Jan Tolsdorf*, Delphine Reinhardt, and Luigi Lo Iacono

Employees' privacy perceptions: exploring the dimensionality and antecedents of personal data sensitivity and willingness to disclose

Abstract: The processing of employees' personal data is dramatically increasing, yet there is a lack of tools that allow employees to manage their privacy. In order to develop these tools, one needs to understand what sensitive personal data are and what factors influence employees' willingness to disclose. Current privacy research, however, lacks such insights, as it has focused on other contexts in recent decades. To fill this research gap, we conducted a cross-sectional survey with 553 employees from Germany. Our survey provides multiple insights into the relationships between perceived data sensitivity and willingness to disclose in the employment context. Among other things, we show that the perceived sensitivity of certain types of data differs substantially from existing studies in other contexts. Moreover, currently used legal and contextual distinctions between different types of data do not accurately reflect the subtleties of employees' perceptions. Instead, using 62 different data elements, we identified four groups of personal data that better reflect the multi-dimensionality of perceptions. However, previously found common disclosure antecedents in the context of online privacy do not seem to affect them. We further identified three groups of employees that differ in their perceived data sensitivity and willingness to disclose, but neither in their privacy beliefs nor in their demographics. Our findings thus provide employers, policy makers, and researchers with a better understanding of employees' privacy perceptions and serve as a basis for future targeted research on specific types of personal data and employees.

Keywords: employee privacy, factor analysis, structural equation modeling, latent class analysis

DOI 10.2478/popets-2022-0036
Received 2021-08-31; revised 2021-12-15; accepted 2021-12-16.

***Corresponding Author: Jan Tolsdorf:** Bonn-Rhein-Sieg University of Applied Sciences, E-mail: jan.tolsdorf@h-brs.de
Delphine Reinhardt: University of Göttingen, E-mail: reinhardt@cs.uni-goettingen.de
Luigi Lo Iacono: Bonn-Rhein-Sieg University of Applied Sciences, E-mail: luigi.lo_iacono@h-brs.de

1 Introduction

The fundamental right to privacy applies to all situations and contexts in life, including the employment context. In Europe, the employment context is also subject to the rules of the General Data Protection Regulation (GDPR). One of its key elements for the processing of personal data is the principle of "prohibition with subject to permission". This means that employers may only process personal data of their employees if this is explicitly permitted. Yet, employees often have limited ability to decide on the nature and extent of disclosures of certain personal data because laws or employers' interests outweigh employees' privacy interests. Indeed, this situation can be at odds with employees' perceived right to privacy [93]. For example, employees in Germany are generally required to disclose information about a potential membership in Christian churches, health insurance and social security numbers, child allowance, and contact details. In addition, the country of birth and disabilities may also be processed. Most of these data are perceived by users as highly sensitive in the online and marketing context [1, 54, 81]. In contrast, knowledge on perceived data sensitivity in the employment context is lacking. Given that privacy is known to be a contextual concept [33, 63, 87], we attempt to address this gap in this paper. In particular, we build on the specifics of the employment context and examine whether and to what extent the fact that disclosure of personal data is mandatory affects employees' perceptions of data sensitivity and their willingness to disclose data. Addressing this issue contributes to filling existing research gaps [7], and is also timely and important as more and more data about employees are collected and processed. This includes, for example, data collected as part of mobile working [46], the use of wearables [58], and the use of analytics and monitoring [2]. As a result, this work lays the basis for a better understanding of employees' perceived data sensitivity and their willingness to disclose data in the employment context. Our results can also be leveraged for the design of transparency tools, i.e., making processing operations transparent to employees, as

required by the GDPR. By understanding which data are perceived as sensitive by employees and what differences exist among employees, our results help to design tools tailored to employees' needs [93] and thereby increase the tools' usability [62].

To this end, we make the following contributions: We conducted a cross-sectional survey with 553 employees from Germany to gain a better understanding of perceived data sensitivity and their willingness to disclose these data in employment relationships. Our research contributes to the body of knowledge in several ways: (1) We found that perceived sensitivity in the employment domain differs significantly from the results of previous studies in other domains. Context also appears to be a significantly more decisive factor for perceived sensitivity than the cultural background. (2) Based on an assessment of perceived sensitivity and willingness to disclose 62 different personal data elements, we identified four groups of personal data with distinct characteristics: Two groups representing data not related to the employment relationship, and two groups representing data either disclosed or arising from the employment relationship. Overall, perceived data sensitivity proved to be a fairly stable moderate predictor of employees' willingness to disclose across all data groups. However, context may have different effects on perceived sensitivity and willingness to disclose, causing employees to be potentially unwilling to share non-sensitive data but willing to disclose data perceived as sensitive. (3) We could only partially confirm frequently used antecedents in online research for the willingness to disclose data in the employment context, thus motivating future research in this direction. (4) We further identified three groups of employees: One group was willing to disclose, depending on the data's sensitivity and contextual appropriateness for employment. Another small group was reluctant to disclose truthful data, even if they were essential to the employment relationship. A third group was very willing to disclose all but the most sensitive data. However, the groups did neither differ in their privacy beliefs nor in their demographic backgrounds, leaving open the question of influencing factors.

To the best of our knowledge, we are the first to conduct a comprehensive study on perceived sensitivity and willingness to disclose in the employment context in the cultural space of Europe. Our study contributes to the general body of knowledge in privacy research by providing new insights into privacy in employment. We highlight differences between contexts and make an important contribution to balancing the existing one-sided focus of research on private contexts.

The rest of this work is structured as follows: first, we summarize previous work and introduce our research model and hypotheses in Section 2. We then outline our methodology in Section 3 and present the results of our study in Section 4. We then discuss our findings in Section 5 and highlight limitations as well as opportunities for future work in Section 6.

2 Background and research model

In this section, we provide background information on our research topics of interest and review previous work. We elaborate existing research gaps and derive our research questions and hypotheses guiding our research.

2.1 Perceived data sensitivity

Numerous studies on online environments, smart device use, and marketing show that different types of personal data are perceived differently by users in terms of sensitivity [1, 13, 52, 54, 55, 59–61, 70, 81, 83]. The "sensitivity" property of personal data is commonly defined as *the perceived negative consequences or (potential) loss associated with data disclosure* [13, 61]. Perceived loss is highly context-dependent, which in turn also makes perceived sensitivity context-dependent [48, 56, 85]. However, recent studies in the marketing and online context show that perceived sensitivity seems unaffected by slight context changes. In more detail, different studies have been conducted with samples from the USA and Brazil [54], from Germany [81], and from Saudi Arabia [1]. All found that the ranking of various personal data by perceived sensitivity was largely unaffected by differences in culture and context. This raises the question whether a global consensus for a ranking of personal data by perceived sensitivity can be reached [54, 81].

Furthermore, findings specifically related to perceived data sensitivity in the employment context are largely limited to work from the USA prior to 2000 [51, 76, 91, 99]. The more recent quantitative studies are embedded in a broader "information privacy" framework that focuses on employers' practices in recruitment and on effects of information systems and workplace surveillance [2, 7]. Although the topics covered are complementary to ours, their emphasis is on investigating (adverse) behavioral effects and developing remediation strategies. Aside from these quantitative studies, previous qualitative research revealed that, while employ-

ees can recognize sensitive data, their actions are based on individual interpretations rather than formal rules [35, 86, 93]. In [35], five clusters of cues used by employees to recognize sensitive data are identified, but the focus was not on employees' own personal data. Furthermore, the findings in [93] show that office workers use privacy spheres and context to decide whether personal data are sensitive.

We supplement previous research with a quantitative survey on perceived data sensitivity in the employment context, partially closing the contextual research gap. We also examine whether perceived sensitivity differs significantly between the employment context and the online and marketing contexts examined in [1, 54, 81]. This leads to the formulation of our first research question:

RQ1. Does the employment context alter the ranking of personal data by perceived data sensitivity compared to other contexts?

2.2 Willingness to disclose personal data

People's willingness to disclose personal data is an overall strong predictor of their actual disclosure behavior [11, 38]. To date, however, investigations of willingness to disclose in employment relationships are strongly limited to applicant procedures and applicants' willingness to provide truthful information to future employers [7]. Nevertheless, employees were overall found to be willing to disclose personal data to their employers, despite being aware of potential privacy invasions [3]. They do so by assessing the relevance and suitability of the requested data [76, 92, 99]. Moreover, willingness to disclose can increase if employees believe to receive adequate gratification in return [58], and their preferences for sharing (sensitive) personal data vary by region [21, 30]. Furthermore, employees' willingness to disclose can also be partially explained by Communication Privacy Management (CPM) theory [69, 86] and privacy as contextual integrity [63]. CPM describes the tension between the desire to reveal and the desire to withhold information based on ownership, control, and turbulence. Contextual integrity emphasizes on the appropriate flow of information. Here, different transmission principles apply, taking into account social norms for a particular context. We study willingness to disclose and its relationship with perceived sensitivity for different data, as described in the upcoming section.

2.3 Groups of personal data

Legislation and international standards distinguish between different groups of personal data based on their sensitivity properties. A commonly used distinction is that between "personal data" and "sensitive personal data". For example, the GDPR defines personal data as "*any information relating to an identified or identifiable natural person*" (Art. 4, GDPR), whereas sensitive personal data are defined as "*special categories of personal data*" which include any data concerning the "*racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation*" (Art. 9, GDPR). The regulation also acknowledges the sensitivity of information on criminal convictions and offenses (Art. 10, GDPR). Another commonly used definition is that of personally identifiable information (PII), which refers to a set of information that allows for mostly direct identification of a natural person (e.g., name, social security number) [32, 57]. Aside from legal distinctions, research has also made efforts to identify groups of personal data. A common approach is subdividing personal data according to perceived sensitivity using conventional clustering methods [1, 54, 81]. Such clusters differ in average perceived sensitivity but are often difficult to interpret in terms of semantic meaning. In contrast, factorization approaches revealed latent groups of personal data with increased interpretability [31, 38, 60]. In the field of employment, a study in 1973 [76] identified five factors of personal data from job interviews and found that questions about religious and ethnic, as well as financial backgrounds were perceived as inappropriate.

This paper aims to provide updated insights into employees' perceptions of groups of personal data by examining whether employees' perceived data sensitivity matches legal distinctions, and by comparing different groups of personal data based on legal and non-legal definitions in terms of employees' perceived sensitivity and willingness to disclose. Furthermore, we are the first to examine and compare the magnitude of the negative correlation between perceived sensitivity and willingness to disclose for different groups of data. As such, our research questions are as follows:

RQ2$_a$. Can latent groups of personal data be identified in the employment context based on employees' willingness to disclose and perceived data sensitivity?

RQ2$_b$. Do perceived data sensitivity and willingness to disclose differ between groups of personal data?
RQ2$_c$. Is the magnitude between perceived data sensitivity and willingness to disclose affected by the group of personal data?

2.4 Antecedents and causal model

Contextual differences in perceived sensitivity and willingness to disclose are subject to numerous influences. Many of them have been studied in scope of the "Antecedents → Privacy Concerns → Outcomes" (APCO) model to explore individuals' privacy behaviors [85]. However, the model is context dependent and different types of personal data or different domains may lead to different results and conclusions. To date, there is a lack of studies targeting the employment context [7]. We therefore review the antecedents used in [24, 48, 85] in relation to perceived sensitivity and willingness to disclose in the employment context using causal modeling. We formulate the following research question:

RQ3. In the employment context, what antecedents influence perceived sensitivity and willingness to disclose personal data, and how do they relate to each other?

Based on this question, we formulate the following hypotheses **H1** − **H5** (cf. Fig. 1):

Risk beliefs refer to the uncertainty that the disclosure of personal information could lead to some kind of material or non-material loss. Thus, *risk beliefs* negatively influence willingness to disclose [12, 53]. Since employees withhold data when they fear adverse consequences [3, 86], we hypothesize that (**H1**) high *risk beliefs* decrease employees' willingness to disclose.

Trust has an opposite effect to *risk beliefs* [10] and recent studies indicate that office workers in Germany trust their employers in the processing of their personal data [92, 93]. We thus hypothesize that high levels of *trust* in employers (**H2$_a$**) increase employees' willingness to disclose and (**H2$_b$**) decrease their *risk beliefs*.

Privacy concerns have indirect effects on privacy behavior by substantially influencing willingness to disclose [6, 11, 24]. This relationship also holds for the employment context [10]. Given that employees are required to disclose large amounts of potentially sensitive personal data to their employers, we hypothesize that (**H3$_a$**) employee concern about the extensive *collection* of personal data decreases willingness to disclose and also (**H3$_b$**) decreases overall *trust* in employers,

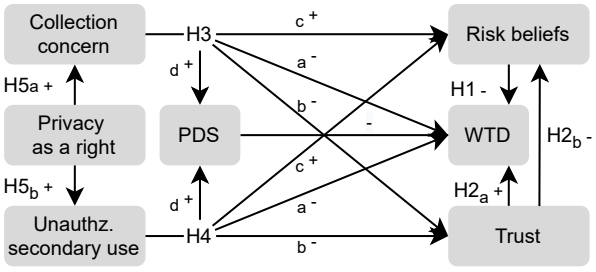

Fig. 1. Anticipated causal model of antecedents for perceived data sensitivity (PDS) and willingness to disclose (WTD) in the employment context. For each hypothesis (H1 - H5), the arrows indicate the expected direction of effect (positive (+) or negative (-)) of the seven latent constructs.

but (**H3$_c$**) increases employees' *risk beliefs*. In addition, based on findings in online privacy research [24, 48], we hypothesize that (**H3$_d$**) high levels of *collection concern* increase employees' perceived sensitivity. Moreover, since office workers have expressed concern that some data could have negative consequences if used for purposes other than those intended [92], we expect the same effects for employees' concerns about the *unauthorized secondary use* of personal information by employers (**H4$_a$**, **H4$_b$**, **H4$_c$**, **H4$_d$**, cf. Fig. 1).

Privacy as a right has been less studied than the aforementioned antecedents. Yet, people tend to perceive the right to privacy differently, which has also an effect on their privacy beliefs [4, 85, 93]. Previous work revealed that office workers' beliefs about this right influence their attitudes toward how data should be used or how much data should be disclosed [93]. We therefore hypothesize that strong beliefs in the right to privacy will lead to both (**H5$_a$**) increased *collection concerns* and (**H5$_b$**) increased concerns about the *unauthorized secondary use* of personal data.

2.5 Employee groups and clusters

Privacy research has put enormous effort in classifying people into groups. Most attempts classify people based on their privacy concerns [4, 9, 15, 17, 42, 43, 73, 80, 83, 94]. Fewer attempts are based on perceived sensitivity, willingness to disclose, and behavior [38, 41, 45, 68, 98]. Segmentations are found useful to (1) assess willingness to disclose in marketing settings [42], (2) study the impact of service features on different users [37], (3) serve developers and service providers in developing products [15], and (4) help resolve the privacy paradox [43]. Some studies have included employees [68], but so far there

are no studies that focus on the employment context [7]. We therefore take a step to close this gap and examine differences among employees based on willingness to disclose and perceived sensitivity. We choose these attributes, because we believe they are the most relevant for employers when attempting to process truthful data.

