Online Appendix to: Making Decisions about Privacy: Information Disclosure in Context-Aware Recommender Systems

BART P. KNIJNENBURG and ALFRED KOBSA, University of California, Irvine

A. STATISTICAL METHODOLOGY

The statistical analysis of the experiment presented in this article consists of two parts. We first validate the measured latent concepts in a series of Confirmatory Factor Analyses (CFAs), one for disclosure behavior and one for the subjective measures. Subsequently, we test the structural relations between the manipulations and the subjectively and behaviorally measured latent concepts using Structural Equation Modeling (SEM). In this section, we provide a detailed explanation of these statistical methods using the data from this article as an example.

A.1. Confirmatory Factor Analysis (CFA)

The first step is to confirm whether users' behavior and subjective evaluations measure the predicted latent factors. A latent factor captures the common variance (or 'shared essence') of the items that are used to measure it. A robust factor will only emerge if the items have enough common variance.

The postexperimental questionnaire items are created in such a way that we have a clear preexisting notion of which items measure which factor (see Tables II and III). We thus set up our model in accordance to this hypothesis, as shown in Figure 5 for the factors 'general privacy concerns', 'collection concerns', and 'control concerns'. In this model, each of the items {gpc1...ctrl5} is essentially a regression outcome¹³, predicted by the unobserved latent variables {gpc, coll, ctrl}. { $I_{gpc,gpc1} \dots I_{ctrl,ctrl5}$ } are the loadings of the items on the factors. The model tries to estimate these loadings so that the paths match the covariance matrix of the items {gpc1...ctrl5} as closely as possible (e.g., $cov_{gpc1,gpc2} \approx I_{gpc,gpc1} * I_{gpc,gpc2}$, and $cov_{gpc1,coll1} = I_{gpc,gpc1} * w_{gpc,coll} * I_{coll,coll1}$). Intuitively speaking, factor analysis tries to model the "overlap" between items. The part of the variance that does not overlap (the 'uniqueness') is excluded (and represented by the arrows at the bottom of Figure 5). The more overlap gets extracted, the more reliable the factors are (the reliability of each factor is represented by the slanted arrows pointing to the factors from the top). The factors may be correlated with each other (e.g., $w_{gpc,coll}$). The solution has no standard scale, so we give the factors a standard deviation of 1 and a mean of 0.

If the model was specified correctly, it has a good least-squares fit, and the paths { $I_{gpc,gpc1} \dots I_{ctrl,ctrl5}$ } are significantly larger than zero. However, the following problems may arise.

- A certain item does not have enough in common with the other items to load significantly on its factor: In this case the item has "low communality" and should be removed from the analysis. We typically aim for at least 30% of the variance of an item to be captured by the factor.
- A certain item loads strongly on the wrong factor (this can be discovered by inspecting the 'modification indices' of the model). In this case, the item has a high

¹³Since these outcomes are measured on a 5-point scale, this regression model is an ordinal response model.

Fig. 5. An example CFA, limited to the factors ‘general privacy concerns’, ‘collection concerns’, and ‘control concerns’.

‘cross-loading’, and unless we can come up with a good reason for this to occur, it should be removed from the analysis.

- Two or more items may have a high residual correlation (discovered by inspecting the modification indices): In this case, the factors do not represent the amount of overlap between those items well enough. We can choose to either remove the items, or to ‘split up’ or redefine the existing factors in a way that better captures the common variance of the items.
- The correlation between two factors may be higher than the average variance extracted (AVE) from each item: In this case, we lack discriminant validity, and we must conclude that these two factors essentially measure the same concept.
- A certain factor has only low-loading items, with an AVE of < 0.50 : In this case, the factor is not measured robustly, and should be reconsidered. Another indicator of low robustness is Cronbach’s alpha, which is acceptable at $.7 < \alpha < .8$, good at $.8 < \alpha < .9$, and excellent at $\alpha > .9$.

In the current article, we conducted this iterative process separately for the disclosure behaviors (yes/no items) and questionnaires (5-point scale items). Tables II and III display the resulting factors. Once an adequate factor solution is established, the remaining items are used in the second step of the analysis.

