

An Investigation into the Sensitivity of Personal Information and Implications for Disclosure: A UK Perspective

Rahime Belen-Saglam¹, Jason R.C. Nurse^{1,*} and Duncan Hodges²

¹*School of Computing, University of Kent, Canterbury, Kent, UK*

²*Centre For Electronic Warfare, Information and Cyber, Cranfield University, Defence Academy of the United Kingdom, Shrivenham, UK*

Correspondence*:
Corresponding Author
j.r.c.nurse@kent.ac.uk

2 ABSTRACT

3 The perceived sensitivity of information is a crucial factor in both security and privacy concerns
4 and the behaviours of individuals. Furthermore, such perceptions motivate how people disclose
5 and share information with others. We study this topic by using an online questionnaire where
6 a representative sample of 491 British citizens rated the sensitivity of different data items in a
7 variety of scenarios. The sensitivity evaluations revealed in this study are compared to prior
8 results from the US, Brazil and Germany, allowing us to examine the impact of culture. In addition
9 to discovering similarities across cultures, we also identify new factors overlooked in the current
10 research, including concerns about reactions from others, personal safety or mental health
11 and finally, consequences of disclosure on others. We also highlight a difference between the
12 regulatory perspective and the citizen perspective on information sensitivity.

13 We then operationalised this understanding within several example use-cases exploring
14 disclosures in the healthcare and finance industry, two areas where security is paramount.
15 We explored the disclosures being made through two different interaction means: directly to a
16 human or chatbot mediated (given that an increasing amount of personal data is shared with
17 these agents in industry). We also explored the effect of anonymity in these contexts. Participants
18 showed a significant reluctance to disclose information they considered 'irrelevant' or 'out of
19 context' information disregarding other factors such as interaction means or anonymity. We also
20 observed that chatbots proved detrimental to eliciting sensitive disclosures in the healthcare
21 domain; however, within the finance domain, there was less effect. This article's findings provide
22 new insights for those developing online systems intended to elicit sensitive personal information
23 from users.

24 **Keywords:** personal information disclosure, information sensitivity, privacy, chatbots, conversational user interfaces, conversational
25 agents

1 INTRODUCTION

26 The internet has enabled people throughout the world to connect with each other in ways that previously
27 would have been considered unimaginable. To enable such interactions, individuals are often required to

28 share various types of information and this can in turn lead to privacy concerns about how their personal
29 information is stored, processed and disclosed to others.

30 From research, we know that a user's privacy concerns and their willingness to disclose information are
31 affected by the perceived sensitivity of that information (Markos et al., 2018). However, it is vague and
32 open to debate as to how 'sensitive' information may be categorised. A risk-oriented definition is adopted
33 by some studies in the literature as seen in the EU's General Data Protection Regulation (GDPR) (European
34 Parliament, 2016) which defines sensitive information as follows:

35 *Personal data which are, by their nature, particularly sensitive in relation to fundamental rights and freedoms*
36 *merit specific protection as the context of their processing could create significant risks to the fundamental*
37 *rights and freedoms.*

38 However, several other dimensions are also introduced to explain how users perceive sensitivity including:
39 perceived risk, possibility of harm or public availability of data can lead information to be perceived as
40 sensitive (Ohm, 2014; Rumbold and Pierscionek, 2018). In addition to studies which explore the factors
41 leading to a high perceived sensitivity, it is possible to report two other research themes in this area. Firstly,
42 studies that report the perceived sensitivity of different data items at granular levels or in different usage
43 contexts (Milne et al., 2017; Markos et al., 2017; Schomakers et al., 2019; Belen Sağlam et al., 2022).
44 Secondly, studies which investigate the relationship between information sensitivity and disclosure (Wadle
45 et al., 2019; Aiello et al., 2020; Belen Sağlam and Nurse, 2020; Treiblmaier and Chong, 2013; Bansal et al.,
46 2016).

47 This research aims to provide a UK perspective on the research areas identified above, a problem that is
48 missing in existing literature. To the best of our knowledge, there is also no study that synthesizes findings
49 associated with the factors that lead certain information to be considered sensitive, sensitivity ratings
50 of different personal data items and the comfort felt while disclosing them under different conditions.
51 Therefore, we formulated our research question as follows: 'What are the perspectives of British citizens
52 regarding the sensitivity of the information and the impact of different factors on the disclosure of personal
53 information?'. To answer this research question and provide key related insights into this issue, the
54 following research objectives (RO) are defined:

- 55 • RO1: Identify the main factors that lead British citizens to regard certain information as sensitive.
- 56 • RO2: Explore the levels of sensitivity associated with the different personal data items
- 57 • RO3: Explore the impact of user factors on levels of sensitivity of the different personal data items.
- 58 • RO4: Explore if there is an international consensus on the level of sensitivity of the personal data items
59 (comparing Germany, the US, Brazil and the UK).
- 60 • RO5: Determine the impact of context/situation (specifically finance or health domains) on an
61 individual's level of comfort in disclosing information.
- 62 • RO6: Determine the impact of interaction means (human or chatbot) while sharing personal information
63 on individual's level of comfort in disclosing information.
- 64 • RO7: Determine the impact of anonymity (identified or anonymous) on individual's level of comfort in
65 disclosing information.

66 Through this research, we contribute to the literature on information sensitivity and disclosure in three
67 novel ways:

- 68 1. We provide insights into the factors that lead to certain information being considered sensitive and
69 provide a UK perspective on these debates.
- 70 2. We provide sensitivity ratings of different data items for UK citizens and explore the international
71 consensus on data sensitivity. Those findings can further help to inform discussions on the process of
72 cross-national data flows.
- 73 3. We empirically investigate the impact of demographic characteristics, anonymity, context (health and
74 finance), and interaction means (human or chatbot) on information sensitivity and comfort to provide
75 information.

76 Our findings, therefore, can also contribute to an understanding of how to design inclusive information
77 systems when sensitive disclosures are required. The assumption we make in this study is that comfort
78 is inversely related to sensitivity; i.e., the more comfortable an individual is in sharing some personal
79 information, the less sensitive that information is perceived to be, this is consistent with prior work (e.g.
80 Ackerman et al., 1999).

81 The remainder of this paper is structured as follows. The Literature Review section summarises the
82 literature relevant to our research question. We present our methodology in the Research Methodology
83 section and following this, we present our descriptive results in Results section. We critically reflect on and
84 consider our findings in the Discussions section, as well as highlighting the implications for research and
85 practice. The paper closes with a discussion of the limitations of the research and future plans.

2 LITERATURE REVIEW

86 This section summarises the relevant literature underpinning this research in following four sub-categories.

87 2.1 What makes information sensitive?

88 A fundamental challenge for protecting personal information is first defining how it can be conceptualised
89 and categorised. While there are several different opinions in the literature about how sensitive personal
90 information may be defined, regulatory frameworks can provide a robust foundation. The European General
91 Data Protection Regulation (GDPR) considers personal data sensitive if it reveals a racial or ethnic origin,
92 political opinions, religious or philosophical beliefs, trade union membership, data concerning health, sex
93 life and sexual orientation. In addition to these data types, genetic data and biometric data also fall into this
94 category. The GDPR covers those data items in a special category defined as '*data that requires specific*
95 *protection as the context of their processing could create significant risks to an individual's fundamental*
96 *rights and freedoms*' (European Parliament, 2016).

97 One notable study on sensitive information, Ohm (2014) aimed to understand what makes information
98 sensitive and focused on a list of categories of information that have been legally treated as sensitive,
99 primarily from the United States. This list of sensitive categories was then employed to infer the
100 characteristics of information types that result in it being considered sensitive. In brief, four factors
101 were reported when assessing whether a given piece of information seems sensitive: the possibility of
102 harm, probability of harm, presence of a confidential relationship, and whether the risk reflects majoritarian
103 concerns.

104 A schema has been proposed for assessing data categories to guide the relative sensitivities of different
105 types of personal information (Rumbold and Pierscionek, 2018). The paper explores several factors that
106 influence the perception of personal data as sensitive, including the public availability of data, the context of

107 the data use and its potential to identify individuals. Contrary to popular belief, researchers stated that data
108 publicly observable is not necessarily non-sensitive data (Rumbold and Pierscionek, 2018). The potential
109 of certain information being used to infer new information when aggregated with others is another factor
110 leading to a perception of sensitivity. Several other issues, such as the risk of re-identification, automated
111 profiling, behavioural tracking and trustworthiness of the person/system with whom the data is shared, are
112 also given as potential problems to affect sensitivity evaluation of particular information types. The massive
113 increase in sensors associated the internet-of-things (IoT) devices (e.g., sensor data, or heart-rate data
114 from wearable devices) within the medical domain has increased the amount of health data collected from
115 citizens. This has raised the risk of third party data access such as health professionals or even insurance
116 companies (Levallois-Barth and Zylberberg, 2017). Sharing data with third parties may increase the risk of
117 discrimination and also make it possible to infer the prevalence of certain pathologies. Therefore, Levallois-
118 Barth and Zylberberg (2017) claim that even though those data items may not be potentially sensitive when
119 considered in isolation, sensitivity evaluations may change in the future. However, surprisingly, Kim et al.
120 (2019) revealed that within healthcare, sensitivity has no statistically significant impact on the willingness
121 to provide privacy information even though it significantly influences the perceived privacy risk. Those
122 conflicting findings highlight some of the challenges in sensitivity evaluations and disclosure which will be
123 explained further in Section 2.3.

124 Finally, the nature of the technology also has an impact on the sensitivity evaluations and data storage
125 decisions accordingly. For instance, due to its immutable nature which prevents data being changed, Kolan
126 et al. (2020) argued that personal medical data should not be stored directly on public blockchain systems.
127 This was confirmed by Zheng et al. (2018) who also preferred not to store health information in blockchain
128 in their proposed solution. Based on that, it can be argued that the concerns regarding the use of data in the
129 future shapes the sensitivity evaluations of personal data.

130 **2.2 What types of information are perceived as sensitive?**

131 In addition to the studies that explore the factors leading individuals to perceive certain information as
132 sensitive, studies have also categorised data types according to the perceived sensitivity.

133 In one of those studies researchers identified two clusters of information that were considered more
134 sensitive: secure identifiers (e.g., social security number) and financial information (e.g., financial accounts
135 and credit card numbers). It is noted that basic demographics (e.g., gender, birth date) and personal
136 'preferences' (e.g., religion, political affiliation) were seen as less sensitive by the survey respondents
137 (Milne et al., 2017).

138 Another study by Markos et al. (2017), used a cross-national survey between consumers in the United
139 States and Brazil to explore the cultural differences in the perception of sensitivity. The authors examined
140 42 information items concluding that US consumers generally rated information as more sensitive and
141 were less willing to provide information to others than their Brazilian counterparts. Financial information
142 and identifiers were observed to have the highest perceived sensitivity with security codes and passwords,
143 financial account numbers, credit card numbers, or formal identifiers such as social security number and
144 driving licence number appeared in a cluster of highly sensitive data.

145 A similar study has been conducted that provided a German citizen perspective on information sensitivity
146 (Schomakers et al., 2019). Researchers compared their results with the results from the US and Brazil (Milne
147 et al., 2017; Markos et al., 2017) and noted that, on average, the perceptions of information sensitivity of
148 German citizens lies between that of US and Brazilian citizens. Cluster analysis revealed that similar data
149 items were considered highly sensitive by the three countries except that German citizens considered the

150 credit score to appear in a medium-sensitive cluster whilst US and Brazilian citizens considered this to be
151 in a higher-sensitivity cluster. However, in general, German citizens were reported to perceive passwords as
152 most sensitive, followed by identifiers such as financial account numbers, passport numbers or fingerprints.

153 In addition to those studies that focus on general items of information, some researchers focused on
154 specific information domains. For example, Bansal et al. (2010) focused on health information and the role
155 of individual differences on perceived information sensitivity and disclosure in this domain. Meanwhile,
156 Ioannou et al. (2020) focused on travel providers and their customers' privacy concerns when sharing
157 biometric and behavioural data and the impact of these concerns on the willingness to share this data. This
158 study highlighted the context-dependence of privacy preferences. It is reported that although travellers
159 worry about the privacy of their data, they are still willing to share their data, and the disclosure decision is
160 dependent upon expected benefits rather than privacy concerns. Confirming the 'privacy paradox' (Norberg
161 et al., 2007), it was found that there was no link between privacy concerns and willingness to share
162 biometric information and that expected benefits outweigh privacy concerns in the privacy decisions made
163 by travellers.