RQ4$_a$. Can employees be classified into groups according to willingness to disclose and perceived sensitivity?
RQ4$_b$. Do these groups differ in terms of demographic factors or privacy attitudes?

3 Methodology

To examine our research questions and hypotheses, we conducted a cross-sectional online survey with 553 employees in Germany between July 2020 and March 2021. The data were analyzed quantitatively using appropriate statistical methods. In what follows, we provide details on ethical considerations, the measurement instrument used, the survey's procedure, the participants' recruitment and demographics, and the data analysis.

3.1 Ethics

Our institutions do not have a formal IRB process, yet we ensured to minimize potential harms from our study by adhering to the Code of Ethics of the German Sociological Association and the Standards of Good Scientific Practice of the German Research Foundation. Our study design was also independently approved by two data protection officers at our institutions. Employees participating in our study were informed about the data collected at the beginning of the survey. After consenting to participate, they could leave the survey at any time and delete their responses. In addition, we collected data anonymously whenever possible. If this was not possible, the data were stored separately from the response data and deleted after the survey was completed. All data were stored on encrypted hard drives.

Participants recruited through online panels were paid according to minimum wage in Germany (€9.60/h) adjusted to the median completion time. Participants recruited via other channels were not paid but invited to participate in a raffle of shopping vouchers. We pointed out the conditions of participation at the beginning of the study. When we contacted organizations to recruit their employees, we provided extensive information about the study and surveys for review. We assured participating organizations that they could not be identified. One organization required approval through employee representation. We assured representatives and employees that we would not share information about participation with their respective employers. Last but not least, we explicitly referred to voluntary participation in our invitation emails and, after consultation with the organizations, explained whether the study may be completed during or outside working hours.

3.2 Measurement instrument

We used validated measurement items from the literature to design our survey and adapted them as needed. For privacy antecedents, we used items from [53, 84] to elicit *trust*, *risk beliefs*, *collection concern*, and *unauthorized secondary use*, as well as items from [4] to elicit *privacy as a right*. All antecedents were measured with three to four items using a six-point scale. To measure perceived sensitivity and willingness to disclose, we used a set of 62 items representing various personal data. Participants rated sensitivity on a six-point scale and willingness to disclose on a four-point scale, respectively. The set of 62 personal data items is composed of the results of a series of workshops conducted in 2019 as part of the preparation of this study. The workshops targeted at eliciting employees' requirements for privacy-enhancing tools. In the workshops, participants were asked to list personal data that are frequently disclosed at work or that they believe should be protected. Details on the workshops are available in [71]. For our survey, we combined the responses from four workshops with a total of 30 participants from four research institutions and one private company in Germany. Workshop participants included works councils, administrative staff, IT professionals, and researchers from the fields of ergonomics, data protection law, and human-computer interaction. From the responses, we created a consolidated list of personal data with 50 unique items. Given the expertise of our workshop participants, we consider the list to reflect a fair representation of personal data relevant for the purpose of our study. To address potential bias through participant recruitment, we have completed the list with items from studies on privacy in the online and marketing context [54, 81]. Some items were omitted, if they have a different meaning in the German-cultural space or if no equivalent exists. The full questionnaire is available in Appendix A. The final list of personal data items is available in Fig. 2 and in Appendix D.

3.3 Participants and procedure

Our survey requires participants to rate a total of 133 items. As a result, the length of the questionnaire and the associated workload may influence employees' willingness to participate, leading to fatigue near the end of the survey, and increasing the risk of unbalanced responses [22]. We have therefore created a two-part questionnaire to make the survey more appealing to employees, easier to complete, and to avoid quality loss due to excessive and repetitive question design. The first part (Part 1) is composed of three sub-parts: (1) demographics related to employment, (2) ratings of perceived sensitivity and willingness to disclose 62 personal data items, and (3) remaining demographics and survey feedback for part one. The second part (Part 2) comprises questions on the variables of our causal model and survey feedback for part two. We recruited our participants via the two online panels Prolific (N = 351) and Respondi (N = 111), as well as via mailing lists of organizations we contacted (convenience sampling), and through social media of the local Chamber of Commerce and Industry (N = 133) (N_{total} = 595). The reason for distributing the survey across multiple channels was to reach a larger number of participants and to reduce demographic bias from individual channels, as response rates via Prolific were low for some demographic groups. First, we invited participants to complete Part 1, and then reinvited them to Part 2 two days later. To avoid methodical artifacts, we screened participants to ensure that they were employed in Germany and spoke German. After survey completion, we linked the responses from both parts by merging the data from the surveys. For the online panels, we used user IDs provided by the panels. For all other recruitment channels, we used passcodes generated by the participants themselves. Passcodes were created in the first survey and had to be re-entered in the second survey. Neither the user IDs nor the passcodes allow us to identify the natural persons. Furthermore, we have removed participants from the data based on timing, the number of missing responses ($\geq 10\%$), and participants' self-assessed quality of the responses, consisting of ratings for honesty and seriousness. We additionally checked the data for multivariate outliers and straightlining response patterns.

In total, we have accepted 553 responses as valid for Part 1, and 393 responses for Part 2. Response times averaged 11.7 minutes (median = 9.8) for Part 1 and 12.8 minutes (median = 11.6) for Part 2. The sample demographics are summarized in Table 1. Overall, our sample is slightly biased in favor of younger male participants as there is a small positive correlation between sex and age (ρ = .17, CI_{95}: [.08, .25]). Nevertheless, participants' ages spanned the typical period of employment ($x \in [18, 67]$, $\bar{x}$ = 39.6, sd = 12.3). At the time of the survey, half the respondents had been employed by their current employer for at least six years ($x \in [0, 46]$, $\bar{x}$ = 8.77, sd = 9.5). Three-quarters had permanent employment. In addition, our sample includes employees from 18 different industries and 12 different occupational groups. The distribution of industries among the top five industries was balanced, but a bias toward the service sector was observed among professional groups. Compared to the overall population of employees in Germany [67], however, our sample is biased toward younger employees with a university degree who work for large organizations, have a slightly shorter job tenure, and higher income. Our participants also primarily worked in the fields of IT, science, business, law, and education. Details on the industries and professional groups, and a separate presentation of the demographics of Part 1 and Part 2, as well as a comparison with the population of employees in Germany, can be found in Appendix C.

Table 1. Participant demographics summary.

Demographic variables			**N = 553**
Age (years)	%	Net income (€ / month)	%
≤ 24	8.7	< 1k	9.2
25 – 34	32.4	1k < 2k	36.7
35 – 44	27.1	2k < 3k	36.9
45 – 54	14.6	3k < 4k	11.4
55 – 64	16.5	≥ 4k	5.8
≥ 65	.7		
Sex	%	Job tenure (years)	%
Diverse	.2	≤ 4	47.3
Female	39.6	5 – 9	24.1
Male	59.7	≥ 10	28.6
Org. size (num. employees)	%	Other	%
< 10	8.0	German nationality	88.2
10 – 249	34.4	University degree	58.2
250 – 999	25.7	Permanent employment	75.8
≥ 1k	31.6	Multiple jobs	7.6

3.4 Analysis

To answer **RQ1** and **RQ2**, we analyzed the data from Part 1 because it contained responses about perceived sensitivity and willingness to disclose. To test the hypotheses **H1** − **H5** under **RQ3** and to investigate **RQ4**, we used the subsample of Part 2, as it contained responses to the latent constructs (i.e., privacy beliefs). All analyses were performed using R (see Appendix B).

3.4.1 Comparison between contexts

To compare results for perceived sensitivity between contexts [1, 54, 81], we have created an intersection of items from all studies and ranked the items according to their mean scores per study. We then compared the pairwise Spearman rank correlation coefficients (ρ) between all studies to verify whether the ranking of the items remain constant. Next, we examined whether the pairwise correlations between our German sample (employment context) and the German sample in [81] (online context) differed by running tests for differences in overlapping correlations. Significance was determined using the percentile bootstrap method of Rousselet et al. [79] at the 95% confidence interval (CI_{95}, $n_{boot} = 2000$).

3.4.2 Latent structure analysis

To identify groups of personal data based on participants' assessments of their willingness to disclose, we conducted an Exploratory Factor Analysis (EFA). We also ran a Confirmatory Factor Analysis (CFA) on willingness to disclose and perceived sensitivity to validate the identified structure. We chose common factor analysis over principal component analysis, because research suggests that willingness to disclose and perceived sensitivity are influenced by latent variables, such as contextual norms [56, 63]. For analysis, we followed guidelines for EFA and CFA with ordinal data. [26, 97, 102]:

First, we removed personal data items with nonresponse rates $\geq$10% [26] and tested for univariate and multivariate normality (UVN, MVN) to assess the suitability of the data for further analysis. Next, we examined the possibility of imputing missing data by testing for missing patterns using Little's test. We also visually examined the data if we expected biased results due to violations of the MVN assumption. To conduct EFA and CFA on different datasets, we split the data in half at random ($N_{EFA} = 277$, $N_{CFA} = 276$) and verified that the demographic properties were similar. Next, we examined the basic factorability assumption using the Kaiser–Meyer–Olkin measure of sampling adequacy, the Kaiser–Meyer–Olkin criterion, and Bartlett's test of sphericity. To account for the ordinal nature of the data, we used polychoric correlations [19, 29]. We then removed items with high ($|r| \geq .8$) or very low ($|r| < .3$) pairwise correlations. The number of factors to retain was determined on the basis of multiple recommended criteria [72, 102]: Parallel analysis, Velicer's Minimum Average Partial, and post-hoc model fit indexes, i.e., the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC).

The estimators for EFA were selected based on recommendations to recover weak factors in rather small samples or when MVN is violated [97, 102], including Minimum Residuals, Unweighted Least Squares, and Principal Axis Factoring. We then compared the resulting solutions to each other, in order to ensure that the results replicated for the different estimators [97]. We used oblique rotation to address the expected correlations between emerging factors. After deciding on a factor solution, we have refined it iteratively using Hair et al. [26]'s three-step procedure.

The latent factors identified from EFA were validated with CFAs using the robust estimator WLSMV [36]. First, we fitted a model to the EFA-subsample to detect severe model misspecification [36]. We then fitted a second model for the CFA-subsample to verify the latent structures' validity and reliability for both willingness to disclose and perceived sensitivity. Discriminant validity was validated using the Fornell-Larcker criteria and the heterotrait-monotrait ratio of correlation (HTMT) ($< .85$) [28].

3.4.3 Comparison of groups of personal data

To compare groups of personal data, we have created four groups according to the distinctions made in legal texts and standards discussed in Section 2.3: (1) The group *ALL* includes all 62 personal data items surveyed; (2) The group *GDPR* represents special categories of personal data under Arts. 9 & 10 GDPR [18]; (3) The group *IDENT* represents secure personal identifiers (e.g., Passport No.) [57]; and (4) the group *MASTER* refers to employee master data (e.g., contact details). The detailed item mapping is available in Appendix D. For each group, we have created sub-scales for perceived sensitivity (PDS) and willingness to disclose (WTD), and run regression analysis to compare the different groups. PDS_{ALL} and WTD_{ALL} served as the baseline for comparisons. Violations of independence for the dependent variables were addressed by including random intercepts for participants and random slopes for perceived sensitivity in Linear Mixed-effects Models (LMMs) and its robust variants. We have verified that the inclusion of random effects increased the model fit using Likelihood Ratio Tests. Verification of normality and homoscedasticity assumptions for residuals failed by visual inspection and using Levene test. All models were therefore fitted using robust LMMs. Significance

checks were done using the robust LMM's t value and the Satterthwaite approximations [50] of degrees of freedom of the corresponding regular LMM (cf. [23, 101]).

3.4.4 Causal model analysis of antecedents

We analyzed the causal model using structural equation modeling (SEM). Based on expected effect sizes in the range $[.2 \leq |\beta| \leq .85]$ [24] and based on common rules of thumb, we decided that $N_{obs} = 393$ was acceptable ($N_{obs} \geq 300$, $N_{obs}/N_{var} \geq 10$, $N_{obs}/N_{params} > 5$ [26, 96]). We further assessed the validity of the measurement model and structural model as well as the constructs' reliability following guidelines in [26]. Privacy beliefs were modeled as reflective constructs, whereas we used composite scores for perceived sensitivity and willingness to disclose. If fit indices indicated inadequate fit, we ran an EFA with Principal Axis Factoring and oblimin rotation to identify items with high crossloadings and marked them for deletion. We then modelled the causal SEM structure *a-priori* based on our hypothesized model (cf. Fig. 1). We have also included our participants' demographic data as control variables.

3.4.5 Employee groups

We performed Latent Class Analysis (LCA) to identify groups of employees based on their response patterns of willingness to disclose. LCA is a type of finite mixture modelling that determines classes ("clusters") based on subpopulations with different sets of attributes. Observations are assigned probabilities belonging to each class. Here, classes are assumed to be unobserved categorical (latent) variables. We determined the optimal number of classes by first estimating a one-class model and then iteratively adding classes up to a maximum of five, as we expected group sizes similar to those in previous studies in other contexts [4, 38, 42, 80]. We evaluated model fit using various fit indices, with a focus on BIC due to its superiority in LCA class selection [64]. To avoid local maxima, we ran 500 replications.

We have fixed participants' class memberships based on posterior probabilities after deciding on the number of classes. To improve the posterior probabilities and reduce estimation attenuation [8], we ran latent class regression analysis with demographic covariates as dichotomous and attitudes as ordinal (3 bins) variables. Before extracting the classes, we ensured that the entropy was greater than .8, indicating a low classification error. We then compared the fit of a constraint to the fit of an unconstrained multigroup SEM to test for differences between the extracted classes. If the Likelihood Ratio Test was significant, we performed distal outcome analysis for privacy antecedents using a Multiple Indicators and Multiple Causes (MIMIC) model and logistic regression for demographic variables.

4 Results

In this section, we present our results according to our different analysis steps outlined above.

4.1 Descriptive results

The average scores for perceived data sensitivity (PDS) and willingness to disclose (WTD) are plotted in Fig. 2. Detailed scores are available in Appendix D.

Consistent with previous work, *passwords* were perceived as the most sensitive and *hair color* as the least sensitive data types. It is striking that eight of the ten items with the lowest PDS can be clearly assigned to employee master data. The ratio of personal data types with PDS < 5 and data types with PDS > 5 is 21:41. This means that two-thirds of the data types were rated as rather sensitive information. Half of the data types have PDS ≥ 6. The proportion of data types with WTD < 5 and WTD > 5 is 32:30, and is therefore balanced. The willingness to disclose personal data was the lowest for *online dating activities*, closely followed by *passwords* and *DNA*. On the other hand, participants were most willing to disclose their *profession*, *education*, and *language skills* to employers. The ten items with the highest WTD are all directly related to employment.