A.2. Structural Equation Modeling (SEM)

The second step is to test the structural relations between the experimental manipulations (justification type and request order) and the latent concepts. In Figure 3 the structural relations are represented graphically as paths (arrows) between manipulations (rectangles) and latent concepts (ovals). Each set of incoming arrows essentially constitutes a linear regression. For example, the arrows pointing to “perceived privacy threat” specify the regression:

$$y_i = \alpha + \beta_1 \eta_{1i} + \beta_2 \eta_{2i} + \varepsilon_i$$

where y_i is the value of “perceived privacy threat” for participant i , η_{1i} is the value of “control concerns” for participant i , η_{2i} is the value of “collection concerns” for participant i , and ε_i is the residual (error). β_1 and β_2 are the regression coefficients (-0.438 and 0.735 respectively). Since the solution has no fixed intercept, α can be ignored. Note that all latent concepts are still measured by their respective indicators such as in Figure 5 (these indicators are hidden in Figure 3 to improve the clarity of the representation).

There are three types of regression paths in Figure 3: paths emanating from a manipulation with two conditions, paths emanating from a manipulation with more than two conditions, and paths emanating from a latent concept.

The request order manipulation has two conditions: demographics first and context first. As is common in linear regressions, this manipulation is represented in the regressions by a single dummy variable: “demographics first”. This means that the baseline model is measured for “context first”, and that the effect of the dummy variable represents the difference between “context first” and “demographics first”. Because all latent factors are standardized, the effect of this difference is measured in sample standard deviations. For example, the number on the arrow from “demographics first” to “context disclosure” (Figure 3) indicates that context disclosure is 0.928 standard deviations lower in the “demographics first” condition than in the “context first” condition. The number in parentheses (0.129) is the standard error of this coefficient. This standard error can be used in a z-test to test whether the path is significant ($p = z[-.928/.129] < .001$).

The justification type manipulation has five conditions, with “no justification” as the natural baseline condition. In the regression equations, a dummy variable is created for each of the four justifications. In Figure 3, the effects labeled A, B, C and D represent the differences between the justifications and the baseline condition. Moreover, the χ^2 -value preceding these effects represents an omnibus test, which verifies whether the effect of all justifications combined is zero (for a significant omnibus effect, this should not be the case).

Finally, there are effects between two latent variables. These are also standardized regression effects, for example, a difference of one standard deviation in “disclosure help” causes a 0.591 SD difference in “trust in the company”.

An *identifiable* structural equation model has to be a directed acyclic graph of causal paths. Such a graph can be constructed by assuming causal “order” between variables. For instance, assuming that $\{A \rightarrow B \rightarrow C \rightarrow D\}$ means that A can cause B, C and D (but B, C and D do not cause A), B can cause C and D (but C and D do not cause B) and C can cause D (but D does not cause C). An initial factor model is set up by defining all possible paths between the (latent) variables that respect the assumed causal order (i.e., the transitive closure). In a typical model, not all initially specified paths are significant, and nonsignificant paths can be removed. If a certain factor has multiple remaining incoming paths, the R^2 -value represents the fraction of variance explained by its predictors. For example, our model explains 27.3% of the variance in context data disclosure (due to the request order manipulation, and the effect of “self-anticipated satisfaction with the system”).

When there are only indirect and no direct effects between two variables, such as for “control concerns” and “trust in the company”, we call this effect “fully mediated”. When there are both direct and indirect effects, such as for “perceived help” and “satisfaction”, the effect is “partially mediated”. Finally, it is possible that the direct effect is positive, but the indirect effects are negative, or vice versa, such as for “justification type” and “trust in the company”. This is called an “inconsistent mediation”. In the case of inconsistent mediations, total effects provide an evaluation of the sum of all positive and negative paths. The total effects can be calculated by adding all effects from a justification to a factor. For instance, the effect of justification A on Trust is the direct effect $A \rightarrow \text{Trust}$, plus the effect $A \rightarrow \text{Help} \rightarrow \text{Trust}$. Using the values in Figure 3: $-0.887 + .998 \cdot .591 = -0.303$.

The final model (after excluding non-significant effects) can be tested as a whole. The model χ^2 -statistic tests the difference in explained variance between the proposed model and a fully specified model. A good model is not statistically different from the fully specified model ($p > .05$). However, this statistic is commonly regarded as

too sensitive, and researchers have therefore proposed other fit indices [Bentler and Bonett 1980]. Based on extensive simulations, Hu and Bentler [1999] propose cut-off values for other fit indices to be: $CFI > .96$, $TLI > .95$, and $RMSEA < .05$, with the upper bound of its 90% CI falling below 0.10.