164 Research has also examined attitudes towards sharing PII and non-PII (anonymous) data (Markos et al.,
165 2018); they differentiated the information that was already public, hypothesising that items associated
166 with the 'private-self' are perceived as more sensitive than public-self items. Their results demonstrated
167 that some anonymous information like diary/journal entries, hygiene habits, home information, and GPS
168 location are considered sensitive and even more sensitive than PII, conflicting slightly with the general
169 societal interpretation and legislative focus. More expectedly, they identified that private-self information
170 items were perceived as more sensitive than public-self items.

171 **2.3 When do we disclose more?**

172 There are multiple debates regarding personal information disclosure in the literature, some of which
173 consider data sensitivity and other factors such as the perception of benefit. For instance, research has found
174 that people are more willing to disclose when their human needs such as health or security are fulfilled
175 (Wadle et al., 2019); thus, explaining the impact of expected benefits on information disclosure.

176 Conversely other research proposed that the perceived privacy risks play a more significant role than
177 the expected benefits (Keith et al., 2013). The difference in their results was explained by the high degree
178 of realism they provided in their experiments, where participants were given a real app that dynamically
179 showed actual data.

180 In another recent study, perceived privacy risks were argued to significantly reduce the intention to
181 disclose information and the disclosure behaviour, whilst privacy concerns were reported to affect disclosure
182 intention but not the actual information disclosure behaviour (Yu et al., 2020).

183 The impact of personal differences has also been studied; for example, less healthy individuals were
184 more concerned about disclosing their health information arguably due to the risk of their status on
185 employment opportunities or social standing (Bansal et al., 2016). This finding confirms previous studies
186 by Treiblmaier and Chong (2013) who demonstrated that a higher level of perceived risk leads to a lower
187 level of willingness to disclose personal information. The same research examined the role of trust in
188 information disclosure and reported that the direct influence of trust in the Internet (as a communication
189 media) is statistically insignificant. However, the trust of an online vendor (the ultimate receiver of the
190 information) impacts the willingness to disclose.

191 It has also been shown that the perceived fairness of a data request also impacts personal information
192 disclosure (Malheiros et al., 2013). The ‘fairness’ of a data request describes the individual’s belief that
193 data being collected will be used for the purpose communicated by the data receiver and in an ethical
194 manner. The study revealed that when participants saw a disconnect between the disclosures they were
195 asked to make and the specified purpose of the disclosure, they consider it unfair and opted not to disclose.

196 The impact of anonymity has also been studied in a recent study (Schomakers et al., 2020) that reported
197 that the critical element of online privacy and privacy in data sharing is the protection of the identity, and
198 thus, anonymity. The most substantial effect associated with data sharing was the anonymisation level,
199 followed by the type of data (how sensitive it is) and how much the person with whom the information
200 is shared is trusted. It was reported that when the participants can understand why the data is useful to
201 the receiver, they are more willing to provide data. Benefits for the self or the society are also reported
202 as important aspects while deciding to share data. It is clear that when it comes to PII, sensitivity plays a
203 greater role in willingness to disclose than it has for anonymous information, i.e. information that is not
204 personally identifiable (Markos et al., 2018).

205 **2.4 How may non-human agents impact disclosure?**

206 A chatbot is an application created to automate tasks and imitates a real conversation with a human
207 in their natural language (whether spoken or through a textual interface). Today, conversational agents
208 are used in various industries, including finance and health care. In these applications, the collection of
209 personal information is essential to provide an effective service. Consequently, research has focused on
210 disclosing information to chatbots and the modulating factors that enable or degrade disclosure. In one
211 of those studies, it was concluded that users disclose as much to chatbots as they would to humans (Ho
212 et al., 2018), resulting in similar disclosure processes and outcomes. The researchers added that relatively
213 neutral questions might not make a difference between chatbots and humans, and when asked a question
214 that may be embarrassing and might result in negative evaluation, users were also found to respond with
215 more disclosure intimacy to a chatbot than a human.

216 Another study highlighted a similar issue and noted that individuals tended to talk more freely with
217 a chatbot, without perceiving they were being judged or making the chatbot bored of listening to them
218 (Bjaaland and Brandtzaeg, 2018). Accessibility and anonymity are given as other characteristics of chatbots
219 that encourage self-disclosure. ‘Icebreaker questions’ (e.g. ‘how are you doing?’, ‘how is the weather?’) or
220 human-like fillers (e.g. ‘um’, ‘ahh’) are also reported to lead to more effective communication and a sense
221 of a shared experience (Bell et al., 2019; Bhakta et al., 2014).

222 Other research has considered the importance of context and investigated the effects of socio-emotional
223 features on the intention to use chatbots (Ng et al., 2020). While a preference for a technical and mechanical
224 chatbot for financially sensitive information was identified, no significant differences were observed in the
225 disclosure of socially attributed items (such as name, date-of-birth and address) between the chatbots with
226 and without socio-emotional traits.

227 The lack of coherence in the scope of the studies that investigate the impact of employing chatbots on
228 information disclosure has encouraged us to design this study. We systematically investigate the comfort in
229 disclosing sensitive information to a chatbot, varying the context of the domain and the sensitivity levels
230 of data items. We aim to present a rigorous and systematic understanding of the impact on information
231 disclosures from conversational agents.

3 RESEARCH METHODOLOGY

232 In order to answer our research question and achieve the individual research objectives, a rigorous
233 methodology was defined, this was oriented around an online questionnaire and robust qualitative and
234 quantitative data analysis. The questionnaire engaged a sample of 500 British participants and critically
235 explored the topic of information sensitivity. We opted for a questionnaire (e.g., instead of interviews
236 or focus groups) to reach a census representative sample of UK citizens. The questionnaire design (i.e.,
237 questions asked, sequence of questions) and subsequent data analysis techniques were composed specifically
238 to allow us to address each research objective, and address the research question. In what follows, we
239 explain the questionnaire design, present the participant recruitment strategy, and detail the techniques
240 used to analyse the data gathered.

241 3.1 Questionnaire design

242 The questionnaire was implemented on the Survey Monkey platform, and participants were asked to
243 respond to questions posed across five sections. First, we posed questions to collect informed consent from
244 participants. In the second section, demographic characteristics of the participants (age group, gender, and
245 educational level) were gathered. Having gathered this biographic information, the next sections were
246 closely associated with the research objectives. The third section targeted RO1 specifically and therefore
247 asked participants for the reasons or factors that might lead them to consider certain personal information
248 more sensitive than other personal information. This was presented as an open-ended question to allow
249 participants to present any factors they viewed appropriate.

250 The fourth section asked participants questions about the sensitivity of a range of personal data items.
251 These questions provide the basis for achieving RO2 (i.e., exploring the levels of sensitivity of the different
252 personal data items), RO3 (i.e., exploring the impact of user factors on sensitivity of the different personal
253 data items) and RO4 (i.e., enabling a comparison of British citizens' sensitivity perceptions with perceptions
254 from citizens from the US, Brazil and Germany (Markos et al., 2017; Schomakers et al., 2019)).

255 To determine the data items for our study, we decided to use data items covered in existing studies as a
256 basis and enrich those lists in accordance with our research objectives. Some of the original data items by
257 Markos et al. (2017) and Schomakers et al. (2019) were not appropriate for our scenarios and therefore
258 were eliminated, for example: DNA profile, fingerprint, digital signature or browsing history are not easily
259 shared with chatbots due to their nature. We paid particular attention to the differences in the sensitivity
260 classification of Schomakers et al. (2019) to that of Markos et al. (2017). We included the data items that
261 were assigned different sensitivity levels between those two studies. We also expanded our list with data
262 items considered sensitive by the GDPR or any data protection acts of EU countries, the US, China and the
263 UK. These regulations were reviewed, and any data items that were identified as requiring extra controls or
264 given as 'special categories' were added to our list.

265 The complete list of data items is in Table 1. In order to better understand these data items within the
266 context of the domains we considered (health and finance), these data items were manually categorised as
267 either General data items, Health-related information, or Financial information.

268 To examine participants' opinions on the sensitivity of these 40 data items, participants were asked to
269 rank each data item on a 6-point symmetric Likert scale which ranged from 'not sensitive at all' (1) to 'very
270 sensitive' (6). Throughout the study, we used a 6-point scale as done by Schomakers et al. (2019) to enable
271 a direct comparison between nationalities. A 6-point scale has also been shown to avoid overloading the
272 participants' discrimination abilities (Lozano et al., 2008).

Table 1. The full list of data items used in the study

Category	Data item
General data items	Passwords, Passport Number, Formal Identification Number, IP Address, Private Phone Number, Current Location, Home Address, Criminal Records, Face Picture, Online Dating Activities, Sex Life, Sexual Orientation, Email Address, Social Network Profile, License Plate Number, Shopping habits, Political Affiliation, Weight, Mother's Maiden Name, Post Code, Place Of Birth, Number Of Children, Religion, Height, Hair Colour, Name Of Pet, Trade Union Membership, Social Welfare Needs, Racial or Ethnic Origin, Full Name, Education Records, Date of Birth, Citizenship, Marital Status, Gender
Health Information	Alcohol Consumption, Smoking Habits, Substance Abuse Conditions, Mental Health, HIV and/or other sexually transmitted diseases, Medical Diagnoses, Chronic Diseases
Financial Information	Credit Card Number, Credit Score, Income Level, Occupation, Bank Account Credentials

273 For the fifth and final section of the questionnaire, a set of questions was posed to assess the effects of
 274 three variables, i.e., identification (anonymous or identified), context (finance or health) and interaction
 275 means (a human or chatbot), on the comfort in disclosing personal information (RO5-7); thus, was a 2x2x2
 276 factorial design. Participants were asked to rate their comfort level while disclosing particular data items in
 277 each of the scenarios summarised below in Table 2. For example, in scenario 1 (S1) the question was given
 278 as follows: 'Assume that you are speaking to a person on an online health service website where you do
 279 not need to identify yourself (i.e., you can be anonymous). How comfortable would you feel disclosing
 280 (i.e., sharing) the personal information listed below?'. Comfort levels were assessed again on a 6-point
 281 Likert scale ranging from 1 'Not comfortable at all' to 6 'Very comfortable'.

Table 2. Scenarios used in the study

ID	Interaction Means	Context	Anonymity
S1	Person	Health	Anonymous
S2	Person	Finance	Anonymous
S3	Person	Health	Identified
S4	Person	Finance	Identified
S5	Chatbot	Health	Anonymous
S6	Chatbot	Finance	Anonymous

282 In order to reduce the possible overload of participants, two scenarios have been eliminated from the
 283 study. These would be S7 and S8 to complete the 2x2x2 design where participants would be asked to
 284 disclose personal information to a chatbot where they needed to identify themselves. When piloting the
 285 study, it became apparent that the quality of the responses was significantly reduced beyond six scenarios.
 286 This pragmatic decision allowed us to focus on the six scenarios which would supply the most value to
 287 practitioners.

288 To determine the data items to use for this final part of the questionnaire, we abridged the original list
 289 of data items and selected 20 items; ten were general data items, five were health related, and five were
 290 finance related. This abridging was another pragmatic choice to reduce the load on our participants whilst
 291 still delivering a solid evidence base for practitioners. While shortening the list, we retained data items that
 292 are frequently subject to debates in the literature. Personal identifiers, data items in the special category of

293 the GDPR or personal information related to health and finance were maintained in this list for this reason
294 (See Table 3).

Table 3. Reduced set of 20 data items used in the final stage of the study

Category	Data item
General data information	GPS Location, Criminal Records, Sex Life, Social Network Profile, License Plate Number, Political Affiliation, Mother's Maiden Name, Religion, Trade Union Membership, Racial or Ethnic Origin
Health information	Alcohol Consumption, Mental Health, HIV and/or other sexually transmitted diseases, Medical Diagnosis, Chronic Diseases
Finance information	Credit Card Number, Credit Score, Income Level, Occupation, Bank Account Credentials

295 We included six attention checking questions to ensure the quality of our data. The scenarios in the
296 second step were randomised in the questionnaire software to avoid any sequence bias. The data items (i.e.,
297 the lists of 40 and 20 items) in the questions were also randomised for the same purposes. The study has
298 been reviewed and ethically approved by the Research Ethics & Governance department of University of
299 Kent and Cranfield University Research Ethics Committee.

300 3.2 Participants

301 Participants were recruited using Prolific in order to reach a census representative sample of UK citizens.
302 Since this study's ultimate goal is to understand UK citizens' perspective, it was essential to gather
303 responses from a representative set of the public. This platform was also selected since it has good quality
304 and reproducibility compared to other crowdsourcing platforms (Peer et al., 2017).