Furthermore, visually comparing the scores from this study with scores from previous studies conducted in the online [1, 81] and marketing [54] contexts reveals differences in the rating of perceived sensitivity (cf Fig. 2). With a few exceptions, the scores for less sensitive data are lower in this study than in other studies, whereas scores for sensitive data are almost always higher in this study. A detailed comparison of perceived sensitivity between the different studies is further summarized in Fig. 3. The scatter plots show a clear linear relationship between the ranked PDS scores across all studies except for our study. This implies that personal data items were ranked similarly in previous studies despite the different contexts. This is also con-

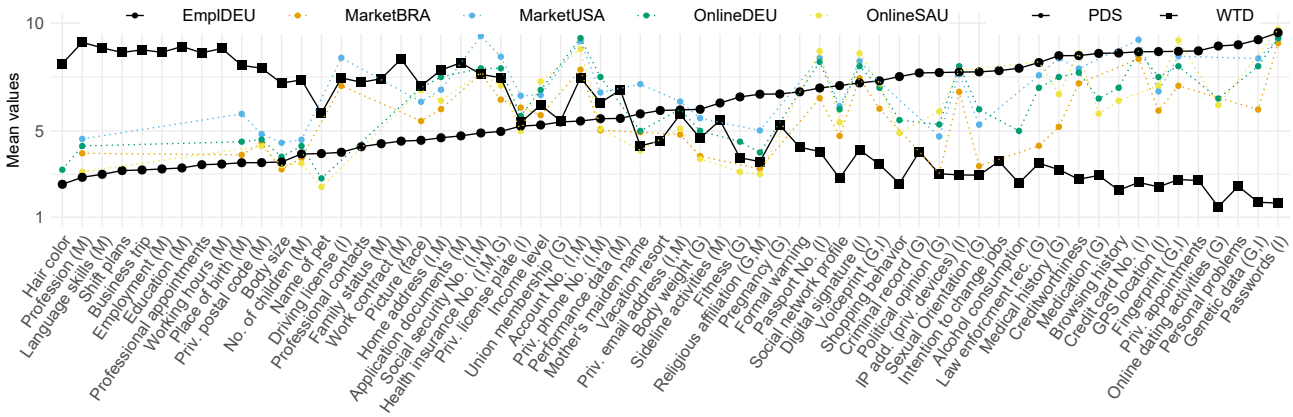

Fig. 2. Mean values for perceived data sensitivity (PDS) (dots) and willingness to disclose (WTD) (square) of 62 personal data items, sorted by PDS (this study, EmplDEU). Adjusted PDS scores from studies in other contexts and cultural backgrounds are included for comparison (MarketBRA [54], OnlineDEU [81], OnlineSAU [1], MarketUSA [54]). Missing assignments indicate that the item was not surveyed in the corresponding study. "M" marks employee master data, "I" marks secure identifiers, and "G" marks data under GDPR.

firmed by the high correlation coefficients. Here, the scores determined in [81] for the German sample have the highest correlation coefficients of all studies conducted. In contrast, the ranking of PDS scores in our study correlates only moderately to weakly with the ranking in previous studies. The pairwise comparisons further confirm that the differences between the correlation coefficients of our study and those of Schomakers et al. [81] are significant in all cases (cf. Fig. 3).

Summary: Referring to **RQ1**, our analysis revealed that perceived data sensitivity varied more between the employment context and other contexts than between online and marketing or cultural contexts.

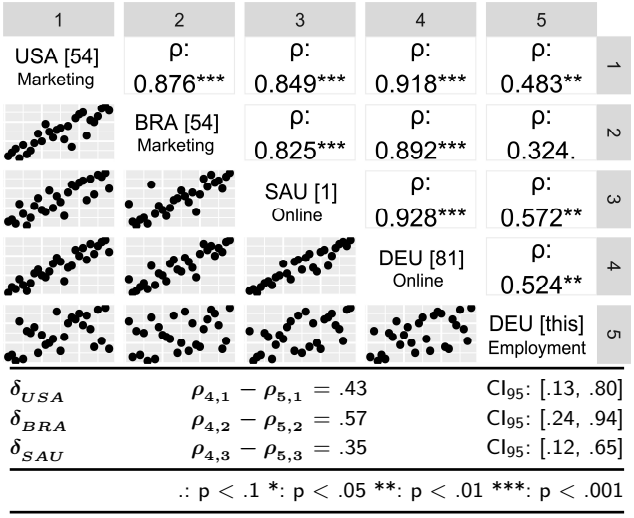

Fig. 3. Pairwise rank correlations (Spearman) of perceived data sensitivity investigated in different studies, including pairwise comparisons between the two studies with samples from Germany.

4.2 Latent groups of personal data

Our tests indicated violation of the normality assumption. Based on insignificant results of Little's test ($\chi^2(2196) = 1784.212$, n. s.) and visual inspection of the data, we concluded that data were missing completely at random. We have therefore imputed missing data using the non-parametric method *missForest* suitable for ordinal data [89]. Furthermore, the basic factorability assumption was confirmed by all items having acceptable values for the Kaiser–Meyer–Olkin measure of sampling adequacy ($\geq .85$) and by the Kaiser–Meyer–Olkin criterion ($\geq .91$) indicating "meritorious" factorability of the correlation matrix. The Bartlett's test of sphericity was also significant ($\chi^2(1830) = 8394.02$, $p < .001$), implying that the correlation matrix was appropriate for factor analysis. Factor retention criteria suggested retaining between three and six factors, which is consistent with the range of dimensions proposed in previous work [31, 38, 54, 60, 61, 76, 81]. Due to the high number of items, the skewed data, and the sample size, we focused particularly on avoiding bias towards overfactoring [95]. After comparing different solutions, a four factor solution using Principal Axis Factoring and Promax rotation achieved the best partitioning in terms of acceptable loading height ($> .45$), low number of cross-loadings (relative magnitude of variance [26]), acceptable commonality ($\geq .5$), and interpretability of the factors. Iterative refinement resulted in a set of 18 items. The final CFAs had good to acceptable model fits, and indicators for construct reliability were in acceptable range. Discriminant validity has also been confirmed. The results of the analysis are reported in Table 2.

Table 2. Latent groups of personal data and results of CFA.

	WTD			PDS		
Model fit						
Scaled fit indices						
χ^2(df), ***: p < .001	(129): 207.9***			(98): 188.9***		
CFI	.98			.99		
GFI	.99			.99		
RMSEA	.05			.06		
Recommended values [26]:	CFI > .94,	GFI > .95,	RMSEA < .7			
Identified latent constructs and their items						
SENS	λ			λ		
Genetic data	.87	α	.83	.59	α	.81
Personal problems	.71	ω	.84	.68	ω	.82
GPS location	.70	AVE	.58		AVE	.58
Medication	.73			.83		
Creditworthiness	.74			.77		
Medical history	.82			.90		
NOTSENS	λ			λ		
Hair color	.82	α	.74	.90	α	.70
Body size	.83	ω	.77	.74	ω	.73
Body weight	.73	AVE	.63		AVE	.68
PII	λ			λ		
Home address	.80	α	.81	.87	α	.90
Social security No.	.80	ω	.82	.91	ω	.91
Health insurance No.	.82	AVE	.63	.89	AVE	.78
Account number	.77			.86		
WORK	λ			λ		
Employment	.87	α	.79	.86	α	.91
Profession	.85	ω	.81	.85	ω	.92
Professional appointments	.72	AVE	.61	.86	AVE	.76
Shift plans	.77			.89		
Business trip	.67			.90		
Recommended values [26]:	$\lambda \geq .7$,	$\alpha \geq .7$,	$\omega \geq .7$,	AVE $\geq .5$		

Summary: With respect to $\mathbf{RQ2_a}$, using factor analysis, we identified four latent groups of personal data. The data groups emerged directly from our participants' response patterns and represent distinct dimensions of employees' perceived sensitivity and willingness to disclose personal data in the employment relationship.

4.3 Comparison groups of personal data

We compared the relationship between willingness to disclose and perceived sensitivity for eight groups of personal: the four predefined groups and the four latent groups. To better distinguish the latent groups of personal data in further analysis, we have assigned them names to reflect the groups' characteristics of sensitivity and context. A summary of all the data groups examined is provided in Table 3. In the regression analysis performed, we included random effects, as this significantly increased the fit of the LMMs (Fit_{PDS}: $\chi^2(1) = 462.28$, $p < .001$; Fit_{WTD}: $\chi^2(2) = 457.5$, $p < .001$). The results of regression analysis are reported in Table 4 and Fig. 4, respectively.

We found that all groups of personal data were perceived as having significantly different levels of sensitivity. Employees' willingness to disclose also differed significantly between the eight groups. For one thing, the assessed scores deviated significantly from the baseline (*ALL*). Second, Tukey post-hoc analysis further revealed that perceived sensitivity and willingness to disclose differed also significantly among all data groups ($p < .001$), except for *MASTER* and *PII*, between which no significant difference was found. Furthermore, data groups clearly related to employment (*PII*, *MASTER*, *WORK*) had significantly lower perceived sensitivity and higher willingness to disclose, whereas this effect was reversed for all other data groups. This finding confirms the context-dependence of willingness to disclose. A notable exception is the group *NOTSENS*, which behaved like employment-related data groups. Contrary to intuition, perceived data sensitivity (PDS) was even much lower compared to PDS_{PII} and PDS_{MASTER}, despite being unrelated to the employment context (cf. Table 4). This highlights that *NOTSENS* represents personal data of very low sensitivity. In contrast, however, willingness to disclose (WTD) was also significantly lower compared to WTD_{PII} and WTD_{MASTER}. This indicates that significantly lower perceived sensitivity does not imply significantly higher willingness to disclose. Regarding the sensitive data groups, we note that the latent group *SENS* reflects data considerably

Table 3. Groups of personal data examined.

Data group	Set of personal data items contained
Predefined groups (cf. Appendix D)	
ALL	All 62 items
GDPR	Special categories under Arts. 9 & 10 GDPR [18]
IDENT	Secure personal identifiers (e.g., Passport No.) [57]
MASTER	Employee master data (e.g., contact details)
Latent groups (cf.Table 2 or Appendix D)	
SENS	Sensitive data types from the private sphere not related to employment and potentially harmful
NOTSENS	Least sensitive data types from the private sphere and not related to employment
PII	Personal and secure identifiers, usually disclosed to employers in Germany
WORK	Data types arising directly from employment

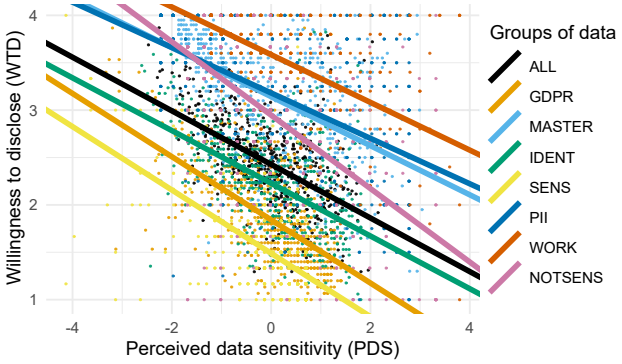

Fig. 4. Robust LMM's fixed effects for different groups of data.

more sensitive than the group *GDPR*. Indeed, the effects of perceived sensitivity and willingness to disclose were almost twice as strong as those of *GDPR*.

Examining the magnitude between perceived sensitivity and willingness to disclose, we found that perceived sensitivity had a notable significant negative effect on willingness to disclose for all data groups. However, visual inspection of the regression lines (cf. Fig. 4) reveals that this magnitude is significantly steeper for the *NOTSENS* group. Tukey post-hoc analysis confirmed this observation ($p < .001$). The analysis also revealed that the magnitudes for the groups *GDPR* and *SENS* are steeper than for *PII* (-.09, CI_{95}:[-.15, -.03]).

Summary: Referencing $\mathbf{RQ2_{b\text{-}c}}$, our results suggest that employees perceive different groups of personal data as having different levels of sensitivity, and that their willingness to disclose also differs significantly by group. Data more related to the employment context had significantly lower perceived sensitivity as well as higher willingness to disclose. Moreover, the magnitude between perceived sensitivity and willingness to disclose appears to be largely constant, except for personal data with low perceived sensitivity.

Table 4. Results of robust LMMs for perceived data sensitivity (PDS) and willingness to disclose (WTD) with random effects by participants. Data groups are compared with ALL as a baseline.

	PDS		WTD	
Predictors	Est.	CI_{95}	Est.	CI_{95}
(Intercept)	3.78***	3.69, 3.88	2.43***	2.43, 2.46
GDPR	.84***	.72, .96	-.59***	-.63, -.54
MASTER	-.85***	-.97, -.73	.72***	.67, .76
IDENT	.43***	.31, .55	-.20***	-.24, -.16
SENS	1.58***	1.46, 1.70	-.94***	-.98, -.89
PII	-.68***	-.80, -.56	.75***	.71, .80
WORK	-1.68***	-1.80, -1.55	1.16***	1.11, 1.20
NOTSENS	-1.18***	-1.30, -1.06	.52***	.48, .56
PDS[c]			-.28***	-.32, -.24
NOTSENS x PDS[c]			-.10***	-.15, -.06
R_m^2 / R_c^2	.458 / .586		.749 / .843	

Note. N = 553 ***: $p < .001$
R_m^2: marginal; R_c^2: conditional [c]centered

4.4 Antecedents and causal model

An initial CFA of the measurement model indicated overall adequate fit and an EFA revealed clearly emerging factors. However, we removed an item for *risk beliefs* that cross-loaded onto *trust*, as well as an item for *collection concern* with $\rho > .9$ on multiple items of *risk beliefs*. While this relaxed the variance shared between the constructs, their correlation remained strong. Nevertheless, the adjusted measurement model had good fit ($\chi^2(172) = 292$, $p < .001$, CFI = .99, GFI = .99, RMSEA = .043), and indicators for construct reliability and validity were in acceptable range (cf. Table 5). In addition, all subsequent structural equation modeling (SEM) analyses also showed adequate model fit. The detailed analysis results for the different groups of personal data are reported in Table 6.

First, SEM analysis has confirmed significant moderate negative effects of perceived sensitivity on willingness to disclose for all groups of personal data. In contrast, our hypotheses regarding the effects of antecedents on willingness to disclose were confirmed only for some groups but not for others. For the anticipated positive effect of *trust* on willingness to disclose, we found a small ($|\beta| \leq .3$) significant effect for the group *ALL* and a small significant effect for the group *GDPR*. We also found a small significant negative effect of *collection concern* on willingness to disclose for the group *SENS*. Regarding effects on perceived sensitivity, we found small significant positive effects of *collection concern* on PDS for the four data groups *ALL*, *GDPR*, *IDENT* and *MASTER*. There were also small significant effects of *unauthorized secondary use* on PDS for the data groups *ALL*, *GDPR*, and *IDENT*.

Regarding the relationships between antecedents, we found support for most anticipated effects. First, *privacy as a right* had significant moderate ($.3 \leq |\beta| \leq .5$) positive effects on both *collection concerns* and *unauthorized secondary use*. This means that employees with strong convictions about a right to privacy had significantly more privacy concerns. Furthermore, *trust* in employers had a significant small negative effect on *risk beliefs*, while *collection concern* had a strong ($|\beta| > .5$)

Table 5. Construct reliability measures, validity measure, and correlations.