A.3. Limitations of CFA and SEM

Using confirmatory factor analysis, one may run the epistemological risk of confirming a certain model when a better model exists unknowingly. The number of factors, and which items load on them, is not always straightforward. A look at the correlations between the measured items can give insights into the selection of a factor model; another option is to compare several factor models in terms of model fit. In our case, for example, we explicitly explored various factor solutions for disclosure behavior. The two-factor solution in Table II of the main text, which contrasts demographics disclosure and context data disclosure, had the best fit though.

Like any regression model, structural equation models make assumptions about the direction of causality in the model. From a modeling perspective, an effect (“disclosure help” → “trust in the company”) and its reverse (“trust in the company” → “disclosure help”) are equally plausible. By including manipulations into our model, we are able to “ground” the causal effects: participants are randomly assigned to a condition, so condition assignment cannot be caused by anything in the model (i.e., “justification type” → “disclose help” is possible, but not “disclosure help” → “justification type”). Furthermore, the Knijnenburg et al. [2012] framework provides hypotheses for the directionality of causal effects (since in the framework we hypothesize that $SSA \rightarrow EXP$ and not $EXP \rightarrow SSA$, we can limit ourselves to testing “disclosure help” → “trust in the company”).

Furthermore all modeled structural relations are linear, although it is possible to include transformed variables, such as quadratic effects. In practice, however, such transformations are rarely needed, because latent factors are modeled to have a normal distribution.

B. OVERVIEW OF INTERVIEW FINDINGS

We conducted semi-structured interviews with 17 participants. The interviews lasted between half an hour and one hour, and consisted of two parts. In the first part, we showed participants paper versions of the Applause information disclosure screens. Participants were asked whether they would disclose the requested information, and to elaborate on this decision (i.e., to give a personal reason for disclosing or not disclosing the information). We showed participants a combination of the different justification types (see Table I, the “explanation” justification was added to the online experiment based on comments from the first 11 interview participants, and was therefore only tested in the last 6 interviews). At one point during the interview, we put screenshots of the different justification types next to each other, and explicitly asked participants to compare the justifications and to indicate which one they liked best.

After participants completed their disclosure decisions, we asked them to reflect on the application. We asked them to freely elaborate on their experience, but where appropriate, we primed them with specific questions on satisfaction, trust, threat, perceived help, and general privacy concerns, in line with the questions in the online study (Table III).

The interviews were then transcribed and structured using an ‘open-coding’ process, in which statements from participants are grouped and emerging groups are merged and split until a coherent and useful grouping remains. In the following we do not

strive to capture all the insights that emerged from the interview study, but rather give an overview of the discussed topics and the spectrum of participants' answers.

B.1. Security Issues

- Users think about security issues when faced with the decision to disclose personal information. All participants mentioned at least once that they would not disclose a certain piece of information due to security concerns.
- Recent security breaches can influence users' perception of the security of online services. 4 participants mentioned the recent hacking of the PlayStation Network, and 1 participant said he would have answered the questions differently after he was told about this recent event.
- Users believe that nothing is 100% secure. Although 2 participants specifically stated that they believed the system to be secure, 6 participants expressed the belief that anything can be hacked.
- Users believe that mobile phones can be lost or stolen, and this poses additional security concerns. 3 participants said that they were afraid to lose their phone and that their data could thereby fall into the wrong hands.
- Companies are usually not blamed for security breaches. 3 participants mentioned that as long as the company does everything in their power to protect the data, they are not to blame for eventual breaches.
- Disclosure of credit card purchases was most influenced by beliefs about security. 8 participants gave security as a reason not to disclose credit card information, because they almost instantly linked this to identity theft.
- Location tracking was also influenced by security beliefs. 4 participants mentioned security in this context. These participants were afraid that an unauthorized third party could use the information to find out if they were at home or not (which could help in planning a burglary attempt).
- Email tracking was also influenced by security beliefs. 3 participants mentioned security in this context. These participants mentioned that this was because their email contained passwords and financial information.
- Some users try to cope with security issues by providing less detailed information to the system. 2 users mentioned that due to security issues, they would rather disclose less detailed information than what the system requested.