305 Before running our questionnaire, we conducted a pilot study with 50 participants to ensure that the
306 questionnaire design and time limits were appropriate and usable for the intended/target audience. We
307 then released the complete questionnaire on a sample of 500 participants (i.e., representative of the UK
308 population based on age, sex and ethnicity), paying £8.72 per hour, which is at least the UK minimum
309 wage. In total, the questionnaire took 15 minutes to complete.

310 From the 500 responses gathered, nine participants failed more than one attention question and thus were
311 excluded from the data analysis. We present the demographics of the final 491 participants in Table 4.

312 3.3 Data Analysis

313 To analyse the data gathered, we used techniques most appropriate for the respective question set (See
314 Figure 1). After collecting consent and demographic characteristics of the participants at the beginning
315 of the questionnaire, in the first step, to achieve RO1 we asked reasons or factors that lead participants
316 to consider certain personal information as more sensitive than other personal information. We used
317 thematic analysis to analyse this qualitative data (Braun and Clarke, 2006). Firstly, brief labels (codes) were
318 produced for each response, and when all data had been initially coded, themes were identified, grouping
319 responses with similar codes into the same category. Finally, the themes were reviewed to check whether
320 the candidate themes appeared to form a coherent pattern.

321 The analysis conducted to achieve RO2 was descriptive and we ordered the data items by computing their
322 average sensitivity ratings. For RO3, we built proportional-odds logistic regression models for each data
323 type to model the effects of age, gender and education. This modelling approach allows us to build a model

Table 4. Demographic Profile of Participants (GCSEs are the qualifications taken in Years 10 and 11 of secondary school in the UK. A-levels are a subject-based qualification offered by the educational bodies in the UK to students completing secondary or pre-university education)

Age	18-24	10.4%
	25-34	19.2%
	35-44	15.9%
	45-54	18.9%
	55-Over	35.6%
Gender	Female	50.3%
	Male	49.7%
Education	GCSE	15.5%
	A-level or equivalent	28.1%
	Undergraduate degree	34.4%
	Postgraduate degree	18.7%
	Doctorate	3.3%

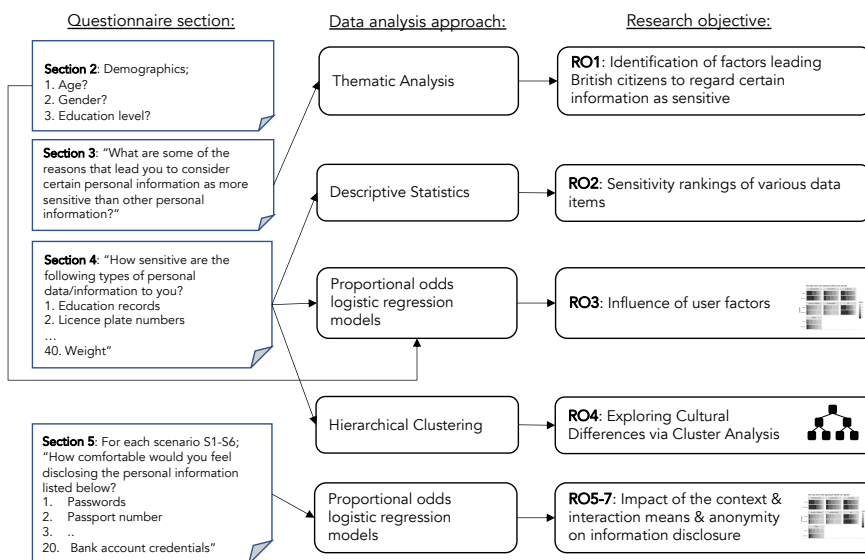


Figure 1. Study design

324 that predicts a particular participant’s probability of giving a data item a particular sensitivity rating based
 325 on their age, gender, and education level. By exploring these model coefficients, we can gain insight into
 326 the effects of these variables on how comfortable people are disclosing sensitive information.

327 To achieve RO4, we used hierarchical cluster analysis (Bridges Jr, 1966) to group data types based on
 328 their perceived sensitivity. Initially, each data item is assigned to an individual cluster before iterating
 329 through the data items and at each stage merging the two most similar clusters, continuing until there is one
 330 remaining cluster. At each iteration, the distance between clusters is recalculated using the Lance-Williams
 331 dissimilarity (Murtagh and Contreras, 2012). This clustering allowed us to build a tree diagram where the
 332 data items viewed as being of similar sensitivity are placed on close together branches.

333 Finally, for Research Objectives 5 to 7 we used proportional-odds logistic regression modelling to analyse
 334 the effects of anonymity, context and interaction means, using these three variables to predict the comfort
 335 level while disclosing personal information.

4 RESULTS

336 This section describes the results from both the open-ended qualitative question and the quantitative results
 337 from the Likert scale questions. Further discussion of the results is explored in Section 5.

338 4.1 RO1: Identification of factors leading British citizens to regard certain information 339 sensitive

340 As mentioned previously, we asked our participants an open-ended question regarding the factors that
 341 lead them to consider a data item to be sensitive. A thematic analysis of the responses led to several factors
 342 being identified. These included some of the factors reported in the literature, such as the risk of harm, trust
 343 of interaction means, public availability of data, context, and identification. However, we identified several
 344 other areas that have been overlooked or not dealt with comprehensively. These new themes included
 345 concerns regarding the reactions from the listener, concerns regarding personal safety or mental health,
 346 consequences of disclosure on beloved ones or careers, or concerns regarding sharing information about
 347 others such as family members or friends.

348 The complete set of themes and codes are presented in Table 5 with the number of responses related to
 349 each theme and code. These summaries provide a useful indicator of the themes emerging from the study
 350 and the popularity of each theme.

Table 5. Thematic analysis of what makes data sensitive

Themes	Codes
Privacy (181)	Identity (64), Private information (45), Identity theft (35), Access to more (18), Third party sharing (9), Personal life (5), Tracing (5)
Context (135)	Finance (80), Health (55)
Financial Problems (100)	Risk of fraud (69), Financial loss (18), Impact on career (12), Financial exposure (1)
Reactions (84)	Embarrassment (31), Discrimination (17), Judgement (15), Reputational harm (12), Cultural conditioning (5), Reactions in general (4)
Consequence of disclosure on me (84)	Personal security (18), Misuse (18), Harm (18), Personal safety (8), Risk of crime (7), Mental Health (6), Legal issues (3), Harassment (2), Cost & Benefit (1)
Nature of information (43)	Relevance (17), Public Availability (10), Information of others (7), Value (5), Group (2), Stability (1), Delicacy (1)
Interaction means (26)	Concerns regarding the recipient (20), Trust (6)
Consequence of disclosure on others (21)	Impacts on others (15), Security of others (3), Safety of other (2), Child grooming (1)

351 In the remainder of this section, we provide details of the most pertinent themes emerging from our study.
 352 The names of the themes and the codes under themes are written in italics.

353 4.1.1 Privacy concerns

354 Privacy concerns expressed by the participants while evaluating the sensitivity level of information often
 355 focused on *identity theft*. In our study, 35 participants expressed their concerns in a finance context where
 356 credentials or some other identifiers were given as examples to sensitive personal information due to their
 357 potential exploitation for identity fraud. Identifiers or other information used to identify individuals when
 358 used together were also considered sensitive by several participants even if identity theft was not explicitly

359 mentioned. For some participants, it was enough to consider a piece of personal information as sensitive if
360 it could reveal their *identity*.

361 Another concern that emerged under the privacy theme was *private information*. Within this code,
362 data items were reported to be considered more sensitive if the owners of them preferred to keep them
363 private. Medical histories and financial status are mainly considered private and, hence sensitive by those
364 participants. These participants also mentioned unsolicited emails, phone calls or customised advertisements
365 as an effect of sharing information about themselves. A particular category under this privacy concern
366 pertained to *personal life* where preferences in life, family information or relations with partners were
367 considered sensitive by participants.

368 Interestingly, respondents found some publicly available information to be sensitive due to the potential
369 use to *access more information* about the individuals. Again this was most notable when that new
370 information was related to the health or financial status of the individuals. One poignant example in
371 this category was the name of a pet or mother's maiden name, information commonly used for security or
372 password questions.

373 Other emergent concerns included the fear of being physically traced; data items that would allow
374 individuals to be traced were considered sensitive by a group of participants: '*People being able to find
375 where I live or work or steal my identity.*', '*you can use it to track somebody, find out other information
376 related to what you have ...*'.

377 The final code related to privacy violations was the risk of third-party sharing. Some participants
378 considered personal information sensitive when they thought it might be shared with other groups and
379 become more widely available than expected. This concern around third-party sharing is increasingly in
380 line with the studies that argue that third-party access leads to privacy concerns (e.g. Pang et al., 2020).

381 4.1.2 Two main contexts of sensitive personal information: Health and Finance

382 In addition to the themes that led participants to consider certain information as more sensitive, our
383 analysis also identified two primary contexts that heavily dominated the responses; health and finance.
384 Hence, it is possible to report a consensus on the sensitivity of the health and finance-related information.
385 Participants noted that these data items were expected to be given a higher standard of protection by the
386 systems that process them. Some responses exhibited concerns regarding health information being sold
387 or passed to insurance companies or other bodies interested in this information. Conversely, some others
388 worried about the impact of disclosing their health status on their financial creditworthiness or career. Some
389 participants also found health-related information inherently very private and thus sensitive, without giving
390 any consequence as a reason.

391 Finance is a significantly more common response to our question when compared with health data.
392 Several participants provided finance-related personal information as an example of sensitive information.
393 In addition, several other data items, outside of a finance context, were considered sensitive by participants
394 due to their impact on participants' financial status. Even though financial loss dominates the responses,
395 some other factors such as impacts on career and financial confidentiality also led participants to find
396 information more sensitive.

397 4.1.3 Financial Problems

398 As discussed previously, financial concerns dominated the responses. Consequences under this theme
399 centre around *financial loss*, *financial exposure*, *risk of fraud* and *negative impacts on career*. The risk of

400 fraud appeared to be the largest concern as many participants reported information to be more sensitive if it
401 could enable fraudulent activities. More specific responses were given by some participants where *financial*
402 *loss* was explicitly given as a concern while evaluating the sensitivity level of information. *Financial*
403 *exposure*, which could be considered an overlapping area between the themes *Privacy* and *Financial*
404 *problems*, was another code that emerged in the responses. Finally, when evaluating the sensitivity level,
405 several participants reflected on the impacts on their career of disclosing financial information. Political
406 and religious affiliations, and medical histories, were popular examples given as sensitive information that
407 participants believed could compromise their careers or aspirations.

408 4.1.4 Concerns regarding the reactions of people

409 Another concern of participants observed was the interpersonal *reactions* between the individual sharing
410 the information and the individual to whom the information was disclosed. Under this theme, the
411 most common reaction was *embarrassment* with participants reporting that information that they found
412 embarrassing to disclose was considered sensitive.

413 Medical records or being a member of protected characteristics were given as examples of sensitive
414 information since they were considered embarrassing for themselves or their families. Similarly,
415 *discrimination* was another code that emerged under this category. A group of participants reported
416 a data item to be sensitive if they believed it would invoke the prejudice or bias of others. Religious or
417 political affiliation, sexual orientation, race, disability or genetic defects and health information were
418 examples given as sensitive due to this concern. Disclosure of personal health information has been known
419 to result in discrimination by employers and insurance agencies if they gain access to such information
420 (Rindfleisch, 1997).

421 Participants also reported finding information sensitive if it may cause them to be judged by others. In
422 addition to *judgement*, *reputational harm* was another factor that led participants to consider a data item
423 sensitive. We also identified *cultural conditioning*, which some participants highlighted as ‘taboo’ subjects
424 within society and considered items related to those taboos more sensitive (e.g., sex life, political leanings)
425 purely because of this societal/cultural conditioning.

426 4.1.5 Consequences of disclosure on the individual

427 A majority of responses under this theme exhibited answers where participants defined sensitive
428 information as the information that could be *misused/used against them* or cause them *harm*. Some
429 participants provided more specific answers and negative effects on *mental health* and *personal safety* or
430 feelings such as *harassment* and *fear*.

431 *Personal security* was one of the most popular responses with participants linking sensitivity to a resulting
432 security risk. It was not possible to differentiate in the majority of the responses if the given concern
433 was about the individuals’ physical security or digital security (e.g., ‘I have concerns about security’,
434 ‘Things which might compromise my security’). However, some responses implicitly covered it where
435 participants gave ‘home address’ or ‘bank account number’ as examples. *Risk of crime* is another code in
436 this category. Participants were aware that some personal details could be used fraudulently and considered
437 those sensitive. It is worthy of note here that almost all the concerns given in this category were in a
438 financial context.