Construct		$\overline{x}$	sd	α^a	ω^a	1.	2.	3.	4.	5	λ^a (final selection only)				Sources
1. Collection concern	COLL	2.57	1.24	.82	.83	**.68**	.62^u	.47^u	.01^u	.19^u	.835	.774	.857		[27, 53]
2. Risk beliefs	RSKB	2.04	1.01	.75	.79	.79^l	**.63**	.5^u	.0^u	.07^u	.745	.852	.779		[27, 53]
3. Trust	TRST	5.07	.91	.88	.89	-.69^l	-.7^l	**.75**	0^u	.04^u	.916	.756	.890	.886	[27, 53]
4. Unauthz. sec. use	UNAU	5.39	.80	.74	.77	.08^l	.0^l	.04^l	**.63**	.22^u	.810	.795	.772		[27, 53]
5. Privacy as a right	PRGT	4.08	1.18	.75	.76	.43^l	.26^l	-.2^l	.46^l	**.56**	.809	.663	.768		[4]

Note. N = 393, ^{l}: inter-construct correlation (ρ), bold: AVE, ^{u}: ρ^2, Discriminant validity (Fornell-Larcker) requires $\rho^2 <$ AVE,
^{a}: Recommended values [26]: $\lambda \geq .7$, $\alpha \geq .7$, $\omega \geq .7$, AVE $\geq .5$

positive effect on *risk beliefs*. Moreover, *collection concern* also had a strong negative effect on *trust*. In other words, employees who were concerned about their employer's collection of personal data had significantly less trust and also anticipated greater privacy risks. Concerning *unauthorized secondary use*, no anticipated effect was confirmed. Contrary to our expectation, *unauthorized secondary use* even had a significant small positive effect on *trust*, rather than a negative effect.

With respect to demographic differences, we found very few and only small significant effects. For perceived sensitivity and willingness to disclose, we found

Table 6. Results SEM analysis for eight different groups of personal data.

Regressions antecedents (ANT) on willingness to disclose (WTD) and perceived data sensitivity (PDS): ANT → {WTD, PDS}

	ALL			GDPR			IDENT			MASTER		
Model fit	χ^2 : 293.22	CFI	: .99	χ^2 : 293.37	CFI	: .99	χ^2 : 296.50	CFI	: .99	χ^2 : 296.56	CFI	: .99
	df : 251	GFI	: .99	df : 251	GFI	: .99	df : 251	GFI	: .99	df : 251	GFI	: .99
	p : .03	RMSEA	: .02	p : .03	RMSEA	: .02	p : .03	RMSEA	: .02	p : .03	RMSEA	: .02
Hypothesized effect	Est.	CI_{95}	β	Est.	CI_{95}	β	Est.	CI_{95}	β	Est.	CI_{95}	β
PDS $-\rightarrow$ WTD	-.49	[-.58, -.40]	-.43***	-.62	[-.71, -.53]	-.52***	-.49	[-.58, -.39]	-.41***	-.45	[-.56, -.35]	-.40***
H1 RSKB $-\rightarrow$ WTD	.10	[-.06, .26]	.14	.14	[-.03, .30]	.18	.09	[-.07, .25]	.12	.03	[-.12, .19]	.05
$H2_a$ TRST $+\rightarrow$ WTD	.16	[.01, .31]	.18*	.26	[.10, .43]	.28***	.13	[-.02, .27]	.15	.05	[-.08, .19]	.06
$H3_a$ COLL $-\rightarrow$ WTD	-.24	[-.50, .03]	-.21	-.13	[-.42, .17]	-.10	-.24	[-.51, .02]	-.22	-.23	[-.47, .01]	-.21
$H3_d$ COLL $+\rightarrow$ PDS	.25	[.14, .35]	.25***	.22	[.12, .33]	.22***	.20	[.10, .31]	.21***	.23	[.12, .34]	.24***
$H4_a$ UNAU $-\rightarrow$ WTD	-.10	[-.24, .04]	-.09	-.08	[-.22, .06]	-.07	-.14	[-.29, .00]	-.13	-.06	[-.19, .07]	-.06
$H4_d$ UNAU $+\rightarrow$ PDS	.22	[.11, .33]	.23***	.24	[.14, .34]	.24***	.17	[.07, .28]	.19***	.08	[-.02, .19]	.09

	NOTSENS^L			PII^L			SENS^L			WORK^L		
Model fit	χ^2 : 291.52	CFI	: .99	χ^2 : 294.81	CFI	: .99	χ^2 : 286.13	CFI	: .99	χ^2 : 287.74	CFI	: .99
	df : 251	GFI	: .99	df : 251	GFI	: .99	df : 251	GFI	: .99	df : 251	GFI	: .99
	p : .04	RMSEA	: .02	p : .03	RMSEA	: .02	p : .06	RMSEA	: .02	p : .06	RMSEA	: .02
Hypothesized effect	Est.	CI_{95}	β	Est.	CI_{95}	β	Est.	CI_{95}	β	Est.	CI_{95}	β
PDS $-\rightarrow$ WTD	-.56	[-.65, -.47]	-.49***	-.43	[-.55, -.32]	-.40***	-.58	[-.65, -.51]	-.50***	-.44	[-.53, -.35]	-.40***
H1 RSKB $-\rightarrow$ WTD	.04	[-.13, .21]	.06	.04	[-.13, .20]	.06	.09	[-.09, .26]	.13	-.05	[-.22, .12]	-.08
$H2_a$ TRST $+\rightarrow$ WTD	.05	[-.11, .22]	.07	.03	[-.11, .17]	.03	-.09	[-.24, .06]	-.11	.00	[-.16, .16]	.00
$H3_a$ COLL $-\rightarrow$ WTD	-.01	[-.31, .29]	-.01	-.03	[-.30, .23]	-.03	-.31	[-.61, -.02]	-.29*	.09	[-.17, .34]	.09
$H3_d$ COLL $+\rightarrow$ PDS	-.01	[-.11, .09]	-.01	-.04	[-.14, .07]	-.04	-.05	[-.15, .06]	-.05	-.02	[-.13, .08]	-.02
$H4_a$ UNAU $-\rightarrow$ WTD	-.06	[-.19, .08]	-.05	.08	[-.05, .20]	.08	.10	[-.04, .24]	.09	-.01	[-.15, .12]	-.01
$H4_d$ UNAU $+\rightarrow$ PDS	-.01	[-.12, .10]	-.01	-.02	[-.12, .09]	-.02	-.06	[-.16, .03]	-.07	-.03	[-.14, .08]	-.03

Regressions between antecedents (ANT) valid for all groups of personal data depicted above: ANT → ANT

Hypothesized effect	Est.	CI_{95}	β	Hypothesized effect	Est.	CI_{95}	β
$H2_b$ TRST $-\rightarrow$ RSKB	-.35	[-.49, -.22]	-.28***				
$H3_c$ COLL $+\rightarrow$ RSKB	.99	[.74, 1.23]	.61***	$H3_b$ COLL $-\rightarrow$ TRST	-.93	[-1.1, -.76]	-.72***
$H4_c$ UNAU $+\rightarrow$ RSKB	-.17	[-.35, .00]	-.11	$H4_b$ UNAU $-\rightarrow$ TRST	.22	[.07, .37]	.18**
$H5_a$ PRGT $+\rightarrow$ COLL	.43	[.31, .56]	.40***	$H5_b$ PRGT $+\rightarrow$ UNAU	.54	[.37, .71]	.48***

Note. N = 393, β: standardized path coefficient (measure of effect size [16]), ^{L}: Latent groups *: p < .05 **: p < .01 ***: p < .001

significant effects for sex, nationality, company size, education, and job tenure. In addition, age and industry had small effects on *unauthorized secondary use*. Also, perceived *trust* differed significantly by industry. However, because the effects are small and vary greatly by data group and antecedent, no clear conclusions can be drawn. The details are therefore reported in Appendix E and Appendix F, respectively.

Summary: Our analysis has partially confirmed hypotheses $\mathbf{H2_{a\text{-}b}}$, $\mathbf{H3_{a\text{-}d}}$, $\mathbf{H4_{d}}$, and $\mathbf{H5_{a\text{-}b}}$. In contrast, there was no support for hypotheses **H1** and $\mathbf{H4_{a\text{-}c}}$. Moreover, the antecedents' effects on perceived sensitivity and willingness to disclose were largely inconsistent across the groups of data and were especially insignificant for latent groups of data. Furthermore, demographics had mostly negligible, if any, effects. As a result, effects of privacy antecedents in the employment context also appear to differ from other contexts.

4.5 LCA and clusters of employees

The fit indices of the five repeated LCAs indicated that a three-class solution was the best model. Because entropy was greater than .8, we have fixed the class membership of participants and assigned them to clusters. 85% of participants were assigned to one of these groups with a probability of $\geq 90\%$. When comparing the restricted SEM to the unrestricted SEM, the fit decreased significantly ($\Delta(\chi^2) = 946.03$, $p < .001$), indicating that the clusters were significantly different. An overview of the clusters is provided in Fig. 5. This reveals a clear low-medium-high cluster structure:

Cluster 1 ($N = 174$) is the mid-cluster. Here, willingness to disclose follows the anticipated order among the latent data groups based on context and sensitivity. Tukey post-hoc analysis revealed significant differences between all latent groups for both willingness to disclose (WTD) and perceived sensitivity (PDS) ($p < .001$).

Cluster 2 is the smallest ($N = 74$). It represents the low-cluster with an overall low willingness to disclose across all latent groups of data. Tukey post-hoc analysis revealed that neither WTD_{PII} and $\text{WTD}_{NOTSENS}$ nor $\text{PDS}_{NOTSENS}$ and PDS_{WORK} differed significantly ($p = .72$, $p = .99$). In conclusion, employees in this cluster only distinguished between three levels of perceived data sensitivity and willingness to disclose.

Cluster 3 ($N = 145$) represents the high-cluster. Tukey post-hoc analysis revealed that willingness to disclose and perceived sensitivity did not differ for *NOTSENS* and *PII*. However, while $\text{WTD}_{NOTSENS}$ was lower than WTD_{WORK} ($p < .001$), perceived sensitivity did not differ ($p = 0.16$). This effect was reversed for *PII*: willingness to disclose did not differ whereas PDS_{PII} was higher than PDS_{WORK} ($p < .001$). Employees thus only distinguished between two levels of willingness to disclose and perceived sensitivity.

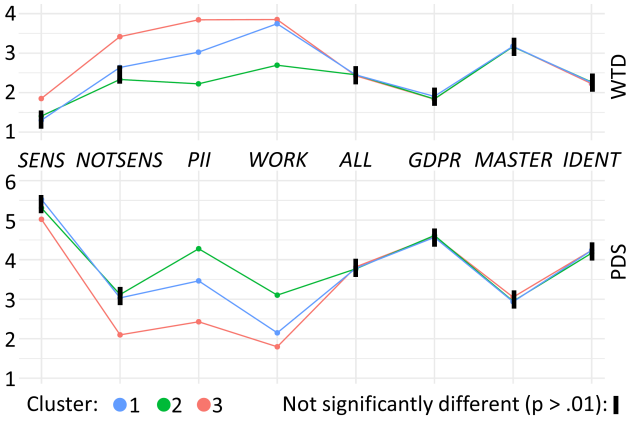

Fig. 5. Willingness to disclose (WTD) and perceived sensitivity (PDS) composite scores for various groups of personal data by clusters of employees identified using LCA on latent groups of personal data ($N_1 = 174$, $N_2 = 74$, $N_3 = 145$).

Furthermore, all clusters differed significantly in perceived sensitivity and willingness to disclose *PII* and *WORK*. However, cluster 1 and cluster 2 did not significantly differ for the data groups *SENS* and *NOTSENS* (cf. Fig. 5). There were also no significant differences between clusters for the non-latent data groups. Similarly, all analysis for demographics returned insignificant results. This means that the participants' demographic characteristics did not differ by cluster.

Summary: Our results show that employees can be clustered according to different willingness to disclose and perceived sensitivity. However, the clusters have no differences in terms of demographic properties or privacy beliefs. Second, neither privacy beliefs nor demographic background can predict cluster membership. As a result, the differences appear to be explained solely by employees' perceptions of the four latent data groups.

5 Discussion and implications

Under **RQ1**, we analyzed whether perceived data sensitivity in the employment context differs from other contexts. We find that previous studies' assumption of a global consensus appears to be supported for the on-

line and marketing contexts [1, 54, 81], but becomes obsolete when results are compared to the employment context. We emphasize that our results are very different from other contexts, even when compared to another sample from Germany of similar size ($N_{[81]} = 592$). This supports our assertion that privacy in the employment context deserves dedicated consideration, and that further research is needed to investigate such differences.

Referring to **RQ2$_{a–c}$**, we investigated whether groups of personal data can be identified based on employees' perceived sensitivity and willingness to disclose, and how these variables differ among these groups. We successfully identified a meaningful set of four latent groups that captures the subtleties of perceived sensitivity and willingness to disclose specific to the employment context. The latent groups' characteristics differ significantly from one another, as well as from data groups defined by law and international standards. Perceived sensitivity also seems to be a fairly stable moderate predictor of willingness to disclose across different data groups. Nevertheless, we find that these variables are not equally influenced by context.

As part of **RQ3**, we examined several antecedents frequently studied in the privacy literature, along with their effects on perceived sensitivity and willingness to disclose. Our findings show that employees with strong beliefs of a right to privacy are fairly concerned about collection and unauthorized secondary use. Our findings also support previous assumptions [93] that employees in Germany generally trust their employers to process their personal data. At least for the industries studied, our findings show that overall risk perceptions are low and overall trust is high. Both factors, however, appeared to have little or no effect on willingness to disclose. Instead, depending on the type of data, antecedents differed between trust and concerns, whereas perceived sensitivity seems to be primarily influenced by concerns. Perceived sensitivity also varied according to participant sex, nationality, and company size. With a few exceptions, neither privacy beliefs nor demographics had notable effects on latent groups of personal data.

As part of **RQ4$_{a–b}$**, we clustered employees into groups according to their willingness to disclose and examined the clusters for differences in demographics and privacy attitudes. We identified three clusters that capture various attitudes toward perceived sensitivity and willingness to disclose. Unlike similar approaches in online privacy research [38], however, clusters are not associated with any of the surveyed demographics or privacy beliefs. In parallel, the clusters do not differ for non-latent groups of personal data.

5.1 Theoretical implications

Our results support findings from previous studies that examined contextual differences for willingness to disclose [21, 30, 56]. However, our results strongly suggest that the context affected perceived sensitivity and willingness to disclose differently, or its effect was obscured by other (maybe unknown) factors. For one thing, this is supported by the observed low willingness to disclose personal data perceived as particularly insensitive. In addition, employees in cluster 3 ranked the sensitivity of some data from the private context (e.g., *NOTSENS*) similarly to data directly related to the employment relationship. This resulted in a dichotomy of sensitive and non-sensitive data, which, however, cannot be explained by context alone. Indeed, our findings suggest that context appears to be more important for willingness to disclose (WTD) than for perceived sensitivity (PDS). For example, WTD_{PII} is higher than or equal to both WTD_{WORK} and $WTD_{NOTSENS}$ in all clusters, whereas PDS_{PII} is also higher than both PDS_{WORK} and $PDS_{NOTSENS}$. If context was the strongest driving factor, PDS_{PII} should have been lower than $PDS_{NOTSENS}$. One possible explanation is that perceptions of personal data are influenced not only by general privacy attitudes [48], but also by specific attitudes and norms with varying effects on perceived sensitivity and willingness to disclose in different contexts.