B.2. Other Uses of Users' Data

- All users worry about possible other uses of the information they disclose. Several participants mentioned at least once that they would not disclose a certain piece of information due to the fear that the information would be used in an unintended way. Others mentioned that they would not disclose certain information or even abandon the system if they found out that the information was being used in an unintended way.
- Most users dislike it when a company sells their personal information. 12 participants mentioned this explicitly, 5 of them were primed about it. However, 4 participants mentioned stated that they would not mind a company selling their personal information.
- Users think that advertisement is the most likely unintended use for their information. 11 participants explicitly mentioned advertisement, 2 of them were primed about it. Advertisement is however not a problem for all users. 6 of participants were okay with the system giving them advertisements based on their personal information. They said that this was just part of the deal. However, 4 participants mentioned that if they would pay for the app, they would expect the company not to sell their information or use it for advertisement.

- Surveillance is another unintended use mentioned by users. 5 participants mentioned this in the context of tracking the microphone. 2 other participants mentioned this in the context of tracking location.

B.3. Privacy Concerns

- Users consider privacy when deciding what requests to answer. 13 participants would not disclose at least one piece of information because they deemed it too private, and 14 participants would disclose at least one piece of information because they did not deem it private. For 11 participants both situations occurred at least once. Only 1 participant did not use privacy as a reason for any of the disclosure decisions. 6 participants mentioned that some information is so private that the system should not even ask for it.
- Users talk about privacy in different ways. 12 participants mentioned that something was (not) “personal”. 6 participants referred to a gut feeling of discomfort. 5 participants mentioned the word “private” or “privacy”. 4 participants mentioned that something was “too much”, “too invasive”, or “too intrusive”. 4 participants said that they did not want the system to know a certain thing, or that they thought the system did not need to know a certain thing. 3 participants mentioned that something was (not) “secret” or “confidential”. 3 participants said that disclosing something “doesn’t bother me”.
- Users say that privacy has priority over usefulness. 8 participants disclosed at least one piece of information despite the fact that they thought it would not be useful for the system, because they did not deem it to be private. On the other hand, 9 participants felt that at least one piece of information was so private that they decided to not disclose it, even though they thought it would be useful for the system.
- The combination of several pieces of information can cause additional privacy concerns. 5 participants mentioned that at some point, they felt that the combination of several items they disclosed caused additional privacy concerns.
- Some users deal with privacy concerns by providing less detailed information. 7 participants mentioned that due to privacy issues, they would rather disclose less detailed information than what the system requested.
- Some users want more manual control over what they disclose to the system. 4 participants mentioned that they wanted to be able to control what the system would track.

B.4. Benefits of Disclosure

- The benefits of disclosure are an important factor in deciding whether to disclose information, but not always. 13 participants explicitly stated that they would disclose something because it would be useful for the system to know, and 11 participants explicitly stated that they would not disclose something because it would not be useful for the system to know. However, 7 participants disclosed something despite thinking that it would not be useful, and 2 participants did not disclose something despite thinking that it would be useful.
- Users will occasionally give their intuitive idea of the benefits of providing certain information priority over the system-stated benefits (i.e. the justification message). 10 participants at least once gave their intuition priority over the system-stated benefits. 7 of them decided that something was useful despite the low percentage in the justification message, while 5 of them decided that something was not useful despite the high percentage in the system message. Only 1 participant stated that

she changed her mind about her initial idea of the usefulness of allowing the system to track her accelerometer when the system showed a low percentage.

B.5. Control

- Users like to have the option to choose what to disclose. 8 participants liked the fact that they could decide what to disclose per specific piece of information, and 7 participants said that they would have problems with a system that uses their information without asking.
- Some users think that the system protects their privacy by giving them control over their settings. 7 participants say that the system protects their privacy because they can choose what to disclose. However, 6 participants did not think that the system protects their privacy at all.
- Giving users control makes some of them more cautious, but others less cautious. 8 participants mentioned that by asking specific questions, the system made them more cautious. On the other hand, 3 participants said that the fact that they could change their level of disclosure allowed them to be less cautious initially. 3 participants did not think that the system influenced their caution.
- Users may refuse to disclose something in fear that the system would make incorrect inferences based on the information. 7 participants did not disclose something because they were afraid it would lead to incorrect recommendations, 6 participants did not disclose something because they believed that the information misrepresented them, and 3 participants mentioned that some people might not disclose something because they could be ashamed of it. Conversely, 6 participants decided to disclose something because it represented them well.
- Many users like to be able to change their disclosure. 16 participants explicitly mentioned this, 9 of them said they would like to be able to change their answers after seeing the actual recommendations.
- Some users may lie as a way to control the system. 3 participants said they would lie on one of the requests.
- Some users want to manage their identity by giving information at a different level of detail. 4 participants mentioned this as a way to give the system a better representation of them.

