systems 21*, 6 (2012), 621–642.
30. Lukaszewski, K.M., Stone, D.L., and Stone-Romero, E.F. The Effects of the Ability to Choose the Type of Human Resources System on Perceptions of Invasion of Privacy and System Satisfaction. *Journal of Business and Psychology 23*, 3-4 (2008), 73–86.
31. MacKinnon, D. *Introduction to Statistical Mediation Analysis*. CRC Press, 2007.
32. Mayer, R.C., Davis, J.H., and Schoorman, F.D. An Integrative Model Of Organizational Trust. *Academy of Management Review 20*, 3 (1995), 709–734.
33. Miyazaki, A.D. and Fernandez, A. Internet Privacy and Security: An Examination of Online Retailer Disclosures. *J Public Policy & Marketing 19*, 1 (2000), 54–61.
34. Mossholder, K.W., Giles, W.F., and Wesolowski, M.A. Information privacy and performance appraisal: An examination of employee perceptions and reactions. *Journal of Business Ethics 10*, 2 (1991), 151–156.
35. Mulligan, D. and Schwartz, A. Your Place or Mine?: Privacy Concerns and Solutions for Server and Client-Side Storage of Personal Information. *Tenth conference on Computers, Freedom and Privacy*, (2000), 81–84.
36. Muslukhov, I., Boshmaf, Y., Kuo, C., Lester, J., and Beznosov, K. Understanding Users' Requirements for Data Protection in Smartphones. ICDEW 2012, 228–35.
37. Newman, G.H. and Enscoe, C.J. System and method for providing client side personalization of content of web pages and the like. 2000. http://www.google.com/patents?id=VI0EAAAAEBAJ.
38. Norberg, P.A., Horne, D.R., Horne, D.A. The Privacy Paradox: Personal Information Disclosure Intentions versus Behaviors. *J Consum Aff 41*, 1 (2007), 100–126.
39. Oulasvirta, A. and Sumari, L. Mobile kits and laptop trays: managing multiple devices in mobile information work. *CHI* 2007, 1127–1136.
40. Pirim, T., James, T., Boswell, K., Reithel, B., and Barkhi, R. An Empirical Investigation of an Individual's Perceived Need for Privacy and Security. *Int'l Journal of Information Security and Privacy 2*, 1 (2008), 42–53.
41. Ponemon. *2012 Most Trusted Companies for Privacy*. Ponemon Institute, 2013. http://www.ponemon.org/local/upload/file/2012%20MTC%20Report%20FINAL.pdf
42. Schmidt, E. *36h MacTaggart Lecture*. 2011. http://www.youtube.com/watch?v=hSzEFsfc9Ao#t=1224s.
43. Shen, X., Tan, B., and Zhai, C. Implicit user modeling for personalized search. ACM ICIKM 2005, 824–831.
44. Shin, D.-H. The effects of trust, security and privacy in social networking: A security-based approach to understand the pattern of adoption. *Interacting with Computers 22*, 5 (2010), 428–438.
45. Smith, H.J., Dinev, T., and Xu, H. Information Privacy Research: An Interdisciplinary Review. *MIS Quarterly 35*, 4 (2011), 989–1016.
46. Solove, D.J. A Taxonomy of Privacy. *University of Pennsylvania Law Review 154*, 3 (2006), 477–564.
47. Spiekermann, S., Grossklags, J., Berendt, B. E-privacy in 2nd Generation E-Commerce: Privacy Preferences versus Actual Behavior. *ACM EC* 2001, 38–47.
48. Teltzrow, M. and Kobsa, A. Impacts of User Privacy Preferences on Personalized Systems: a Comparative Study. In C.-M. Karat, J. Blom and J. Karat, eds., *Designing Personalized User Experiences for eCommerce*. Kluwer, 2004, 315–332.
49. Toch, E., Wang, Y., and Cranor, L.F. Personalization and Privacy: A Survey of Privacy Risks and Remedies in Personalization-Based Systems. *User Modeling and User-Adapted Interaction 22*, 1-2 (2012), 203–220.
50. Toubiana, V., Narayanan, A., Boneh, D., Nissenbaum, H., and Barocas, S. Adnostic: Privacy Preserving Targeted Advertising. *NDSS*, (2010) http://www.nyu.edu/pages/projects/nissenbaum/papers/adnostic.pdf.
51. Xu, H., Luo, X. (Robert), Carroll, J.M., Rosson, M.B. The personalization privacy paradox: An exploratory study of decision making process for location-aware marketing. *Decis Support Syst 51*, 1 (2011), 42–52.