Moreover, our findings suggest that the magnitude between perceived sensitivity and willingness to disclose is largely stable across different groups of personal data. Instead of treating perceived sensitivity as an indirect driver of willingness to disclose [48], its direct effects are also apparent and should be considered. Based on the explanations provided above, this relationship seems particularly well suited to identifying pitfalls where context does not have an equal impact on perceived sensitivity and willingness to disclose. This also implies that examining perceived sensitivity or willingness to disclose in isolation could lead to incorrect conclusions about the specific construct not considered in a study.

Furthermore, our results show that a dichotomous distinction between sensitive and non-sensitive data for studying privacy preferences, as is common in privacy research [53, 61], is also viable in the employment context. However, our cluster analysis showed that perceived sensitivity and willingness to disclose may differ substantially for some data but not at all for others. Strictly dichotomous views cannot capture such subtleties. Therefore, considering the multidimensionality of personal data is clearly preferable [38, 60, 76].

Moreover, our results of the analysis of commonly used antecedents in privacy research [24, 48, 85] confirm that their effects may indeed vary depending on the situation [100] and the sensitivity of the data [61]. In addition, the results of this study also indicate differences by personal data groups, which are likely attributable to the composition of the specific types of data (items) they encompass. The fact that hardly any significant effects were found for the latent data groups suggests that frequently observed effects of antecedents [24, 48] disappear for smaller and more homogeneous data groups. For example, the groups *PII* (latent) and *MASTER* (non-latent) had similar scores for perceived sensitivity and willingness to disclose, and the magnitude and direction of effects between perceived sensitivity and willingness to disclose were identical for the groups *SENS* (latent) and *GDPR* (non-latent) as well as for *WORK* (latent) and *MASTER* (non-latent). This emphasizes these groups' strong similarities. Privacy concerns, however, almost exclusively had significant effects on the non-latent data groups, i.e., those that were considerably more heterogeneous than the latent data groups we identified in the factor analysis. This suggests that privacy concerns are less related to perceived sensitivity and willingness to disclose, and more related to the actual personal data than previously thought. This stresses the importance to make the type of personal data explicit in privacy research [61]. For example, previous studies using the IUIPC and CFIP privacy scales in the employment context have indeed found significant effects of *trust* and *risk beliefs* on *behavioral intentions* [10]. However, they did not explicitly indicate any personal data. Therefore, their results could be attributed to the imprecise questions of the scales and are thus subject to interpretation by the employees. In our own study, we also assessed *trust* in a non-specific way. At the same time, we found significant effects of *trust* on WTD_{ALL} and WTD_{GDPR}. This outcome could be attributed to the fact that the groups *ALL* and *GDPR* reflect specific types of personal data that are salient to employees when being asked general questions about privacy. This means that employees may have intuitively thought about data items contained in these groups when responding. Precise questions about *trust* in handling specific data might have yielded different results.

Contrary to our expectations, *unauthorized secondary use* had a weak positive effect on *trust*, which could be explained by the fact that employees in Germany do indeed make strong demands for their privacy, but at the same time trust their employers to comply with them. This would be supported by the fact that we found extremely high levels of trust combined with low levels of risk beliefs and collection concerns. This would also be consistent with qualitative studies finding that employees suspected violations and infringements with other employers, but not with their own employers [92].

5.2 Practical implications

The study findings also have practical implications for the employment context and its stakeholders.

First, legal and international standards' definitions of what constitutes sensitive personal data may serve as broad guidelines for employers to distinguish between different levels of sensitivity. However, the definitions may not necessarily reflect the data that employees consider to be the most sensitive. Recent studies revealed similar results for the private context [82]. A distinction based on "private" data thus may better reflect the perceptions of both employees [93] and consumers [55].

Furthermore, our cluster analysis shows that a noticeable group of employees is unwilling to disclose truthful data, even if they are highly relevant to the employment context. Employers should be aware that data which are critical to the employment relationship may be perceived as sensitive. However, this view does not seem to be shared equally by all employees. For example, two thirds of our sample perceived *PII* as significantly more sensitive and were less likely to share compared to *WORK*, whereas one third made no difference between the two data groups. Besides, although we did not find significant effects of *trust* on willingness to disclose, it is an essential factor in the relationship between employer and employee [2]. Thus, employers should not take our results as a reason to abandon trust-building measures.

Moreover, our results provide insights for the development of tools that facilitate the exercise of the right to privacy in employment. In the EU, employees are primarily granted far-reaching transparency rights and employers are required to provide information on data processing. Assuming that data with high perceived sensitivity or low willingness to disclose are associated with higher information needs by employees, different levels of detail could be provided for different types of personal data. This might help address the challenge that employees desire comprehensive information on the one hand, but find exercising their rights complex on the other [93]. The identified clusters of employees also suggest that a "one-size-fits-all" solution may not be a satisfactory solution in this regard. Instead, transparency enhancing tools should allow for personalization.

6 Limitations and future work

Like any study in the field of privacy research, our study has some limitations. First, although our results constitute an important step towards complementing the results of prior studies that had U.S.-biased samples [6, 48], the demographic characteristics of our sample limit the generalizability of our results. In particular, results are clearly limited by macro-environmental factors such as the cultural background and the existing strong governmental regulation framework in Germany [48]. Results thus may vary for employees from different regions and cultures [30], and depend on the organization types where our participants worked [90]. Our sample's skew discussed in Section 3.3 represents another limitation in the cultural context of Germany.

Furthermore, our study took place during the COVID-19 pandemic. Possible bias in our results due to a larger number of employees working from home during this time cannot be ruled out. However, effects, if any, are likely to be small, as very few data types in our survey would be affected (e.g., IP address). Moreover, because participation was voluntary, sampling is likely affected by a self-selection bias and limited to the population of employees registered with the panels and employed at the organizations we contacted for recruitment. Previous studies also noted a privacy fatigue in their recruitment of employees [93], which may also affected our study. Nevertheless, our findings suggest that our sample incorporates employees with sufficiently different privacy beliefs and perceptions.

In addition, we acknowledge the justified criticism of the used APCO macro model [14]. However, given that disclosure of some categories of personal data is indispensable in the employment context, the focus on finding inconsistencies (i.e., a paradox) may also require a different interpretation. Our findings may therefore reveal employees' desire for privacy rather than actual behavior. Unlike research on online privacy behavior, it is likely impossible to measure actual disclosure behavior in a cross-sectional study in an employment context [47]. Since employees were found to have limited and erroneous knowledge about what personal data they disclose or are processed by employers [93], research cannot rely on self-reported information on this aspect. If researchers have the resources available, study designs with individual large organizations and unambiguous insights into personal data disclosure (e.g., through management reports) would provide a valuable contribution in this respect, since much of the research on actual disclosure behavior in employment is outdated [91, 99].

Based on our findings, we recommend that future studies, particularly those in non-English speaking countries, should exercise caution in applying the same measurements and assumptions to employment that have been used in previous research [27, 53]. Also because recent work revealed validation problems for such scales, even when used in the original (online) context and with native English speakers [25]. Further limitations of our study are the limited set of examined individual factors and antecedents, missing control variables for employees position, and the set of personal data types focusing on employees from research institutions. Responses to privacy beliefs are also biased by having been asked non-specifically. Participants might have responded differently if questions about privacy beliefs had been asked for specific types of data. The latent data groups identified in this study could form the basis for future research to examine any differences. While our results constitute a step toward closing prevailing research gaps in the employment context [7], future work may take our study' limitations into consideration and also further examine the effects of perceived control, expected benefits, or personality traits.

7 Conclusion

We conducted a cross-sectional survey with 553 employees from Germany to gain insight into perceived sensitivity and willingness to disclose in the employment context, and contribute toward closing prevailing research gaps in privacy research. Our results contribute to the body of knowledge by providing advanced insights into the relationship between perceived sensitivity and willingness to disclose, and its dependence on context and employees' perceptions of personal data. Our findings revealed clear differences between the employment context and other contexts, as well as different effects of antecedents used in online privacy research. We further found that groups of employees can be formed with different levels of perceived sensitivity and willingness to disclose. We consider this study to be an important step in the development of tools that focus on the needs of employees and support them in exercising their right to privacy. Nevertheless, further work is required to address remaining research gaps and improve understanding of the specifics of the employment relationship.

Acknowledgement

We thank our study participants and the participating organizations for supporting our work. We thank the anonymous reviewers and our shepherd for their guidance and insightful comments. This research was supported by the German Federal Ministry of Education and Research (BMBF) under the contract number 16KIS0899.

References

[1] Khaled Almotairi and Bilal Bataineh. Perception of Information Sensitivity for Internet Users in Saudi Arabia. *Acta Informatica Pragensia*, 9(2):184–199, 2020.

[2] Nils Backhaus. Context Sensitive Technologies and Electronic Employee Monitoring: A Meta-Analytic Review. In *2019 IEEE/SICE International Symposium on System Integration (SII)*, pages 548–553, 2019.

[3] Kirstie Ball, Elizabeth M. Daniel, and Chris Stride. Dimensions of employee privacy: An empirical study. *Information Technology & People*, 25(4):376–394, 2012.

[4] Lemi Baruh and Zeynep Cemalcılar. It is more than personal: Development and validation of a multidimensional privacy orientation scale. *Personality and Individual Differences*, 70:165–170, 2014.

[5] Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1):1–48, 2015.

[6] France Bélanger and Robert E. Crossler. Privacy in the Digital Age: A Review of Information Privacy Research in Information Systems. *MIS Quarterly*, 35(4):1017–1042, 2011.

[7] Devasheesh P. Bhave, Laurel H. Teo, and Reeshad S. Dalal. Privacy at Work: A Review and a Research Agenda for a Contested Terrain. *Journal of Management*, 46(1): 127–164, 2020.

[8] Bethany C. Bray, Stephanie T. Lanza, and Xianming Tan. Eliminating Bias in Classify-Analyze Approaches for Latent Class Analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, 22(1):1–11, 2015.

[9] Laura Burbach, Chantal Lidynia, Philipp Brauner, and Martina Ziefle. Data protectors, benefit maximizers, or facts enthusiasts: Identifying user profiles for life-logging technologies. *Computers in Human Behavior*, 99:9–21, 2019.

[10] Shawn F. Clouse, Ryan T. Wright, and Ronald E. Pike. Employee Information Privacy Concerns with Employer Held Data: A Comparison of Two Prevalent Privacy Models. *Journal of Information Privacy and Security*, 6(3): 47–71, 2010.

[11] Tobias Dienlin and Sabine Trepte. Is the privacy paradox a relic of the past? An in-depth analysis of privacy attitudes and privacy behaviors. *European Journal of Social Psychology*, 45(3):285–297, 2015.

[12] Tamara Dinev and Paul Hart. An Extended Privacy Calculus Model for E-Commerce Transactions. *Information Systems Research*, 17(1):61–80, 2006.

[13] Tamara Dinev, Heng Xu, Jeff H Smith, and Paul Hart. Information privacy and correlates: An empirical attempt to bridge and distinguish privacy-related concepts. *European Journal of Information Systems*, 22(3):295–316, 2013.

[14] Tamara Dinev, Allen R. McConnell, and H. Jeff Smith. Research Commentary—Informing Privacy Research Through Information Systems, Psychology, and Behavioral Economics: Thinking Outside the "APCO" Box. *Information Systems Research*, 26(4):639–655, 2015.

[15] Janna-Lynn Dupree, Richard Devries, Daniel M. Berry, and Edward Lank. Privacy Personas: Clustering Users via Attitudes and Behaviors Toward Security Practices. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 5228–5239, 2016.

[16] J. A. Durlak. How to Select, Calculate, and Interpret Effect Sizes. *Journal of Pediatric Psychology*, 34(9):917–928, 2009.

[17] Isioma Elueze and Anabel Quan-Haase. Privacy Attitudes and Concerns in the Digital Lives of Older Adults: Westin's Privacy Attitude Typology Revisited. *American Behavioral Scientist*, 62(10):1372–1391, 2018.

[18] European Union. General Data Protection Regulation. May 2016. Regulation (EU) 2016/679.

[19] Njål Foldnes and Steffen Grønneberg. The sensitivity of structural equation modeling with ordinal data to underlying non-normality and observed distributional forms. *Psychological Methods*, 2021.

[20] John Fox. *polycor: Polychoric and Polyserial Correlations*, 2020. R package version 0.8-0/r22.

[21] Sandra Gabriele and Sonia Chiasson. Understanding Fitness Tracker Users' Security and Privacy Knowledge, Attitudes and Behaviours. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–12, 2020.

[22] Mirta Galesic and Michael Bosnjak. Effects of Questionnaire Length on Participation and Indicators of Response Quality in a Web Survey. *Public Opinion Quarterly*, 73(2): 349–360, 2009.

[23] Shawn N. Geniole, Valentina Proietti, Brian M. Bird, Triana L. Ortiz, Pierre L. Bonin, Bernard Goldfarb, Neil V. Watson, and Justin M. Carré. Testosterone reduces the threat premium in competitive resource division. *Proceedings of the Royal Society B: Biological Sciences*, 286(1903): 20190720, 2019.

[24] Nina Gerber, Paul Gerber, and Melanie Volkamer. Explaining the privacy paradox: A systematic review of literature investigating privacy attitude and behavior. *Computers & Security*, 77:226–261, 2018.

[25] Thomas Groß. Validity and Reliability of the Scale Internet Users' Information Privacy Concerns (IUIPC). *Proceedings on Privacy Enhancing Technologies (PoPETs)*, 2021(2): 235–258, 2021.

[26] Joseph F. Hair, William C. Black, Barry J. Babin, and Rolph E. Anderson. *Multivariate Data Analysis*. Eighth edition, 2019.

[27] David Harborth and Sebastian Pape. German Translation of the Concerns for Information Privacy (CFIP) Construct.

Technical Report SSRN 3112207, 2018.

[28] Jörg Henseler, Christian M. Ringle, and Marko Sarstedt. A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science*, 43(1):115–135, 2015.

[29] Francisco Pablo Holgado–Tello, Salvador Chacón–Moscoso, Isabel Barbero–García, and Enrique Vila–Abad. Polychoric versus Pearson correlations in exploratory and confirmatory factor analysis of ordinal variables. *Quality & Quantity*, 44 (1):153, 2008.

[30] Hsiao-Ying Huang and Masooda Bashir. Privacy by region: Evaluation online users' privacy perceptions by geographical region. In *2016 Future Technologies Conference (FTC)*, pages 968–977, 2016.

[31] Athina Ioannou, Iis Tussyadiah, and Graham Miller. That's Private! Understanding Travelers' Privacy Concerns and Online Data Disclosure. *Journal of Travel Research*, page 0047287520951642, 2020.