439 There were very few responses where participants shared their concerns regarding *legal issues*. Those
440 participants reported perceiving information as sensitive if used legally against them (e.g., ‘official bodies

441 can use it to deny services.’). On the other hand, one participant explicitly reported considering the *costs*
442 *and benefits* of disclosing information into account while evaluating its sensitivity.

443 4.1.6 Nature of the information

444 Some participants reported data as more sensitive due to its very nature. For example, characteristics
445 can be given as *intimacy of data* which are generally exemplified with sexual life or other information
446 related to personal life. Participants found these data items sensitive due to their intimate nature. Another
447 characteristic reported was the *value of the data*, which determines to what extent others can use it as it is
448 disclosed. For instance, passwords or passport numbers were seen as more sensitive than social media data
449 since they are perceived as having a higher impact if misused. The *relevance* is another code that emerged
450 which defines the relevance of the information request in the given scenario. Fairness of the request was
451 also given as a pertinent factor: ‘*There are certain details I would not wish to share as I do not feel they are*
452 *of relevance to the data handler*’.

453 A small group of participants considered data items that are costly to change (e.g., home address) more
454 sensitive than items where the cost is lower (e.g., email address). Another response, albeit relatively rare,
455 was when the data item was related to a particular *group* identity. For example, information about minors
456 or vulnerable groups were considered sensitive. Existing research reported that a particular data item might
457 only be sensitive where the individual belongs to a group that often faces discrimination (Rumbold and
458 Pierscionek, 2018). For example, gender at birth is likely to be less sensitive for those who are cisgender
459 compared to those who are transgender.

460 Some participants also considered the *public availability of information* while evaluating the sensitivity
461 of it and considered that data items that were already publicly known were less sensitive.

462 4.1.7 Interaction means

463 Disregarding the content of the information, some participants reported another essential factor; *the*
464 *person/system that the information is shared with*. We identified several participants for whom the sensitivity
465 of information is related to the receiver of the information. For some participants, it was explicitly a matter
466 of *trust*, a data item as more sensitive if they did not trust the person or the system to whom they are
467 disclosing it.

468 4.1.8 Consequences of disclosure on others

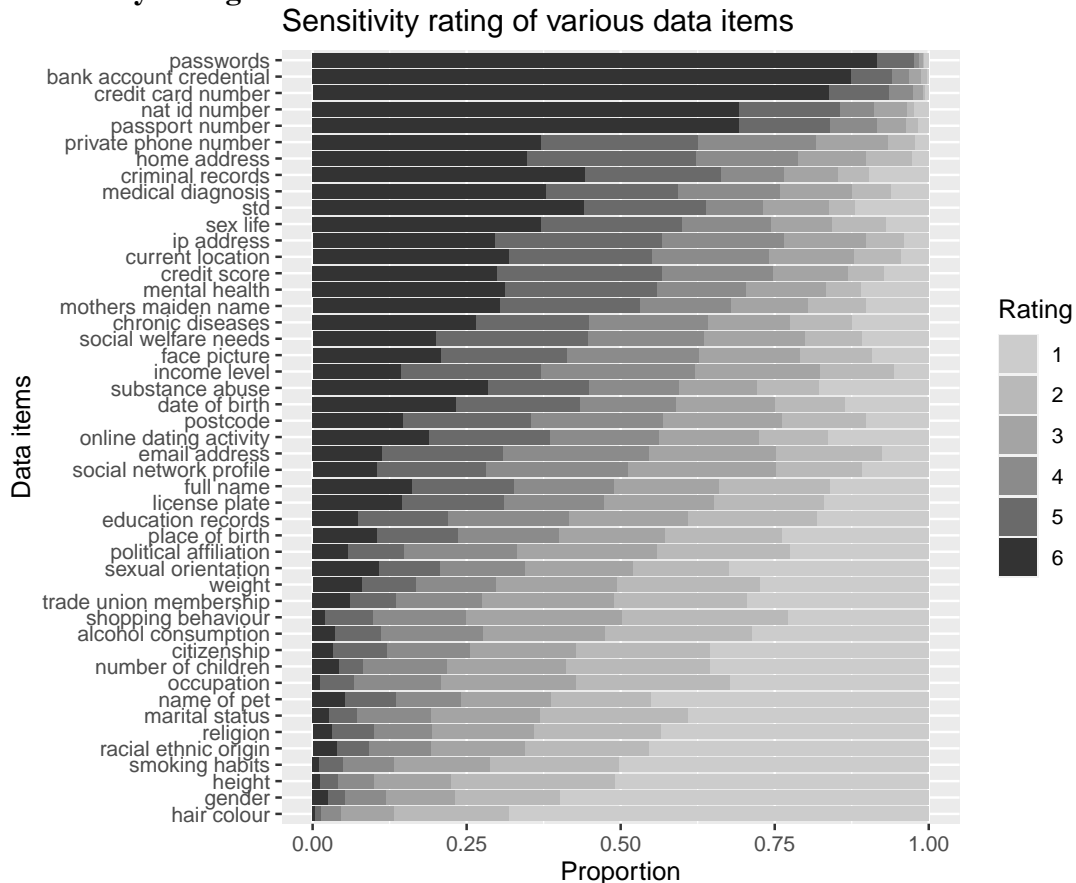
469 In addition to the previous concerns associated with the personal consequences, several responses showed
470 a more altruistic concern. They reported considering *Consequences of disclosure on others* while evaluating
471 the sensitivity of data items. They expressed their concerns regarding the *security and safety of their*
472 *families or beloved ones*. They perceived information sensitive that could cause a risk to the security
473 and safety of others. We have combined the generic concerns under the code *Impact on others* where
474 participants provided their concerns without explicitly defining the impact. Most of these respondents
475 stated that they would not share any information that would put people they know in trouble and consider
476 these data items sensitive.

477 4.2 RO2: Sensitivity rankings of various data items

478 Beyond the factors that are taken into account while assessing the sensitivity of the information, we asked
479 participants to rate 40 data items on a 6-point symmetric Likert scale from ‘not sensitive at all’ (1) to ‘very
480 sensitive’ (6).

481 The participants' ratings for each data item are displayed in Figure 2, the data items are ordered by the
 482 average rating. Our results showed that passwords represented the most sensitive data type for UK citizens,
 483 with 92 % of participants giving it the highest rating, followed by *bank account credentials* and *credit card*
 484 *number*, with 87 % and 83 % of respondents giving it the highest rating. The following items are formally
 485 identifiable information, namely national ID number and passport number, which match the concerns given
 486 regarding identity from the first part of the questionnaire. The least sensitive items were hair colour, gender
 487 and height, which are typically observable human characteristics.

Figure 2. Sensitivity ratings of data items.



488 4.3 RO3: Influence of user factors

489 In order to examine the influence of user factors (age group, gender, education) on the perception of
 490 sensitivity, we built a proportional odds logistic regression model for each data type. We identified those
 491 data items which demonstrated a sensitivity that had a statistically significant effect (using a p-value less
 492 than 0.05) from one of these factors.

493 The gender of the respondents was a modulating factor on the perception of the sensitivity of an *income*
 494 *level*, with female respondents typically considering the sensitivity higher than male participants, see
 495 Figure 3. This was also true for *IP address*, *criminal records*, *weight* and *sexually transmitted disease*.
 496 Conversely, male participants considered *smoking habits* and the *number of children* to be more sensitive
 497 than female participants.

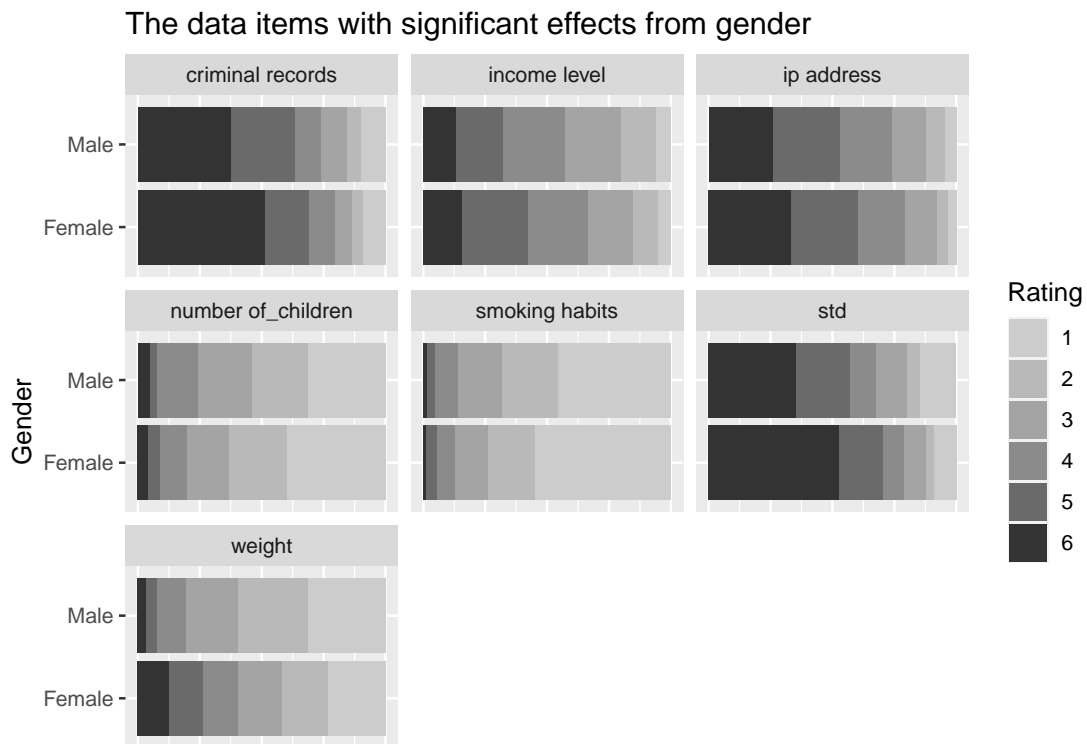


Figure 3. The data items with significant effects from gender.

498 The data items on which education has significant impact are *current location*, *political affiliation* and *sex*
 499 *life*. The level of education led to the sensitivity being perceived as higher for *political affiliation*. Education
 500 also modulated the perceived sensitivity of the *current location* with those who left education before
 501 achieving a post-16 qualification identifying a significantly lower sensitivity, also seen in the sensitivity of
 502 the *sex life* data item. Note, this analysis controlled for the age variable, so this is not an artefact from age
 503 measures.

504 The respondents' age was also observed to have significant effects on perceived sensitivity. The *Credit*
 505 *score* was considered significantly less sensitive by the majority of the participants aged between 18–24.
 506 This age group also tends to consider *date of birth*, *email address* and *mothers maiden name* less sensitive
 507 when compared to other older groups. Looking across these final three data items with factors that have a
 508 relationship with age, there tends to be an increase in sensitivity with age until the 45–54 age group before
 509 decreasing in the 55 plus age group.

510 4.4 RO4: Exploring Cultural Differences via Cluster Analysis

511 We conducted a cluster analysis on the sensitivity of the data items as done by Markos et al. (2017)
 512 and Schomakers et al. (2019). However, we used hierarchical clustering in order to gain a high-fidelity
 513 understanding of the relationship between data items; the result is shown in Figure 4. Using a silhouette
 514 analysis, we found four clusters to be the most appropriate for our data set. Each cluster was cross-
 515 referenced with the ranking in Figure 2 to label the four clusters of data categories (very highly, highly,
 516 medium and low sensitive) as shown in Table 6. Previous work heuristically categorised data items into
 517 three groups as highly, medium and less sensitive. However, our empirical clustering result differentiated a
 518 small group of data types from the other highly sensitive data. We grouped those items under the title of
 519 'Very highly sensitive data' in our categorisation.