B.6. Justifications

- Users may only skim the justification message. 8 participants said they did not completely read the message, and just looked at the percentage. 2 of them mentioned that the message was too wordy.
- Most users appreciate the justification message, but some are aware that it may trick them into disclosing more. When asked, 12 participants commented they liked to receive such a message. 5 participants believed that the messages could potentially trick them into disclosing more information, whereas 4 participants said that the system would not be able to trick them. 9 participants said that they would be skeptical about the veracity of the stated percentage.
- Most users are only occasionally influenced by the justification message. 11 participants at least once used a low percentage in the message as a reason not to disclose something. Conversely, 5 participants at least once used a high percentage in the message as a reason to disclose something. On the other hand, all participants at least once said they would disclose something despite a low percentage in the message, and 14 participants at least once said they would not disclose something despite a high percentage.

- The typical percentage at which users are convinced differs per user. 6 participants said that a percentage around or below 50% would be too low. 4 participants said that they would be okay with percentages around or somewhat below 50%.
- Which justification message works best differs per user. The “number of others” message is worst: 3 participants like it best, but 11 participants do not like it; some even find it worse than having no message at all. 10 participants prefer the “useful for you” or “useful for others” message, and these were often considered equal. 3 participants prefer the “explanation” message, but this message was only shown to the last 6 participants. In fact, 10 of the first 11 participants wanted an explanation justification, and 6 of them explicitly said they would prefer such an explanation over any other justification message.

B.7. Effort of Disclosure

- Users do not want to answer too many requests. 9 participants mentioned this, and 5 of them wanted an option to skip certain requests and come back to them later.
- Users will not answer requests that take too much effort to answer. 6 participants decided to not answer a certain request because it was too difficult to answer quickly. 4 participants said that a certain request was superfluous (i.e., the system should already know the answer based on answers to previous questions).
- Whether users take the effort to answer a request may depend on their mood. 2 participants said that they would answer more or fewer requests depending on their mood.

B.8. Disclosure Heuristics

- Users decide whether to disclose something by looking at what they normally do, or what they did before. 13 participants based their disclosure decision on what they normally do, or what they would disclose to other people. 11 participants based their disclosure decision on what they did in a specific other situation (5 participants compare with Google, 3 with Amazon, 2 with Yelp, 1 with Facebook, and 1 with a credit card company). Similarly, 7 participants seemed to disclose something because the information was “already out there anyway”.
- Users’ decision depends on the company that provides the system. 15 participants mentioned the reputation of the company as a reason to disclose or not. 6 of them mentioned that they trust some companies more than others. 3 of them mentioned that they believe some companies would provide a higher security than others. 3 of them mentioned that they would trust a company that has this information already anyway (e.g., a bank). 7 participants believe that that manufacturer of the phone or the network provider would be the best company to provide the recommender service, because they have a clear motive to make the phone usage experience better (and they would have no ulterior motive). 4 participants trust the current company (Appy), whereas 3 participants do not.
- When a certain request does not apply to the user, they will skip the request. 6 participants skipped a request that did not apply to them, even when the request had a specific “does not apply to me” answer. They believe that skipping the request is equivalent to indicating this explicitly. However, 1 participant explicitly said she would answer a question about sports with “I do not follow any sports” in order to not get any sports recommendations.

B.9. Privacy and Responsibility

- Most users believe that online companies have too much information about them. 10 participants think this is the case. However, 9 participants believe that they are more careful about their privacy than others.

- Most users believe that they are themselves ultimately responsible for their privacy. 7 participants said that by choosing to use the app, they were themselves responsible for their privacy. However, 3 participants made the side note that the company needed to be clear about its practices. Similarly, while 3 participants mention that they do not have enough control over their online privacy, 3 other participants find it unrealistic to expect further control over their privacy once they choose to use an app.

B.10. Satisfaction and Usage Intention

- Users are generally positive about the idea of a mobile app recommender system. 15 participants like the idea, and said that they would use it. 7 participants said that their current phone lacks a good way to find apps. 5 of them would recommend Applause to others if it worked well, and 3 of them would even pay for using it.