[32] ISO. *ISO/IEC 29100:2011(E): Information Technology — Security Techniques — Privacy Framework*. 2011.

[33] Leslie K. John, Alessandro Acquisti, and George Loewenstein. Strangers on a Plane: Context-Dependent Willingness to Divulge Sensitive Information. *Journal of Consumer Research*, 37(5):858–873, 2011.

[34] Terrence D. Jorgensen, Sunthud Pornprasertmanit, Alexander M. Schoemann, Yves Rosseel, Patrick Miller, Corbin Quick, Mauricio Garnier-Villarreal, James Selig, Aaron Boulton, Kristopher Preacher, Donna Coffman, Mijke Rhemtulla, Alexander Robitzsch, Craig Enders, Ruben Arslan, Bell Clinton, Pavel Panko, Edgar Merkle, Steven Chesnut, Jarrett Byrnes, Jason D. Rights, Ylenio Longo, Maxwell Mansolf, Mattan S. Ben-Shachar, Mikko Rönkkö, and Andrew R. Johnson. semTools: Useful Tools for Structural Equation Modeling, 2021.

[35] Michelle L. Kaarst-Brown and E. Dale Thompson. Cracks in the Security Foundation: Employee Judgments about Information Sensitivity. In *Proceedings of the 2015 ACM SIGMIS Conference on Computers and People Research*, pages 145–151, 2015.

[36] Rex B. Kline. Assumptions in structural equation modeling. In *Handbook of Structural Equation Modeling*, pages 111–125. 2012.

[37] Bart P Knijnenburg. Information Disclosure profiles for Segmentation and Recommendation. In *1st USENIX Workshop on Privacy Personas and Segmentation (PPS)*, pages 1–4, 2014.

[38] Bart P. Knijnenburg, Alfred Kobsa, and Hongxia Jin. Dimensionality of information disclosure behavior. *International Journal of Human-Computer Studies*, 71(12):1144–1162, 2013.

[39] Manuel Koller. RobustImm: An R Package for Robust Estimation of Linear Mixed-Effects Models. *Journal of Statistical Software*, 75(1):1–24, 2016.

[40] Selcuk Korkmaz, Dincer Goksuluk, and Gokmen Zararsiz. Mvn: An r package for assessing multivariate normality. *The R Journal*, 6(2):151–162, 2014.

[41] Takashi Koshimizu, Tomoji Toriyama, and Noboru Babaguchi. Factors on the sense of privacy in video surveillance. In *Proceedings of the 3rd ACM Workshop on Continuous Archival and Retrival of Personal Experences (CARPE)*, pages 35–44, 2006.

[42] Ponnurangam Kumaraguru and Lorrie Faith Cranor. Privacy Indexes: A Survey of Westin's Studies. Research Report CMU-ISRI-5-138, Institute for Software Research, International School of Computer Science Carnegie Mellon University Pittsburgh, 2005.

[43] Marija Kuzmanovic and Gordana Savic. Avoiding the Privacy Paradox Using Preference-Based Segmentation: A Conjoint Analysis Approach. *Electronics*, 9(9):1382, 2020.

[44] Alexandra Kuznetsova, Per B. Brockhoff, and Rune H. B. Christensen. lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13):1–26, 2017.

[45] Nancy K. Lankton, D. Harrison McKnight, and John F. Tripp. Facebook privacy management strategies. *Computers in Human Behavior*, 76(C):149–163, 2017.

[46] Benedikt Lebek, Kenan Degirmenci, and Michael H. Breitner. Investigating the Influence of Security, Privacy, and Legal Concerns on Employees' Intention to Use BYOD Mobile Devices. In *Proceedings of the 19th Americas Conference on Information Systems (AMCIS)*, volume 3, pages 2191–2198, 2013.

[47] Lebek Benedikt. Information security awareness and behavior: A theory-based literature review. *Management Research Review*, 37(12):1049–1092, 2014.

[48] Yuan Li. Empirical Studies on Online Information Privacy Concerns: Literature Review and an Integrative Framework. *Communications of the Association for Information Systems*, 28:453–496, 2011.

[49] Drew A. Linzer and Jeffrey B. Lewis. poLCA: An R Package for Polytomous Variable Latent Class Analysis. *Journal of Statistical Software*, 42(1):1–29, 2011.

[50] Steven G. Luke. Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods*, 49(4): 1494–1502, 2017.

[51] Fred A. Mael, Mary Connerley, and Ray A. Morath. None of Your Business: Parameters of Biodata Invasiveness. *Personnel Psychology*, 49(3):613–650, 1996.

[52] Miguel Malheiros, Sören Preibusch, and M. Angela Sasse. "Fairly Truthful": The Impact of Perceived Effort, Fairness, Relevance, and Sensitivity on Personal Data Disclosure. In *Trust and Trustworthy Computing*, Lecture Notes in Computer Science, pages 250–266, 2013.

[53] Naresh K Malhotra, Sung S Kim, and James Agarwal. Internet Users' Information Privacy Concerns (IUIPC): The Construct, the Scale, and a Causal Model. *Information Systems Research (ISRE)*, 15(4):336–355, 2004.

[54] Ereni Markos, George R. Milne, and James W. Peltier. Information Sensitivity and Willingness to Provide Continua: A Comparative Privacy Study of the United States and Brazil. *Journal of Public Policy & Marketing*, 36(1):79–96, 2017.

[55] Ereni Markos, Lauren I. Labrecque, and George R. Milne. A New Information Lens: The Self-concept and Exchange Context as a Means to Understand Information Sensitivity of Anonymous and Personal Identifying Information. *Journal of Interactive Marketing*, 42:46–62, 2018.

[56] Kirsten E. M. Martin and Helen Nissenbaum. Measuring Privacy: An Empirical Test Using Context To Expose Confounding Variables. *Columbia Science and Technology Law Review (STLR)*, 18:176–218, 2015.

[57] Erika McCallister, Tim Grance, and Karen Scarfone. Guide to Protecting the Confidentiality of Personally Identifiable Information (PII). Technical Report 800-122, NIST, 2010.

[58] Tobias Mettler and Jochen Wulf. Physiolytics at the workplace: Affordances and constraints of wearables use from an employee's perspective. *Information Systems Journal*, 29 (1):245–273, 2019.

[59] Miriam J. Metzger. Privacy, Trust, and Disclosure: Exploring Barriers to Electronic Commerce. *Journal of Computer-Mediated Communication*, 9(4), 2004.

[60] George R. Milne, George Pettinico, Fatima M. Hajjat, and Ereni Markos. Information Sensitivity Typology: Mapping the Degree and Type of Risk Consumers Perceive in Personal Data Sharing. *Journal of Consumer Affairs*, 51(1): 133–161, 2017.

[61] David L. Mothersbaugh, William K. Foxx, Sharon E. Beatty, and Sijun Wang. Disclosure Antecedents in an Online Service Context: The Role of Sensitivity of Information. *Journal of Service Research*, 15(1):76–98, 2012.

[62] Patrick Murmann and Simone Fischer-Hübner. Tools for Achieving Usable Ex Post Transparency: A Survey. *IEEE Access*, 5:22965–22991, 2017.

[63] Helen Nissenbaum. Privacy as Contextual Integrity. *Washington Law Review*, 79(1):1119–157, 2004.

[64] Karen L. Nylund, Tihomir Asparouhov, and Bengt O. Muthén. Deciding on the number of classes in latent class analysis and growth mixture modeling: A monte carlo simulation study. *Structural Equation Modeling: A Multidisciplinary Journal*, 14(4):535–569, 2007.

[65] OECD. STAN industry ISIC rev. 4. 2017.

[66] Federal Labour Office. Klassifikation der Berufe 2010 – überarbeitete Fassung 2020 Band 1: Systematischer und alphabetischer Teil mit Erläuterungen. Official document, Bundesagentur für Arbeit, 2021.

[67] Federal Statistical Office. Homepage. https://www.destatis.de/EN/Home/_node.html, 2021.

[68] Judith S. Olson, Jonathan Grudin, and Eric Horvitz. A study of preferences for sharing and privacy. In *CHI Extended Abstracts on Human Factors in Computing Systems*, pages 1985–1988, 2005.

[69] Sandra Petronio. *Boundaries of Privacy: Dialectics of Disclosure*. Boundaries of Privacy: Dialectics of Disclosure. 2002.

[70] Joseph Phelps, Glen Nowak, and Elizabeth Ferrell. Privacy Concerns and Consumer Willingness to Provide Personal Information. *Journal of Public Policy & Marketing*, 19(1): 27–41, 2000.

[71] Svenja Polst, Patricia Kelbert, and Denis Feth. Company Privacy Dashboards: Employee Needs and Requirements. In *1st International Conference on Human-Computer Interaction for Cybersecurity, Privacy and Trust (HCI-CPT)*, pages 429–440, 2019.

[72] Kristopher J. Preacher, Guangjian Zhang, Cheongtag Kim, and Gerhard Mels. Choosing the Optimal Number of Factors in Exploratory Factor Analysis: A Model Selection Perspective. *Multivariate Behavioral Research*, 48(1):28–56, 2013.

[73] Sören Preibusch. Managing diversity in privacy preferences: How to construct a privacy typology. In *1st USENIX Workshop on Privacy Personas and Segmentation (PPS)*, pages 1–6, 2014.

[74] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2020.

[75] William Revelle. *psych: Procedures for Psychological, Psychometric, and Personality Research*. Northwestern University, Evanston, Illinois, 2021. R package version 2.1.3.

[76] Bernard L. Rosenbaum. Attitude toward invasion of privacy in the personnel selection process and job applicant demographic and personality correlates. *Journal of Applied Psychology*, 58(3):333–338, 1973.

[77] Yves Rosseel. Lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, 48(1):1–36, 2012.

[78] Guillaume A. Rousselet, Cyril R. Pernet, and Rand R. Wilcox. A practical introduction to the bootstrap: a versatile method to make inferences by using data-driven simulations. *PsyArXiv*, 2019.

[79] Guillaume A. Rousselet, Cyril R. Pernet, and Rand R. Wilcox. The Percentile Bootstrap: A Primer With Step-by-Step Instructions in R. *Advances in Methods and Practices in Psychological Science*, 4(1):2515245920911881, 2021.

[80] Eva-Maria Schomakers, Chantal Lidynia, Luisa Vervier, and Martina Ziefle. Of Guardians, Cynics, and Pragmatists - A Typology of Privacy Concerns and Behavior:. In *Proceedings of the 3rd International Conference on Internet of Things, Big Data and Security*, pages 153–163, 2018.

[81] Eva-Maria Schomakers, Chantal Lidynia, Dirk Müllmann, and Martina Ziefle. Internet users' perceptions of information sensitivity – insights from Germany. *International Journal of Information Management*, 46(1):142–150, 2019.

[82] Eva-Maria Schomakers, Chantal Lidynia, Dirk Müllmann, Roman Matzutt, Klaus Wehrle, Indra Spiecker genannt Döhmann, and Martina Ziefle. Putting Privacy into Perspective – Comparing Technical, Legal, and Users' View of Information Sensitivity. In *INFORMATIK 2020*, pages 857–870, 2021.

[83] Kim Bartel Sheehan and Mariea Grubbs Hoy. Dimensions of Privacy Concern among Online Consumers. *Journal of Public Policy & Marketing*, 19(1):62–73, 2000.

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

[85] H. Jeff Smith, Tamara Dinev, and Heng Xu. Information Privacy Research: An Interdisciplinary Review. *MIS Quarterly*, 35(4):989–1016, 2011.

[86] Stephanie A. Smith and Steven R. Brunner. To Reveal or Conceal: Using Communication Privacy Management Theory to Understand Disclosures in the Workplace. *Management Communication Quarterly*, 31(3):429–446, 2017.

[87] Daniel J. Solove. A Taxonomy of Privacy. *University of Pennsylvania Law Review*, 154(3):477–560, 2006.

[88] Markus D. Steiner and Silvia Grieder. EFAtools: An r package with fast and flexible implementations of exploratory factor analysis tools. *Journal of Open Source Software*, 5 (53):2521, 2020.

[89] Daniel J. Stekhoven and Peter Bühlmann. MissForest — non-parametric missing value imputation for mixed-type

data. *Bioinformatics*, 28(1):112–118, 2012.

[90] Eugene F. Stone, Hal G. Gueutal, Donald G. Gardner, and Stephen McClure. A field experiment comparing information-privacy values, beliefs, and attitudes across several types of organizations. *Journal of Applied Psychology*, 68(3):459–468, 1983.

[91] Paul D. Tolchinsky, Michael K. McCuddy, Jerome Adams, Daniel C. Ganster, Richard W. Woodman, and Howard L. Fromkin. Employee perceptions of invasion of privacy: A field simulation experiment. *Journal of Applied Psychology*, 66(3):308–313, 1981.

[92] Jan Tolsdorf and Florian Dehling. In Our Employer We Trust: Mental Models of Office Workers' Privacy Perceptions. In *Proceedings of the 1st Asian Workshop on Usable Security (AsiaUSEC, FC Workshop)*, Lecture Notes in Computer Science, pages 122–136, 2020.

[93] Jan Tolsdorf, Florian Dehling, Delphine Reinhardt, and Luigi Lo Iacono. Exploring Mental Models of Informational Self-Determination of Office Workers in Germany. *Proceedings on Privacy Enhancing Technologies (PoPETs)*, 2021 (3):5–27, 2021.

[94] Jennifer Urban and Chris Jay Hoofnagle. The Privacy Pragmatic as Privacy Vulnerable. In *1st USENIX Workshop on Privacy Personas and Segmentation (PPS)*, pages 1–5, 2014.

[95] Cees van der Eijk and Jonathan Rose. Risky Business: Factor Analysis of Survey Data – Assessing the Probability of Incorrect Dimensionalisation. *PLOS ONE*, 10(3):e0118900, 2015.

[96] Jichuan Wang and Xiaoqian Wang. 7.1 The Rules of Thumb for Sample Size Needed for SEM. In *Structural Equation Modeling: Applications Using Mplus*. 2nd edition, 2012.

[97] Marley W. Watkins. Exploratory Factor Analysis: A Guide to Best Practice. *Journal of Black Psychology*, 44(3):219–246, 2018.

[98] Pamela Wisniewski, Bart P Knijnenburg, and Heather Richter Lipford. Profiling Facebook Users' Privacy Behaviors. In *1st USENIX Workshop on Privacy Personas and Segmentation (PPS)*, pages 1–6, 2014.

[99] Richard W. Woodman, Daniel C. Ganster, Jerome Adams, Michael K. McCuddy, Paul D. Tolchinsky, and Howard Fromkin. A Survey of Employee Perceptions of Information Privacy in Organizations. *Academy of Management Journal*, 25(3):647–663, 1982.

[100] Heng Xu, Tamara Dinev, Jeff Smith, and Paul Hart. Information Privacy Concerns: Linking Individual Perceptions with Institutional Privacy Assurances. *Journal of the Association for Information Systems*, 12(12):798–824, 2011.