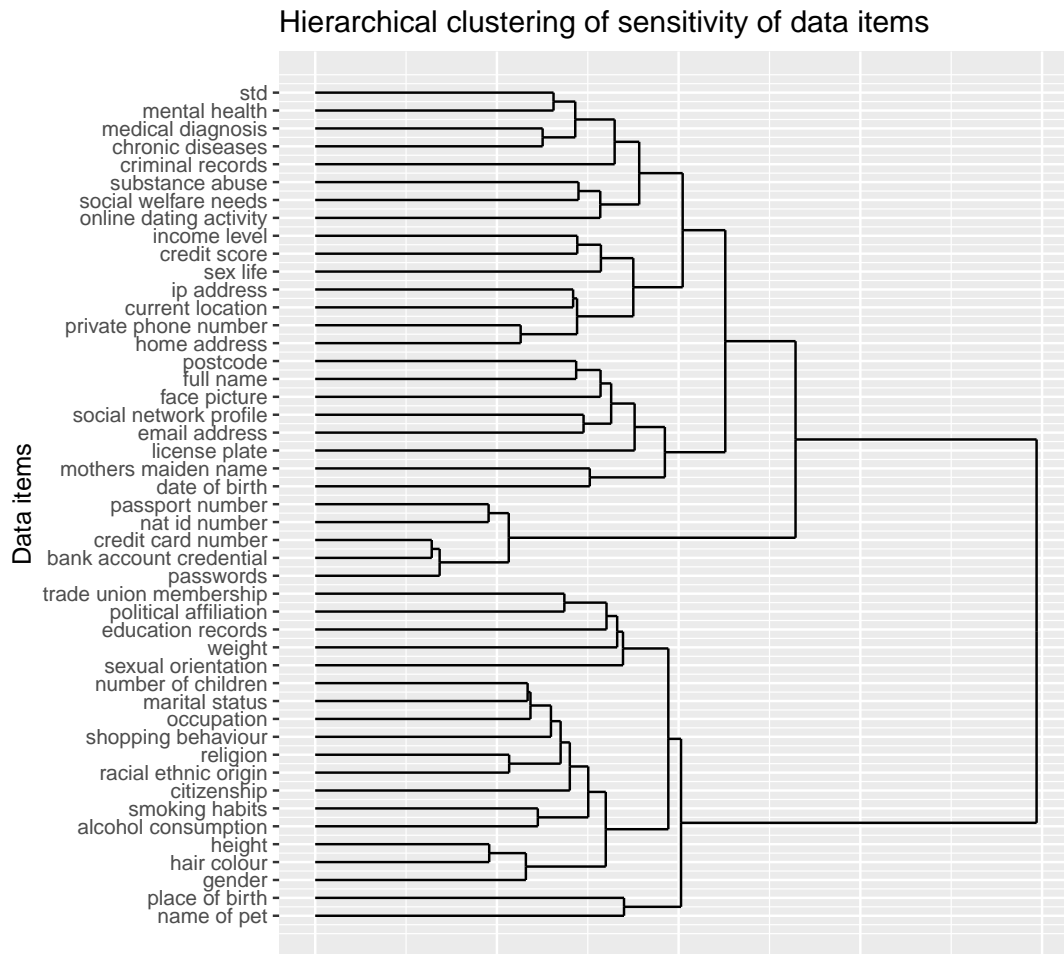


Figure 4. Hierarchical clustering of sensitivity of data items.

520 When previous research compared international measures of data sensitivity (Schomakers et al., 2019)
 521 it was reported that there was only one difference regarding the high sensitivity data category when they
 522 compared their results with Markos et al. (2017), which largely revealed a consensus between three
 523 countries (US, Brazilian and Germany) in this category. We see similar results with data types considered
 524 highly sensitive by those countries also appeared in the same category (or in the ‘Very highly sensitive data’
 525 category) in our study. In our study, several additional items appeared in this category, notably *Income level*,
 526 *current location*, *private phone number*, and *home address* were considered highly sensitive. In contrast,
 527 they belonged to medium or even low sensitive data in the German, Brazilian and US data sets. In our
 528 study, the categorisation for *Credit score* was the same with the Brazilian and US data set, which differs
 529 from the medium sensitivity given by German citizens.

530 Among the items UK citizens placed in a medium sensitive data category, five items (*mothers maiden*
 531 *name*, *license plate number*, *email address*, *social network profile*, *face picture* and *post code*) were in the
 532 low sensitivity data types for German citizens. However, *mothers maiden name*, *social network profile* and
 533 *face picture* were medium sensitive not only for UK citizen but also for US and Brazilian citizens. The
 534 vehicle license plate number appeared in the medium category in our results yet was considered highly
 535 sensitive by US and Brazilian citizens and low by German citizens. The categorisation of the postcode and
 536 email address was identical across all nationalities.

Table 6. Clusters of data items based on sensitivity

Very highly sensitive data	Highly sensitive data	Medium sensitive data	Low sensitive data
Passwords	Private phone number	Date of birth	Name of pet
Bank account credential	Home address	Mothers maiden name	Place of birth
Credit card number	Current location	Licence plate number	Gender
National id number	IP address	Email address	Hair colour
Passport number	Sex life	Social network profile	Height
	Credit score	Face picture	Alcohol consumption
	Income level	Full name	Smoking habits
	Online dating activity	Post code	Citizenship
	Social welfare needs		Racial ethnic origin
	Substance abuse		Religion
	Criminal records		Shopping behaviour
	Chronic diseases		Occupation
	Medical diagnosis		Marital status
	Mental health		Number of children
	Sexually transmitted disease		Sexual orientation
			Weight
			Education records
			Political affiliation
			Trade union membership

537 It is possible to report an international consensus on the low sensitive data items. Almost all data types in
 538 this category in our study were ranked into the same category as previous studies. The only difference is
 539 *sexual orientation* which was given a medium sensitive by German citizens.

540 **4.5 RO5: Impact of the context on information disclosure**

541 The initial analysis focusing on the relationship between context and comfort in disclosing information is
 542 largely in agreement with the literature. The size of the effects is the largest seen in the study. The analysis
 543 of the data items across all scenarios is shown in Figure 5. In this figure, a positive model effect shows
 544 participants being more comfortable disclosing in a health context and a negative model effect showing
 545 participants being more comfortable disclosing in a finance context.

546 There is a clear separation between the information domain and the disclosure domain, with all finance
 547 information showing negative model effects (more comfort in disclosing within a finance domain); however,
 548 there are noteworthy data items with smaller effects. There was a statistically significant effect on ethnic
 549 origin and religion where participants were more comfortable disclosing this within a health context than
 550 in the finance context. Also of interest is the small but significant effect on disclosing a criminal record;
 551 participants were more comfortable disclosing in the finance domain. However, this could be related to
 552 regulations surrounding the requirement for accurate disclosure of information in such cases.

553 Following a similar analysis to the previous section, we considered the pairwise comparison between
 554 scenarios S1 and S2, S3 and S4, and S5 and S6 (from Table 2). This results in the measures of the effect of
 555 the domain in three different scenarios: disclosing anonymously to a chatbot, disclosing anonymously to a
 556 human, and disclosing non-anonymously to a human. The effect of domain across the data items is shown
 557 in Figure 6.

558 This scenario-centred analysis clearly shows the strength of the domain effect. The domain effect is
 559 common throughout all interaction means and degrees of anonymity. An analysis of the models shows no

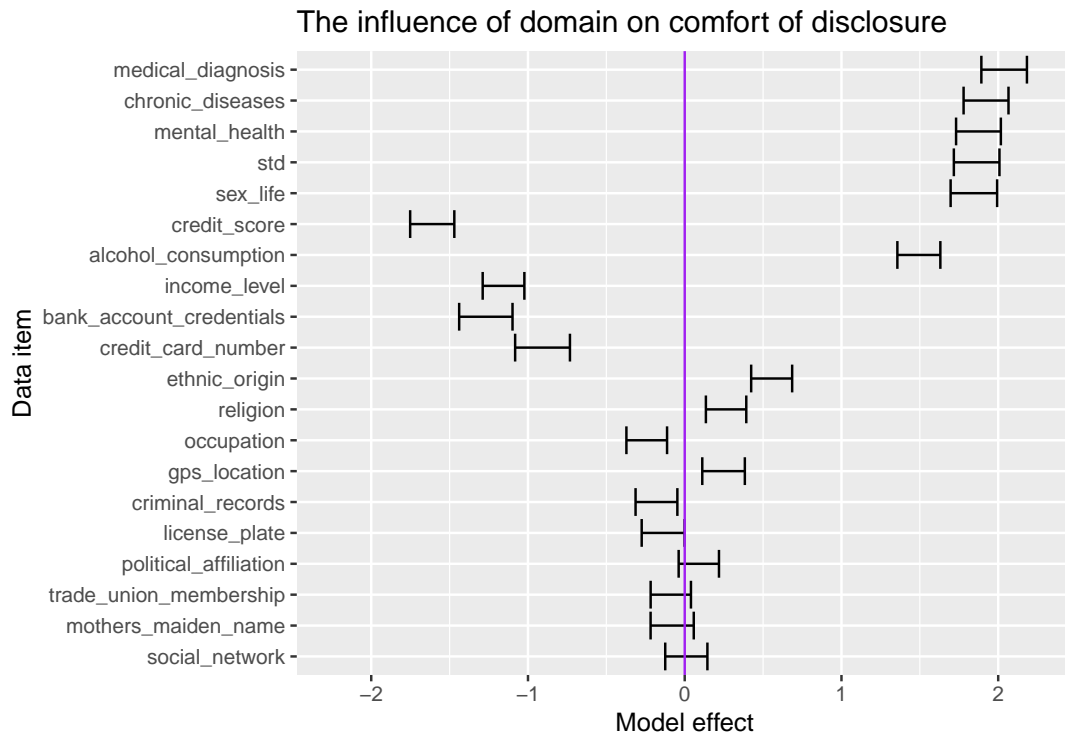


Figure 5. The influence of domain/context.

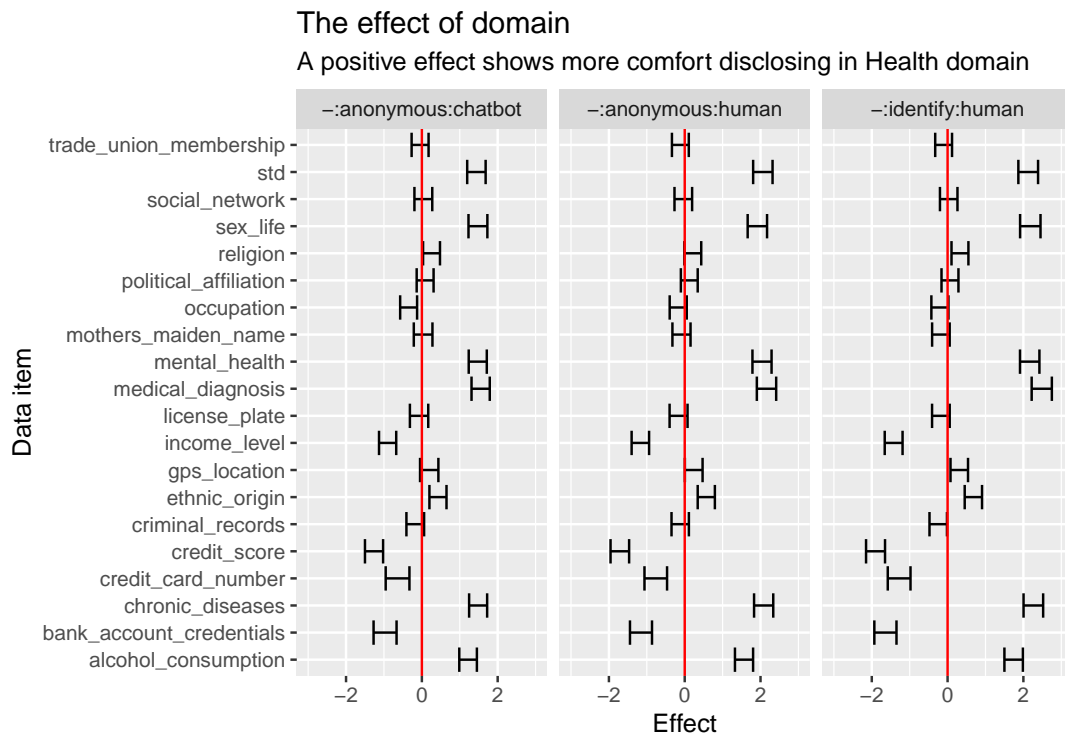


Figure 6. The influence of context across different scenarios of information disclosure.

560 data items where this domain effect is modulated by interaction or anonymity, and there seems to be no
 561 mechanism to significantly override or reduce this effect.

562 4.6 RO6: Impact of interaction means on information disclosure

563 The sixth research objective focused on the interaction means that elicited the disclosure; the model
 564 coefficients from the analysis of each data item are shown in Figure 7. Nearly two-thirds of the data items
 565 show a positive model coefficient (at a 95% confidence), indicating participants were more comfortable
 566 disclosing to a human than a chatbot. There were no data items that participants preferred to disclose
 567 to machines rather than humans. There was no effect from any of the biographic measures (such as age,
 568 gender and education).