[101] Charilaos Yiotis, Jennifer C McElwain, and Bruce A Osborne. Enhancing the productivity of ryegrass at elevated CO2 is dependent on tillering and leaf area development rather than leaf-level photosynthesis. *Journal of Experimental Botany*, 72(5):1962–1977, 2021.

[102] Conrad Zygmont and Mario R. Smith. Robust factor analysis in the presence of normality violations, missing data, and outliers: Empirical questions and possible solutions. *The Quantitative Methods for Psychology*, 10(1):40–55, 2014.

A Items and questions

All questions and items had an option "don't answer".

Demographics:
Are you or have you been employed within the last few months, but not exclusively in self-employment? [Yes; No]

Are you employed by more than one employer? [Yes; No]

In which country are you primarily employed? [List]

In what industry/sector does your employer operate? [list OECD industries [65]]

How many employees work for the company or organization? [< 10; < 50; < 250; < 1000; ≥ 1000]

What professional group do you consider yourself to belong to? [List OECD professions [65]]

How long have you been employed by your current employer? [Number input]

Do you have permanent employment? [Yes; No]

What is your highest level of education? [List]

What was your income (net earnings), i.e. wage or salary after deduction of taxes and social security contributions, in the last month? [< 500; < 1000; < 1500; < 2000; < 2500; < 3000; < 3500; < 4000; ≥ 4000]

What is your age? [number]

What is your biological sex? [Diverse; Male; Female]

What is your citizenship (country)? [List]

Are you currently primarily in education or training? [Yes; No]

Perceived data sensitivity six-point scale "NOT-SENSitive at all" and "Very sensitive":

Assume your current employer has / would have access to the following information and data about you / from you. How sensitive would you rate each of these pieces of information? [62 items in Fig. 2]

Willingness to disclose four-point scale "No, under no circumstances" and "Yes, actually always":

Suppose you were free to decide what data you would provide to your current employer. Would you give them access to the following information and data? [62 items in Fig. 2]

Collection concern six-point scale "Strongly disagree" and "Strongly agree":

(1) It usually bothers me when my employer asks me for my personal data.

(2) When my employer asks me for personal data, I sometimes think twice before providing it.

(3) It bothers me to give personal data to my employer.

(4) I'm concerned that my employer collects too much personal data about me.

Privacy as a right six-point scale "Strongly disagree" and "Strongly agree":

(1) Employee privacy laws should be strengthened to protect personal privacy against employers.

(2) Employees need legal protection against employers' misuse of personal data.

(3) If I were to write a constitution today, I would probably add employee privacy as a fundamental right.

Risk beliefs six-point scale "Strongly disagree" and "Strongly agree":

(1) In general, it would be risky to give my personal data to my employer.

(2) There would be high potential for loss associated with giving my personal data to my employer.

(3) There would be too much uncertainty associated with giving my personal data to my employer.

(4) Providing my employer with my personal data would involve many unexpected problems.

Trust six-point scale "Strongly disagree" and "Strongly agree":

(1) I trust that my employer would keep my best interests in mind when dealing with my personal data.

(2) My employer is in general predictable and consistent regarding the usage of my personal data.

(3) My employer is always honest with me when it comes to using my personal data that I would provide.

(4) My employer handles the personal data they collect about their employees in a proper and confidential way.

Unauthorized secondary use six-point scale "Strongly disagree" and "Strongly agree":

(1) My employer should not use my personal data for any purpose unless I have authorized it.

(2) When I disclose my personal data to my employer for some reason, my employer should never use the data for any other reason.

(3) My employer should never share my personal data with other companies unless it has been authorized by the individuals who provided the information.

Survey feedback:
How did you like this survey? [1 - 5]

How did you process this survey?

Did you work conscientiously on the questions? [No, not at all'; Rather not; Mostly yes; Yes, very]

Did you answer truthfully? [No, not at all; Rather not; Mostly yes; Yes, very]

Is there anything else you would like to tell us or provide feedback on the survey? [free text]

B Analysis environment

Statistical analysis was conducted in R. A detailed list of all packages used for analysis is provided in Table 7.

Table 7. R packages used for analysis.

Analysis	Package	Version	Src
All	R	4.0.3	[74]
Bootstrapping CI	bootcorci	0.0.0.9000	[78]
Exploratory Factor Analysis	psych	2.1.3	[75]
Exploratory Factor Analysis	EFAtools	0.3.1	[88]
Exploratory Factor Analysis	polycor	0.8.0	[20]
Confirmatory Factor Analysis, Structural Equation Models	lavaan	0.6.9	[77]
Confirmatory Factor Analysis, Structural Equation Models	semTools	0.5.5	[34]
Univariate and Multivariate Normality	MVN	5.8	[40]
Imputation	missForest	1.4	[89]
Linear Mixed Models	lme4	1.1.27.1	[5]
Linear Mixed Models	lmerTest	3.1.3	[44]
Robust Linear Mixed Models	robustlmm	2.4.4	[39]
Latent Class Analysis	poLCA	1.4.1	[49]

C Participant demographics

Demographics are reported in Table 8 and Table 9.

Table 8. Participant demographics I.

Description	Part 1	Part 2	*Germany*
Participants	N: 553	N: 393	
Sex	%	%	%
Diverse	0.2	0.0	*n. a.*
Female	39.6	41.7	*46.5*
Male	59.7	58.3	*53.5*
Age (years)	%	%	%
$\leq$ 24	8.7	9.9	*1.3*
25 – 34	32.4	3.5	*22.1*
35 – 44	27.1	29.0	*21.9*
45 – 54	14.6	14.0	*23.6*
55 – 64	16.5	15.8	*29.9*
$\geq$ 65	.7	.8	*1.2*
Job tenure (years)	%	%	%
$\leq$ 4	47.3	46.6	*27.6*
5 – 9	24.1	24.4	*19.1*
$\geq$ 10	28.6	29.0	*44.3*
Org. size (num. employees)	%	%	%
$<$ 10	8.0	7.1	*18.0*
10 – 249	34.4	32.8	*38.0*
250 – 999	25.7	26.7	*44.0*
$\geq$ 1k	31.6	33.1	
Education	%	%	%
University degree	58.2	58.3	*16.9*
Doctorate degree	5.4	4.6	*n. a.*
Master's degree	20.1	23.9	*n. a.*
Diploma's degree	13.9	11.7	*n. a.*
Bachelor's degree	18.8	18.1	*n. a.*
Technical school degree	5.2	3.8	*n. a.*
Apprenticeship / vocational training	14.6	16.8	*n. a.*
Advanced technical college or university entrance qualification	13.0	13.2	*n. a.*
Intermediate diploma	6.1	5.6	*n. a.*
Secondary school leaving certificate	5.4	4.6	*n. a.*
No general school degree	1.1	1.0	*n. a.*
No specification / other	1.8	1.3	*n. a.*
Other	%	%	%
Permanent employment	75.8	75.6	*n. a.*
Multiple jobs	7.6	7.6	*5.4*

Description	Part 1	Part 2	*Germany*
Net income (€ / month)	%	%	%
$<$ 1k	9.2	12.2	*13.0*
1k $<$ 2k	36.7	31.6	*42.0*
2k $<$ 3k	36.9	36.4	*29.0*
3k $<$ 4k	11.4	12.7	*10.0*
$\geq$ 4k	5.8	7.1	*6.0*
Nationality	%	%	%
Germany[1]	88.2	86.0	*87.5*
United States	2.0	2.0	*n. a.*
United Kingdom	1.5	1.5	*n. a.*
Greece	1.1	1.1	*n. a.*
Portugal	0.6	0.6	*n. a.*
Australia	0.5	0.5	*n. a.*
Bulgaria	0.5	0.5	*n. a.*
Egypt	0.5	0.5	*n. a.*
India	0.5	0.5	*n. a.*
Ireland	0.5	0.5	*n. a.*
Ukraine	0.5	0.5	*n. a.*
Argentina	0.3	0.3	*n. a.*
Brazil	0.3	0.3	*n. a.*
Colombia	0.3	0.3	*n. a.*
Estonia	0.3	0.3	*n. a.*
France	0.3	0.3	*n. a.*
Hungary	0.3	0.3	*n. a.*
Indonesia	0.3	0.3	*n. a.*
Italy	0.3	0.3	*n. a.*
Japan	0.3	0.3	*n. a.*
Lebanon	0.3	0.3	*n. a.*
Malaysia	0.3	0.3	*n. a.*
Mexico	0.3	0.3	*n. a.*
Pakistan	0.3	0.3	*n. a.*
Poland	0.3	0.3	*n. a.*
Romania	0.3	0.3	*n. a.*
Russian Federation	0.3	0.3	*n. a.*
Serbia	0.3	0.3	*n. a.*
Spain	0.3	0.3	*n. a.*
Switzerland	0.3	0.3	*n. a.*
Turkey	0.3	0.3	*n. a.*
Turkmenistan	0.3	0.3	*n. a.*
Vietnam	0.3	0.3	*n. a.*
Zimbabwe	0.3	0.3	*n. a.*

Note. Part 1: full sample, Part 2: subsample (Part 1 $\supset$ Part 2), Germany: population of employees in Germany [67]
Percentages include missing responses (omitted for brevity). Maximum non-response rate is $\leq$ 2%.
[1]Germans may have a second citizenship.

Table 9. Participant demographics II.

Description	**Part 1**	**Part 2**	***Germany***
Participants	N: 553	N: 393	
Industry (OECD, [65])	%	%	%
Information and communication	14.6	15.3	3.7
Professional, scientific and technical activities	12.5	14.2	5.8
Education	11.2	11.7	6.8
Human health and social work activities	9.2	10.4	13.2
Financial & insurance activities	9.0	6.9	2.9
Public administration and defense; Compulsory social security	7.8	8.4	6.9
Manufacturing	11.0	9.5	19.0
Wholesale & retail trade	6.0	6.1	13.6
Transportation and storage	3.4	3.3	5.1
Administrative and support service activities	3.3	4.1	5.1
Elecricity, gas, steam, air con. and water supply; sewerage, waste management and remediation activities	1.8	1.3	1.4
Accommodation and food service activities	1.8	1.5	3.7
Arts, entertainment and recreation	1.8	1.8	1.4
Construction	1.4	1.3	6.7
Real estate activities	1.3	1.0	0.5
Other service activities	1.1	0.5	2.8
Agriculture, hunting, forestry and fishing	0.7	1.0	1.2
Professional group (Federal Labour Office, [66])	%	%	%
Science, geography & information technology	21.0	19.3	4.2
Business org., accounting, law & administration	21.0	17.8	20.4
Health, social services, teaching & education	16.5	18.1	18.8
Commercial services, trade, hotel & tourism	10.5	12.2	11.4
Linguistics, literature, humanities, social sciences, economics, media, arts, culture & design	8.1	9.4	2.7
Mining, production & manufacturing	6.0	6.1	21.0
Transport, logistics	3.1	3.1	6.4
Construction, architecture, geodetic surveying and construction engineering	2.7	3.1	6.1
Protection, security and surveillance	1.8	2.0	1.1
Military	0.5	0.3	n. a.
Agriculture, forestry and animal husbandry	0.8	0.8	0.7
Cleaning	0.5	0.5	2.5

Note. Part 1: full sample, Part 2: subsample (Part 1 $\supset$ Part 2), Germany: population of employees in Germany [67]
Percentages include missing responses (omitted for brevity). Maximum non-response rate is $\leq$ 2%.

D Personal data elements and groups of personal data

A summary of the five different studies and contexts compared in Section 4.1 is provided in Table 10. The table also includes the descriptive statistics about the scores for perceived data sensitivity (PDS) and willingness to disclose (WTD) from this study as well as the scores extracted from related work. Furthermore, the average scores for all personal data items and for all studies compared in Section 4.1 and depicted in Fig. 2 are reported in Table 11. In addition, Table 11 also includes a mapping between the different personal data items and the eight different groups of personal data investigated in Section 4.3.

Table 10. Comparison different studies and personal data items.

Description		**This study**	**Markos et al. [54]**	**Markos et al. [54]**	**Schomakers et al. [81]**	**Almotairi and Bataineh [1]**
Study and sample	Year	2021	2017	2017	2018	2020
	Context	Employees	Marketing	Marketing	Online users	Online users
	Country	Germany (DEU)	USA	Brazil (BRA)	Germany (DEU)	Saudi Arabia (SAU)
	N	553	406	401	592	508
Items and scores	All personal data items by study					
	Num. items	62	42	42	40	35
	Perceived data sensitivity					
	min	2.5	4.5	3.0	2.8	2.4
	max	9.6	9.4	9.1	9.3	9.7
	average	6.0	7.0	5.6	6.4	6.1
	median	6.0	6.9	5.7	6.7	6.3
	Willingness to disclose					
	min	1.5	1.7	1.7	n. a.	n. a.
	max	9.1	6	6.3	n. a.	n. a.
	average	5.4	3.8	4.2	n. a.	n. a.
	median	5.4	3.8	4.4	n. a.	n. a.
	Intersection of the studied personal data items between all studies					
	Num. items	28	28	28	28	28
	Perceived data sensitivity					
	min	3.0	2.9	4.5	3.0	3.8
	max	9.7	9.6	9.4	9.1	9.3
	average	6.2	6.5	6.9	5.6	6.6
	median	6.4	6.7	6.9	5.6	7.0
	Willingness to disclose					
	min	1.5	1.7	1.7	n. a.	n. a.
	max	9.1	5.9	6.3	n. a.	n. a.
	average	4.7	3.7	4.1	n. a.	n. a.
	median	4.0	3.5	4.4	n. a.	n. a.

Table 11. Personal data elements' average scores in different studies and assignment to different groups of personal data.