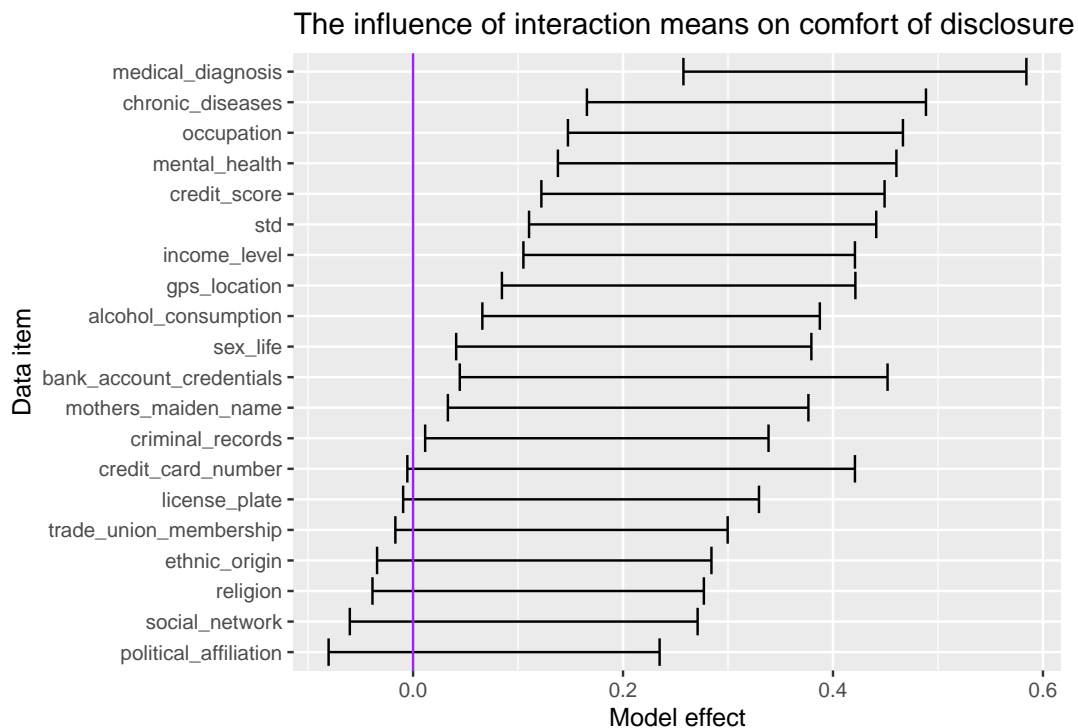


Figure 7. The influence of interacting with a human or chatbot on comfort of disclosure.

569 Using the same modelling approach, we compared the impacts of interaction while disclosing personal
 570 information in health and finance anonymously. To achieve this, we paired the data from scenarios S1
 571 and S5 and scenarios S2 and S6 (shown in Table 2). We then created a multinomial logistic regression to
 572 predict the perceptions of the sensitivity of a data item as a function of the interaction means (chatbot or
 573 human). The model coefficients are shown in Figure 8, with a positive effect being related to more comfort
 574 in disclosing to a human than to a chatbot (the error bounds represent the 95% confidence limit).

575 From these results, we observe that participants felt more comfortable disclosing sensitive information
 576 to humans, particularly in the health context. Sexually transmitted diseases, sex life, mental health,
 577 medical diagnosis or chronic diseases are data items that were preferred to be disclosed to a human by our
 578 participants. However, we can interpret this as preferring to talk to real people rather than chatbots when
 579 they need empathy and rapport in the dyadic.

580 Within the finance domain, only the credit score and income level data items showed a significant effect
 581 (with a 95% confidence) with interaction means. We can argue that using a chatbot will have a more
 582 negligible effect on the disclosures we would expect to be made within the finance domain.

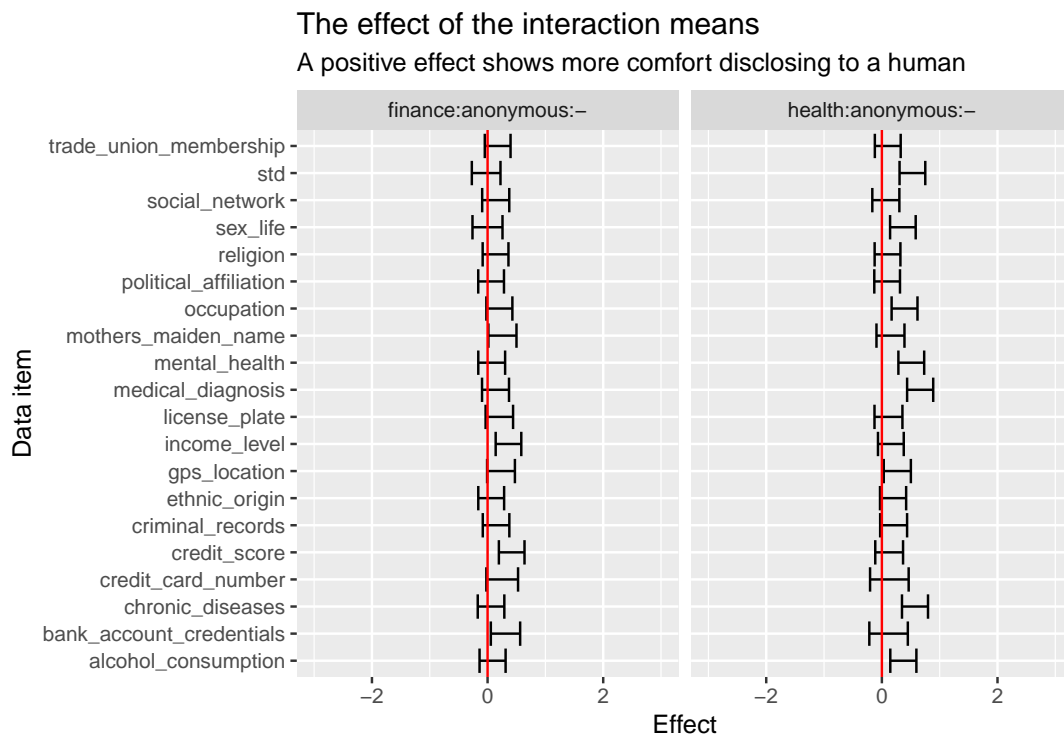


Figure 8. The influence of interaction means across different scenarios.

583 4.7 RO7: Impact of anonymity on information disclosure

584 This analysis considered the effect of anonymity on the disclosure of sensitive information. The logistic
585 regression model coefficients are shown in Figure 9. A positive model effect related to greater comfort
586 in disclosing when non-anonymous (i.e., the individual is identified) and a negative model coefficient
587 demonstrates greater comfort in disclosing when the participant was anonymous.

588 The effect of anonymity is much smaller than other factors in this analysis. However, it does provide
589 statistically significant effects for several data items, most notably sex life and sexually transmitted disease.
590 Interestingly, this also includes political affiliation and alcohol consumption.

591 Two data items that showed a positive model effect (more comfortable in disclosing when done non-
592 anonymously) were the mother's maiden name (something intuitively related to identity) and bank account
593 credentials.

594 Considering the scenario-specific evaluation, we paired scenarios S1 and S3, and S2 and S4 to identify
595 the effect of anonymity within the two contexts when disclosing to a human. The model effect is shown in
596 Figure 10 with a positive model coefficient being related to more comfort in disclosing when identified a
597 negative effect coming from more comfort in disclosing when anonymous.

598 From these results, we can see a small effect from anonymity across the two scenarios. Within the
599 health domain, there is a small effect associated with the sex life data item, but broadly there are very few
600 significant effects associated with this domain. When considering the finance domain in Figure 10 there are
601 minor effects associated with some data items noted in the previous broader analysis. There is also a small
602 negative effect associated with the disclosures associated with sex life in the finance domain; however, this
603 is an out of domain disclosure whilst significant, this is likely to be an unusual disclosure.

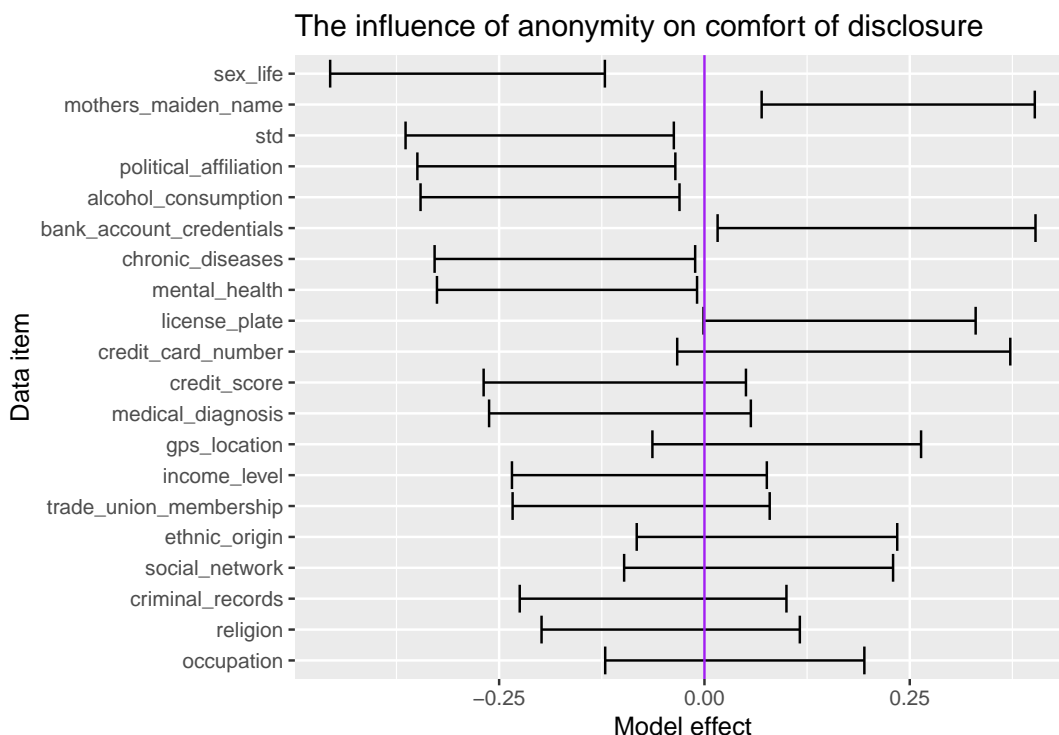


Figure 9. The influence of anonymity on information disclosure.

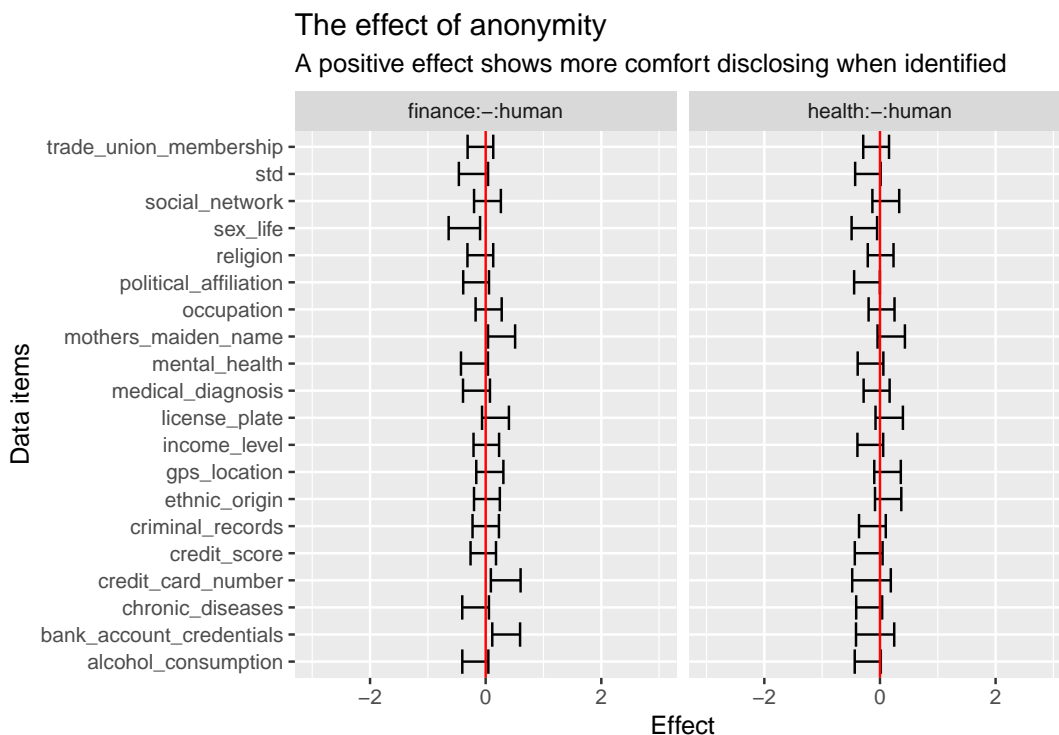


Figure 10. The influence of anonymity across different scenarios of information disclosure.

5 DISCUSSION

604 In this section, we summarise and discuss our key findings for each research objective outlined previously.
 605 Furthermore, we consider the novelty of this work as compared to existing research in the field.

606 5.1 The factors that make information sensitive for UK citizens (RO1)

607 The first research objective was to investigate the primary factors that lead British citizens to regard
608 information as sensitive. Our findings demonstrate that there are three key general topics of note; concerns
609 about the potential consequences of disclosure (this relates to themes *privacy, financial problems, reactions,*
610 *consequences of disclosure on me, consequences of disclosure on others*), the fundamental nature of
611 the information (themes *context, nature of information*), and concerns regarding the person/system the
612 information is shared with (theme *interaction means*).

613 For those with privacy concerns, the main code identified was identity theft. Identity theft, the act of
614 obtaining sensitive information about another person without their knowledge, and using this information
615 to commit theft or fraud, is estimated to cost the UK around £190 billion every year (National Crime
616 Agency, 2021). CIFAS, a UK-based Fraud Prevention Services, stated that in 2019, more than 364,000
617 cases of fraudulent conduct were recorded on their National Fraud Database with an increase of 13 per cent
618 compared to 2018 (CIFAS, 2019). It is promising to observe the degree of awareness of this risk within the
619 UK population; acknowledging that awareness is only the first step to prevention.

620 In addition, we identified several participants' decision-making was related to financial implications,
621 with concerns regarding financial loss being one of the significant codes that emerged from the qualitative
622 analysis. Those findings are reinforced by the items which received the highest sensitivity ratings in the
623 quantitative phase of the study. The bank account credential, credit card number appeared in the top three
624 most sensitive items (see Figure 2). They also confirm prior study which reported the possibility of harm as
625 one of the main factors considered when assessing sensitivity (Ohm, 2014).

626 Our results also uniquely highlight another concern that is generally overlooked by the privacy studies or
627 regulations: disclosure of information belonging to others and impacts on personal information disclosure
628 on others. Responses revealed that some participants consider information sensitive if this information
629 belongs to others. Personal information studies in the literature are generally self-disclosure studies where
630 the information is assumed to belong to the participant. It is also the same for the sensitivity studies where
631 the owner of the information is assumed to be the person whose opinion or behaviour is observed. Our
632 analysis identifies concerns regarding both data belonging to others and the effect of information disclosure
633 on others, particularly the potential harms to others. This observation indicates a societal maturity in
634 identifying the second-order effects of disclosure.

635 As seen in Figure 2, personal data items categorised in a special category by the GDPR were not
636 identified as being sensitive by our participants. We can identify the sensitivity of political affiliation,
637 sexual orientation and trade-union membership as similar and not regarded as very sensitive; for example,
638 a similar ranking was exhibited by weight and a much lower ranking than, for instance, income level or
639 credit score. More interestingly, religion and ethnic origin were considered a very low sensitivity similar
640 levels as marital status or occupation. Here it is worthy of note that, as mentioned before, this research
641 aims to provide a British perspective on information sensitivity. It is well-understood that the perceived
642 sensitivity of a particular type of data varies widely, both between societies or ethnic groups and within
643 those groups (Rumbold and Pierscionek, 2018). The agency individuals have to protect their data, and
644 hence the vulnerability of the individuals data affect the perceived sensitivity. Some of the data items
645 categorised as special category by the GDPR (e.g., racial or ethnic origin or religion) may well have
646 attracted higher sensitivity rankings if this study was constrained to minority ethnic groups rather than the
647 general public.

648 5.2 Influences of user factors on perceived sensitivity (RO2)

649 Our study also allowed us to identify variability in the perceptions of the sensitivity of data items based
650 on the data subjects biographic information. For example, when we considered the age of the data subject,
651 we found several interesting effects. Our findings are partially consistent with the literature that generally
652 report that younger age groups share more information and are less concerned about information privacy,
653 (e.g. Miltgen and Peyrat-Guillard, 2014; Van den Broeck et al., 2015). It is also consistent with the literature
654 that privacy is the most common barrier for older people to use smart technologies (Harris et al., 2022).

655 We can enrich those findings with fine-grained data items; for example, ‘credit score’ was ranked less
656 sensitive by those under 25. We hypothesise that this is because this group do not normally require high
657 credit levels (for example, purchasing a house) and hence are unlikely to be discriminated against based on
658 that level. The same can be said of date-of-birth, which steadily becomes more sensitive during working age
659 until retirement when it becomes less sensitive. Again there is a clear parallel with discrimination within
660 the workplace. We believe that our detailed findings can help develop individually tailored information
661 collection systems that recognise and respect different privacy concerns among different demographics
662 groups.

663 The final two data items that show an effect with age are email address and mothers maiden name,
664 both of which show a low sensitivity for 18–24 years with a higher level across the other age groups but
665 with a peak in the 45–54 cohort. The reduced level of sensitivity associated with young people can be
666 explained by the peak in the group representing Xennials or late Gen X who had an analogue childhood
667 but digital adulthood and have retained some of the understanding of the formative years of digital life.
668 Older participants potentially have come to digital life when the internet and digital socialisation norms are
669 more formed rather than growing up alongside the transformation.

670 When it comes to the impact of education levels on perceived information sensitivity, we found several
671 conflicting findings in the literature. While there are studies that claim that individuals with lower
672 educational levels tend to be less concerned about their personal information, (e.g. Blank et al., 2014;
673 Rainie et al., 2013), there are also those which report no differences in privacy concerns depending on
674 education levels (Li, 2011). Our study highlights that differences in the perception of sensitivity based on
675 education are only prevalent regarding some information types (e.g., current location, political affiliation
676 and sex life). Within the education level, there does appear to be a breakpoint between those who achieved
677 post-16 education, most notably in location and sex life; note this has been controlled for participant age.

678 The final biographic element we explored was the effect of gender on perceptions of sensitivity. Gender
679 provided the largest number of data items that were modulated by this factor. Our study identified an
680 apparent social stigma that female participants felt when disclosing criminal records, sexually transmitted
681 diseases, and weight. We can also explain the higher perceived sensitivity rating of *income level* in female
682 participants by cultural factors, which can be different in a more patriarchal society. Even though the
683 UK is one of the countries where the lowest levels of legal discrimination are measured against women
684 (Georgetown University’s Institute for Women, Peace and Security, 2020) there is still a disconnect between
685 the genders in terms of pay, and it naturally follows that there is a difference in the perceived sensitivity.

686 Our results appear to support Knijnenburg et al. (2013) who hypothesised that information disclosure
687 behaviours consisted of multiple related dimensions and disclosure behaviours do not differ among groups
688 overall, but rather in their disclosure tendencies per type of information. The results are also consistent
689 with the results from RQ1.

690 **5.3 UK perspective on the sensitivity of the different data items and identification of** 691 **cultural differences (RO3 and RO4)**

692 Our results confirmed the consensus on the high perceived sensitivity of the finance-related information
693 and identifiers, which appeared in the same category as Markos et al. (2017) and Schomakers et al. (2019).
694 When we reflect on the least sensitive items (hair colour, gender, height), the common feature is that they
695 are typically visible to the public. These appear consistent with the hypothesis from Markos et al. (2018)
696 who predicted that public information is considered less sensitive compared to private-self information
697 (inner states, personal history, and specific features of the self).

698 We conclude a degree of consensus on what constitutes sensitivity across German, US, Brazilian and UK
699 citizens. However, respondents in our study and our rigorous empirical approach identified several ‘very’
700 highly sensitive data items that formed a discrete cluster above those seen in the other studies. We also
701 saw several elements promoted to the high-sensitivity cluster (e.g., income level, private phone number)
702 compared to other nations, even compared to another western European country. This discontinuity shows
703 that whilst international regulatory frameworks are undoubtedly essential to provide a degree of data
704 protection, we must also have mechanisms to support the cultural differences within individual nations.
705 Considering the internationalised nature of today’s information society, we believe that such findings are
706 important to consider while designing information systems that allow trans-border data flows, or for those
707 systems designed and built in a different socio-economic environment to which they will be deployed.

708 **5.4 Impact of the context on information disclosure (RO5)**

709 Our fifth Research Objective focused on the effect of context on the comfort of disclosing information.
710 Our results broadly align with the literature; however, we highlight the magnitude of this effect; the strength
711 of this effect is nearly ten times greater than any other identified in the study. Figure 5 clearly shows that
712 health-related information is shared with significantly more comfort in a health context. Similarly, the
713 finance-related information is shared more comfortably in a finance context. Also interesting were the
714 data items related to religion and ethnic origin, which exhibited significant preferences for disclosure in
715 the medical domain. It is conceivable that ethnic origin may result in a predisposition to certain illnesses
716 (Cooper, 2004) and justifies a disclosure in the health domain; it is unlikely that the same is true in the
717 financial domain. The effect of context is also not mediated by the scenario and appears to be consistent
718 whether disclosing anonymously to a human or a chatbot or disclosing non-anonymously to a human; this
719 is shown in Figure 6. These findings confirm the impact of relevance on the perceived sensitivity. From a
720 regulatory perspective, this could be interpreted as a clear validation of the *data minimisation principle* of
721 the GDPR, which requires data collection to be adequate and limited to what is necessary.

722 **5.5 Interaction means and comfort to disclose (RO6)**

723 Our penultimate research objective (RO6) focused on the interaction means whether the disclosure was
724 direct to a human or through a chatbot mediated communication. In general, we found participants were
725 more comfortable disclosing directly to a human rather than a chatbot; this was particularly the case with
726 medical diagnosis, chronic diseases and mental health issues, shown in Figure 7. This preference for
727 face-to-face human reporting has been seen in many sensitive domains, for example, within community
728 reporting associated with violent extremism (Thomas et al., 2020). In these cases, it is very often difficult
729 for the individual to make the disclosure. The natural interaction between humans and the perception of
730 control is essential to support and enable these disclosures.

731 When this interaction means is considered in the scenario-specific conditions, we see a slightly more
732 complicated picture. Within the health-based scenario, our participants still prefer disclosing to a human
733 over a chatbot. Again the locus of control and the perception of engaged feedback may encourage
734 participants to be more comfortable disclosing to a human. The other data item that showed a preference
735 was occupation. Those findings contradict with the literature where users were reported to prefer chatbots
736 or to respond with more disclosure intimacy to chatbots than a human (Ho et al., 2018; Bjaaland and
737 Brandtzaeg, 2018). We can hypothesize at this point that within a healthcare setting, the perception of
738 discussing and enriching the disclosure and providing more background as to the day-to-day tasks may
739 drive this preference. When we consider the finance scenario, we generally see little difference between
740 disclosure to a human or a chatbot. An indication that sensitive disclosures in this domain are less likely
741 to be reduced through the use of conversational agents. The only data items that showed a significant
742 effect were the credit score and income level; similarly to the occupation data item within the healthcare
743 setting, we believe that this is a disclosure that the participant may view as requiring more enrichment or
744 explanation. Hence, a factual disclosure with no interaction or feedback may be perceived as less desirable,
745 leading to a perception of more comfort in disclosing to a human.

746 **5.6 Anonymity and comfort to disclose (RO7)**

747 The final research objective (RO7) focused on the effect of anonymity on the person making the disclosure.
748 When considered abstractly, it was clear that several data items demonstrated a preference for anonymous
749 disclosure, such as sex life and sexually transmitted diseases and alcohol consumption and political
750 affiliation, which is inline with the previous findings (Schomakers et al., 2019). This observation would
751 appear to match well to the qualitative results as well, which suggested that the reaction of others was an
752 important element when judging whether items were sensitive or not.

753 As with the previous research objective, when this is contextualised within a real scenario, the results are
754 more nuanced. We can see from Figure 10 that there is no preference for anonymity within the healthcare
755 setting — nearly all data items showed no significant difference in the comfort with being anonymous or
756 identified. We have already demonstrated the strength of the context in the sensitivity of disclosures. We
757 would suggest that the healthcare context and the professional reputation of the National Health Service
758 in the UK lead to participants seeing no value in being anonymous. The only data item that showed a
759 preference for anonymous disclosure was associated with sex life, which was only just significant at the
760 95% level.

761 When considering the finance domain, several preferences for anonymity were observed; these were
762 mostly tied to disclosures related to health, although these effects are minor and only just significant. Hence
763 it is difficult to draw a meaningful conclusion from this domain; however, it may hint that when disclosures
764 are made out of domain, individuals may be more comfortable disclosing if anonymous.

765 **6 CONCLUSION**

765 This final section draws together our research contributions from our rigorous analytical study of this
766 challenging problem.