Personal data	Scaled average scores ($1 \leq \bar{x} \leq 10$)								Groups of personal data						
	Employment DEU [this]		Marketing USA [54]		Marketing BRA [54]		Online DEU [81]	Online SAU [1]	Predefined groups			Latent groups			
	PDS	WTD	PDS	WTD	PDS	WTD	PDS	PDS	IDENT	MASTER	GDPR	SENS	NOTSENS	PII	WORK
Hair color	2.5	8.1					3.2						✓		
Profession	2.9	9.1	4.6	5.9	4.0	6.0	4.3	3.1		✓					✓
Language skills	3.0	8.9								✓					
Business trip	3.2	8.8													✓
Employment	3.2	8.6								✓					✓
Shift plans	3.2	8.6													✓
Education	3.3	8.9								✓					
Professional appoint.	3.4	8.6													✓
Priv. postal code	3.5	7.9	4.9	5.6	4.4	5.2	4.6	4.3		✓					
Place of birth	3.5	8.0	5.8	4.8	3.9	5.8	4.5			✓					
Working hours	3.5	8.8								✓					
Body size	3.6	7.2	4.5	5.9	3.2	6.1	3.8	3.4					✓		
No. of children	3.9	7.4	4.6	5.7	3.7	5.9	4.3	3.5		✓					
Driving license	4.0	7.5	8.4	2.2	7.1	2.8			✓						
Name of pet	4.0	5.8					2.8	2.4							
Professional contacts	4.3	7.3													
Family status	4.4	7.4								✓					
Work contract	4.5	8.3								✓					
Picture	4.6	7.1	6.4	4.0	5.5	4.3	7.0	6.9							
Home address	4.7	7.8	6.9	4.4	6.0	4.5	7.5	6.4	✓	✓				✓	
Application documents	4.8	8.2								✓					
Social security No.	4.9	7.6	9.4	1.7	7.7	2.7	7.9	7.7	✓	✓				✓	
Health insurance No.	5.0	7.5	8.4	2.1	6.5	2.9	7.9	7.1	✓	✓	✓			✓	
Priv. license plate	5.2	5.4	6.6	3.0	6.1	3.1	5.7	5.0	✓						
Income level	5.3	6.2	6.7	4.6	5.7	3.8	6.9	7.3							
Union membership	5.4	5.5									✓				
Account No.	5.5	7.5	9.2	1.7	7.8	2.2	9.3	8.8	✓	✓				✓	
Priv. phone No.	5.6	6.3	6.8	4.1	5.0	5.0	7.5	5.1	✓	✓					
Performance data	5.6	6.9								✓					
Mother's maiden name	5.8	4.3	7.2	3.5	4.9	4.4	5.0	4.1							
Body weight	6.0	4.7	5.6	5.1	3.8	5.6	5.0	3.7			✓		✓		
Priv. email address	6.0	5.8	6.4	5.1	4.8	5.7	6.0	5.1	✓	✓					
Vacation resort	6.0	4.5													
Sideline activities	6.3	5.5								✓					
Fitness	6.6	3.8					4.5	3.1			✓				
Religious affiliation	6.7	3.6	5.0	5.6	3.2	6.3	4.0	3.0		✓	✓				
Pregnancy	6.7	5.3									✓				
Formal warning	6.8	4.2													
Passport No.	7.0	4.0	8.4	2.1	6.5	2.7	8.2	8.7	✓						
Social network profile	7.1	2.8	6.1	4.0	4.8	4.8	6.0	5.4							
Digital signature	7.2	4.1	8.2	2.6	7.4	2.6	8.0	8.6	✓						
Voiceprint	7.3	3.4	7.3	2.7	6.0	3.5	7.0	7.2	✓		✓				
Shopping behavior	7.5	2.6					5.5	4.9							
Political opinion	7.7	3.0	4.7	5.3	3.0	5.7	5.3	5.9			✓				
Sexual Orientation	7.7	3.0	5.3	5.3	3.4	6.0	6.0				✓				
IP add.	7.7	3.0	7.6	3.0	6.8	3.0	8.0	7.8	✓						
Criminal record	7.7	4.0									✓				
Intention changing job	7.8	3.6													
Alcohol consumption	7.9	2.6					5.0								
Law enforcment rec.	8.2	3.5	7.6	2.9	4.3	4.7	7.0	8.2			✓				
Creditworthiness	8.5	2.8	7.9	2.9	7.2	2.4	7.7	8.6				✓			
Medical history	8.5	3.2	8.4	3.0	5.2	4.7	7.5	6.7			✓	✓			
Browsing history	8.6	2.3					7.0	6.4							
Medication	8.6	2.9					6.5	5.8			✓	✓			
GPS location	8.7	2.4	6.8	3.4	5.9	3.4	7.5	7.1	✓			✓			
Fingerprint	8.7	2.7	8.5	2.3	7.1	2.6	8.0	9.2	✓		✓				
Credit card No.	8.7	2.6	9.2	1.9	8.3	2.1	8.7		✓						
Priv. appointments	8.7	2.7													
Online dating activities	8.9	1.5					6.5	6.2			✓				
Personal problems	9.0	2.4										✓			
Genetic data	9.2	1.7	8.4	2.2	6.0	3.3	8.0		✓		✓	✓			
Passwords	9.6	1.7	9.3	1.7	9.1	1.7	9.3	9.7	✓						

E Demographic differences privacy antecedents

Results of our analysis for differences in our participants' demographics between the five different privacy antecedents investigated as part of our causal model are presented in Table 12 and in Table 13, respectively. To test for differences in age (younger vs. older), nationality (German vs. not German), sex (male vs. not male), job tenure ($\leq$ 6 years vs. $>$ 6 years), number of employers (one vs. more than one), permanent employment (yes vs. no), education (university degree vs. no university degree), and company size (working in SME vs. not working in SME), demographics were included as binary exogenous variables in a structural equation model (SEM) that included all five privacy antecedents. The SEM was run with polychoric correlations and the robust estimator WLSMV. Differences with respect to participants' industry and professional group were tested using Kruskal-Wallis test.

Table 12. Results SEM analysis demographics.

Demographics		**Collection concern**			**Privacy as a right**			**Risk beliefs**			**Trust**			**Unauthz. secondary use**		
Regressions		Est.	CI_{95}	β	Est.	CI_{95}	β	Est.	CI_{95}	β	Est.	CI_{95}	β	Est.	CI_{95}	β
Age (is older)	$\rightarrow$	-.12	[-.35, .11]	-.07	-.18	[-.40, .04]	-.11	-.03	[-.25, .18]	-.02	.12	[-.13, .36]	.06	.38	[.12, .63]	.22**
Is German	$\rightarrow$	.00	[-.31, .30]	.00	-.27	[-.54, .00]	-.12	-.08	[-.36, .20]	-.04	.00	[-.33, .32]	.00	-.08	[-.41, .25]	-.03
Is male	$\rightarrow$	.13	[-.08, .35]	.08	.21	[.00, .43]	.13	.17	[-.04, .38]	.11	-.02	[-.26, .21]	-.01	-.10	[-.33, .13]	-.06
Job tenure (longer)	$\rightarrow$	-.08	[-.31, .15]	-.05	.03	[-.19, .24]	.02	.00	[-.21, .21]	.00	.09	[-.16, .33]	.05	-.01	[-.26, .23]	-.01
Multiple employers	$\rightarrow$	-.26	[-.61, .08]	-.08	.44	[.09, .78]	.14*	-.35	[-.71, .01]	-.12	.22	[-.18, .62]	.06	-.04	[-.41, .34]	-.01
Permanent empl.	$\rightarrow$	-.07	[-.31, .17]	-.04	-.03	[-.26, .20]	-.02	-.06	[-.29, .17]	-.03	.05	[-.21, .30]	.02	-.07	[-.32, .19]	-.03
University deg.	$\rightarrow$	.08	[-.12, .28]	.05	-.11	[-.30, .09]	-.07	-.09	[-.27, .10]	-.06	-.01	[-.24, .21]	-.01	-.18	[-.41, .05]	-.10
Works for SME	$\rightarrow$	-.06	[-.26, .13]	-.04	-.05	[-.25, .14]	-.03	-.08	[-.27, .11]	-.05	.11	[-.11, .33]	.06	-.12	[-.34, .10]	-.07

Note. N = 393, β: standardized path coefficient (measure of effect size [16]) *: p $<$ .05 **: p $<$ .01 ***: p $<$ .001

Table 13. Results Kruskal-Wallis test demographics.

Demographics	**Collection concern**		**Privacy as a right**		**Risk beliefs**		**Trust**		**Unauthz. secondary use**	
Industry	H	23.126	H	16.96	H	24.913	H	31.33	H	32.674
	df	17	df	17	df	17	df	17	df	17
	p	0.145	p	0.457	p	0.097	p	0.0182	p	0.0124
	η^2	0.017	η^2	0.038	η^2	0.021	η^2	0.038	η^2	0.043
Professional group	H	9.347	H	10.895	H	8.067	H	2.933	H	15.526
	df	11	df	11	df	11	df	11	df	11
	p	0.59	p	0.452	p	0.707	p	0.992	p	0.16
	η^2	-0.004	η^2	-0.004	η^2	-0.008	η^2	-0.004	η^2	0.012

Note. N = 393

F Covariates SEM analysis

Results of our analysis for differences in our participants' demographics between the willingness to disclose personal data (WTD) and the perceived data sensitivity (PDS) for different groups of personal data are presented in Table 14. To test for differences in age (younger vs. older), nationality (German vs. not German), sex (male vs. not male), job tenure (≤ 6 years vs. > 6 years), number of employers (one vs. more than one), permanent employment (yes vs. no), education (university degree vs. no university degree), and company size (working in SME vs. not working in SME), demographics were included as binary exogenous variables in the structural equation model (SEM) used for analysis in Section 4.4. The results presented in Table 14 thus complement the results presented in Table 6 above.

Table 14. Results SEM analysis demographic variables.

Regressions demographics on WTD and PDS for different groups of personal data: DEMO → {WTD, PDS}													
		ALL			GDPR			IDENT			MASTER		
Regressions		Est.	CI_{95}	β	Est.	CI_{95}	β	Est.	CI_{95}	β	Est.	CI_{95}	β
Is male	→ PDS	-.32	[-.58, -.06]	-.14*	-.32	[-.57, -.06]	-.14*	-.22	[-.47, .03]	-.10	-.28	[-.52, -.04]	-.13*
Is German	→ PDS	.03	[-.32, .39]	.01	.40	[.06, .74]	.13*	-.13	[-.46, .21]	-.04	-.10	[-.42, .22]	-.03
Works for SME	→ PDS	-.17	[-.41, .06]	-.08	-.25	[-.48, -.01]	-.11*	-.04	[-.27, .19]	-.02	-.09	[-.31, .14]	-.04
University deg.	→ PDS	-.06	[-.29, .17]	-.03	.25	[.01, .48]	.11*	-.18	[-.41, .05]	-.08	-.14	[-.37, .08]	-.07
Age (is older)	→ PDS	-.12	[-.40, .16]	-.05	-.07	[-.35, .21]	-.03	-.20	[-.47, .08]	-.09	-.14	[-.40, .12]	-.07
Permanent empl.	→ PDS	.11	[-.17, .38]	.04	.09	[-.19, .38]	.03	.12	[-.15, .39]	.05	.12	[-.15, .38]	.05
Job tenure (longer)	→ PDS	-.09	[-.35, .17]	-.04	-.18	[-.45, .08]	-.08	-.08	[-.35, .18]	-.04	-.02	[-.26, .23]	-.01
Multiple employers	→ PDS	.02	[-.39, .42]	.00	-.18	[-.59, .23]	-.04	-.13	[-.58, .32]	-.03	.04	[-.40, .49]	.01
Is male	→ WTD	-.12	[-.37, .14]	-.05	-.09	[-.34, .16]	-.03	-.13	[-.37, .12]	-.05	-.12	[-.36, .13]	-.05
Is German	→ WTD	.48	[.09, .88]	.14*	.36	[-.08, .79]	.10	.35	[-.02, .72]	.10	.57	[.20, .94]	.17***
Works for SME	→ WTD	.11	[-.11, .33]	.04	.14	[-.07, .36]	.05	.12	[-.11, .35]	.05	.08	[-.14, .30]	.03
University deg.	→ WTD	.11	[-.12, .35]	.05	-.12	[-.34, .11]	-.04	.13	[-.10, .36]	.05	.16	[-.08, .39]	.06
Age (is older)	→ WTD	.08	[-.18, .34]	.03	-.04	[-.27, .19]	-.02	.08	[-.19, .35]	.03	.04	[-.22, .30]	.02
Permanent empl.	→ WTD	.25	[-.03, .53]	.09	.28	[-.03, .60]	.09	.25	[-.05, .54]	.09	.07	[-.21, .34]	.02
Job tenure (longer)	→ WTD	-.05	[-.32, .21]	-.02	.08	[-.16, .32]	.03	.03	[-.24, .30]	.01	-.10	[-.36, .15]	-.04
Multiple employers	→ WTD	.00	[-.44, .43]	.00	.20	[-.23, .64]	.04	-.04	[-.52, .43]	-.01	-.22	[-.68, .23]	-.05

		NOTSENS^{L}			PII^{L}			SENS^{L}			WORK^{L}		
Regressions		Est.	CI_{95}	β	Est.	CI_{95}	β	Est.	CI_{95}	β	Est.	CI_{95}	β
Is male	→ PDS	.00	[-.22, .22]	.00	.15	[-.07, .38]	.07	.03	[-.22, .27]	.01	.10	[-.13, .33]	.05
Is German	→ PDS	.08	[-.22, .39]	.03	-.23	[-.55, .09]	-.08	.08	[-.19, .36]	.03	-.14	[-.47, .20]	-.05
Works for SME	→ PDS	.09	[-.13, .30]	.04	.18	[-.04, .39]	.09	-.01	[-.22, .21]	.00	.04	[-.18, .26]	.02
University deg.	→ PDS	-.04	[-.26, .18]	-.02	-.01	[-.23, .21]	-.01	-.03	[-.25, .19]	-.01	.00	[-.22, .23]	.00
Age (is older)	→ PDS	-.07	[-.32, .18]	-.03	.04	[-.20, .29]	.02	-.11	[-.37, .14]	-.06	.13	[-.13, .38]	.06
Permanent empl.	→ PDS	.11	[-.16, .38]	.05	.00	[-.25, .25]	.00	-.05	[-.33, .23]	-.02	-.04	[-.29, .21]	-.02
Job tenure (longer)	→ PDS	.13	[-.12, .38]	.06	-.04	[-.29, .21]	-.02	.22	[-.03, .48]	.11	.00	[-.26, .26]	.00
Multiple employers	→ PDS	-.23	[-.64, .18]	-.06	-.36	[-.76, .05]	-.09	-.24	[-.61, .14]	-.06	-.43	[-.88, .02]	-.11
Is male	→ WTD	-.18	[-.42, .06]	-.07	-.08	[-.32, .16]	-.03	.01	[-.23, .26]	.01	-.06	[-.33, .21]	-.03
Is German	→ WTD	.13	[-.17, .42]	.04	-.05	[-.39, .29]	-.01	.08	[-.23, .39]	.02	.21	[-.11, .52]	.07
Works for SME	→ WTD	.01	[-.21, .23]	.01	-.22	[-.43, -.01]	-.10*	.07	[-.16, .30]	.03	-.06	[-.28, .15]	-.03
University deg.	→ WTD	.16	[-.06, .38]	.07	.04	[-.18, .25]	.02	-.09	[-.34, .16]	-.04	.14	[-.08, .35]	.06
Age (is older)	→ WTD	-.07	[-.31, .17]	-.03	-.07	[-.31, .17]	-.03	-.21	[-.48, .05]	-.09	-.01	[-.25, .23]	-.01
Permanent empl.	→ WTD	-.13	[-.43, .17]	-.05	-.31	[-.58, -.04]	-.12*	-.31	[-.56, -.05]	-.11*	-.11	[-.38, .16]	-.04
Job tenure (longer)	→ WTD	.20	[-.04, .44]	.08	-.03	[-.26, .21]	-.01	.33	[.05, .61]	.14*	.11	[-.12, .35]	.05
Multiple employers	→ WTD	-.21	[-.63, .20]	-.05	.04	[-.33, .40]	.01	-.21	[-.66, .25]	-.05	.07	[-.42, .56]	.02

Note. N = 393, β: standardized path coefficient (measure of effect size [16]), ^{L}: Latent groups, *: $p < .05$ **: $p < .01$ ***: $p < .001$