767 **6.1 Theoretical Contributions**

768 Our study presents a detailed capture of the perspective of UK citizens regarding the sensitivity of
769 personal information. Three main factors lead British citizens to assign higher sensitivity scores to data

770 items; consequences of disclosure, nature of the information and the concerns regarding with whom the
771 person/system the information is shared. Identity theft and financial loss are the main concerns of the
772 individuals, which is consistent with the risk-based definition of sensitive personal information in regulatory
773 documents. In addition, high sensitivity scores assigned to health and financially related information indicate
774 that there is a consensus on what constitutes sensitivity across German, the US, Brazilian and the UK.
775 However, British citizens regard some items as highly sensitive as compared to the other three countries.
776 These discrepancies highlight the challenge of providing trans-national regulation and should be noted by
777 those managing information security where data flows cross regulatory borders.

778 We also identified individual characteristics that modulate perceptions of sensitive data. We identified
779 age, gender and education level as influencing the sensitivity of particular data items; these modulating
780 characteristics mapped well to the qualitative explanations of the factors that made data items sensitive.

781 The context or the fairness of the request has the most significant impact on the comfort level felt while
782 disclosing personal information. Disclosure of highly sensitive personal information such as sex life,
783 sexually transmitted disease or alcohol consumption was observed to be affected by anonymity. Participants
784 reported disclosing those items with significantly more comfort when they do not have to reveal their
785 identities.

786 This study has developed a systematic understanding of UK citizens' perceptions of sensitive information,
787 showing a degree of consensus with previous studies and some unique insights. We particularly note the
788 effect of the relevance of the disclosure and the effect of the interaction means, whether a human-mediated
789 disclosure or a disclosure mediated by a conversational agent. In general, we highlighted the preference
790 to disclose sensitive personal information to a human rather than a conversational agent. These findings
791 should be considered in the design and management of information within systems that involve sensitive
792 disclosures and hence sensitive data, particularly in the healthcare domain, where our findings are most
793 significant.

794 **6.2 Managerial Contributions**

795 We contribute to the literature by investigating the impact of emerging technologies, particularly
796 conversational agents (or chatbots), on the disclosure of personal data. Such disclosure is a key security
797 concern for both those disclosing their data and for organisations seeking to facilitate accurate, high-
798 integrity disclosures. Despite the existence of studies that show the facilitator role of chatbots on information
799 disclosure, no study, to our knowledge, has evaluated the perceived sensitivity of data items at granular
800 level when they are disclosed to a chatbot. We also consequently identify the contexts where chatbots can
801 enable individuals to disclose sensitive information more comfortably. In addition to providing general
802 insights into how persons in the UK perceive sensitive information, our findings can contribute to the
803 design of chatbots; most notably, defining an evidence-base to support agent use in the most appropriate
804 usage contexts increasing the comfort of disclosing and ultimately ensuring more accurate responses.

805 We specifically investigate two main contexts in our research; health and finance. These contexts have a
806 regulatory demand for high levels of security and data protection, and are traditionally where chatbots are
807 heavily adopted and sensitive personal information is frequently collected and processed (Ng et al., 2020;
808 Stiefel, 2018). Our findings help demonstrate the relationship between the disclosed personal information
809 and the context in which it is disclosed, ultimately uncovering the impact of usage context on disclosure of
810 different data items. Finally, we explore the effect of anonymity, specifically identifying what personal
811 data the UK public prefer to disclose anonymously. These observations provide novel insights for the

812 information collection systems used in the UK by uncovering the factors that lead to perceptions of high
813 sensitivity and hence the comfort (or discomfort) in the disclosure process.

814 6.3 Limitations and Future Work

815 While we believe our study was robust and has made several substantial contributions to the research,
816 some limitations must be acknowledged. Firstly, our results represent self-reported sensitivity evaluations
817 and may not reflect the lived behaviours of our participants. However, this approach allowed us to obtain and
818 compare several sensitivity evaluations across several contexts. It also compares well with previous works
819 in the field (e.g. Schomakers et al., 2019; Markos et al., 2017)), which followed a similar methodological
820 approach. However, we are aware that it might be possible to collect more accurate results when the
821 participants assess their comfort levels while practising the given scenarios.

822 Consequently, to validate our findings, our next step will explore the disclosure behaviours in an
823 experimental context involving both human and chatbot mediated disclosures. Another issue faced in this
824 study is the vagueness regarding the benefits of the disclosure and the perceived risk/trust to the interaction
825 means. In our experimental approach, we intend to ensure a clear and consistent perception of the benefit
826 of disclosure.

827 We also removed two scenarios from our 2x2x2 study; this meant that we could not fully explore all
828 combinations of factors. However, this pragmatic decision has significantly improved the quality of the
829 results and allowed us to draw some robust conclusions from the remaining six scenarios. Future work
830 could consider the value in exploring all scenarios and thereby fully understanding all factors.

ACKNOWLEDGMENT

831 This research was conducted as a part of a UK EPSRC-funded project, A Platform for Responsive
832 Conversational Agents to Enhance Engagement and Disclosure (PRoCEED) (Grant number: EP/S027297/1
833 and EP/S027211/1).

REFERENCES

- 834 Ackerman, M. S., Cranor, L. F., and Reagle, J. (1999). Privacy in e-commerce: examining user scenarios
835 and privacy preferences. In *Proceedings of the 1st ACM conference on Electronic commerce*. 1–8
- 836 Aiello, G., Donvito, R., Acuti, D., Grazzini, L., Mazzoli, V., Vannucci, V., et al. (2020). Customers'
837 willingness to disclose personal information throughout the customer purchase journey in retailing: the
838 role of perceived warmth. *Journal of Retailing* 96, 490–506
- 839 Bansal, G., Gefen, D., et al. (2010). The impact of personal dispositions on information sensitivity, privacy
840 concern and trust in disclosing health information online. *Decision support systems* 49, 138–150
- 841 Bansal, G., Zahedi, F. M., and Gefen, D. (2016). Do context and personality matter? trust and privacy
842 concerns in disclosing private information online. *Information & Management* 53, 1–21
- 843 Belen Sağlam, R. and Nurse, J. R. C. (2020). Is your chatbot GDPR compliant? Open issues in agent
844 design. In *Proceedings of the 2nd Conference on Conversational User Interfaces*. 1–3
- 845 Belen Sağlam, R., Nurse, J. R. C., and Hodges, D. (2022). Personal information: Perceptions, types and
846 evolution. *Journal of Information Security and Applications* 66, 103163. doi:10.1016/j.jisa.2022.103163
- 847 Bell, S., Wood, C., and Sarkar, A. (2019). Perceptions of chatbots in therapy. In *Extended Abstracts of the*
848 *2019 CHI Conference on Human Factors in Computing Systems*. 1–6

- 849 Bhakta, R., Savin-Baden, M., and Tombs, G. (2014). Sharing secrets with robots? In *EdMedia+ Innovate*
850 *Learning* (Association for the Advancement of Computing in Education (AACE)), 2295–2301
- 851 Bjaaland, M. and Brandtzaeg, P. (2018). *Youth and News in a Digital Media Environment* (Nordicom),
852 chap. Chatbots as a new user interface for providing health information to young people. 59–66
- 853 Blank, G., Bolsover, G., and Dubois, E. (2014). A new privacy paradox: Young people and privacy on
854 social network sites. In *Prepared for the Annual Meeting of the American Sociological Association*.
855 vol. 17, 1–35
- 856 Braun, V. and Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology*
857 3, 77–101
- 858 Bridges Jr, C. C. (1966). Hierarchical cluster analysis. *Psychological reports* 18, 851–854
- 859 [Dataset] CIFAS (2019). Annual report 2019. [https://www.cifas.org.uk/about-cifas/
860 annual-reports/annual-report-2019](https://www.cifas.org.uk/about-cifas/annual-reports/annual-report-2019)
- 861 Cooper, R. S. (2004). Genetic factors in ethnic disparities in health. *Critical Perspectives on Racial and*
862 *Ethnic Disparities in Late Life* 267, 269–309
- 863 [Dataset] European Parliament (2016). Regulation (EU) (2016) 2016/679 of the European Parliament and
864 of the Council of 27 April on the protection of natural persons with regard to the processing of personal
865 data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection
866 Regulation). Official Journal of the European Union 59(L 119)
- 867 [Dataset] Georgetown University's Institute for Women, Peace and Security (2020). Women peace
868 and security index 2019/20. [https://giwps.georgetown.edu/wp-content/uploads/
869 2019/12/WPS-Index-2019-20-Report.pdf](https://giwps.georgetown.edu/wp-content/uploads/2019/12/WPS-Index-2019-20-Report.pdf)
- 870 Harris, M. T., Rogers, W. A., and Blocker, K. A. (2022). Older adults and smart technology: Facilitators
871 and barriers to use. *Frontiers in Computer Science*, 41
- 872 Ho, A., Hancock, J., and Miner, A. S. (2018). Psychological, Relational, and Emotional Effects of
873 Self-Disclosure After Conversations With a Chatbot. *Journal of Communication* 68, 712–733
- 874 Ioannou, A., Tussyadiah, I., and Lu, Y. (2020). Privacy concerns and disclosure of biometric and behavioral
875 data for travel. *International Journal of Information Management* 54, 102122
- 876 Keith, M. J., Thompson, S. C., Hale, J., Lowry, P. B., and Greer, C. (2013). Information disclosure on
877 mobile devices: Re-examining privacy calculus with actual user behavior. *International journal of*
878 *human-computer studies* 71, 1163–1173
- 879 Kim, D., Park, K., Park, Y., and Ahn, J.-H. (2019). Willingness to provide personal information: Perspective
880 of privacy calculus in iot services. *Computers in Human Behavior* 92, 273–281
- 881 Knijnenburg, B. P., Kobsa, A., and Jin, H. (2013). Dimensionality of information disclosure behavior.
882 *International Journal of Human-Computer Studies* 71, 1144–1162
- 883 Kolan, A., Tjoa, S., and Kieseberg, P. (2020). Medical blockchains and privacy in Austria - technical and
884 legal aspects. In *Proceedings of the 2020 International Conference on Software Security and Assurance*
885 (IEEE), 1–9. doi:10.1109/ICSSA51305.2020.00009
- 886 Levallois-Barth, C. and Zylberberg, H. (2017). A purpose-based taxonomy for better governance of
887 personal data in the internet of things era: The example of wellness data. In *Data Protection and*
888 *Privacy:(In) visibilities and Infrastructures* (Springer). 139–161
- 889 Li, Y. (2011). Empirical studies on online information privacy concerns: Literature review and an integrative
890 framework. *Communications of the Association for Information Systems* 28, 28
- 891 Lozano, L. M., Garcia-Cueto, E., and Muñiz, J. (2008). Effect of the number of response categories on the
892 reliability and validity of rating scales. *Methodology* 4, 73–79. doi:10.1027/1614-2241.4.2.73

- 893 Malheiros, M., Preibusch, S., and Sasse, M. A. (2013). 'Fairly truthful': The impact of perceived effort,
894 fairness, relevance, and sensitivity on personal data disclosure. In *International Conference on Trust and*
895 *Trustworthy Computing* (Springer), 250–266
- 896 Markos, E., Labrecque, L. I., and Milne, G. R. (2018). A new information lens: The self-concept and
897 exchange context as a means to understand information sensitivity of anonymous and personal identifying
898 information. *Journal of Interactive Marketing* 42, 46–62
- 899 Markos, E., Milne, G. R., and Peltier, J. W. (2017). Information sensitivity and willingness to provide
900 continua: a comparative privacy study of the united states and brazil. *Journal of Public Policy &*
901 *Marketing* 36, 79–96
- 902 Milne, G. R., Pettinico, G., Hajjat, F. M., and Markos, E. (2017). Information sensitivity typology: Mapping
903 the degree and type of risk consumers perceive in personal data sharing. *Journal of Consumer Affairs* 51,
904 133–161
- 905 Miltgen, C. L. and Peyrat-Guillard, D. (2014). Cultural and generational influences on privacy concerns: a
906 qualitative study in seven european countries. *European journal of information systems* 23, 103–125
- 907 Murtagh, F. and Contreras, P. (2012). Algorithms for hierarchical clustering: an overview. *WIREs Data*
908 *Mining and Knowledge Discovery* 2, 86–97
- 909 [Dataset] National Crime Agency (2021). Fraud - the threat from fraud. <https://www.nationalcrimeagency.gov.uk/what-we-do/crime-threats/fraud-and-economic-crime>
- 910
911
- 912 Ng, M., Coopamootoo, K. P., Toreini, E., Aitken, M., Elliot, K., and van Moorsel, A. (2020). Simulating
913 the effects of social presence on trust, privacy concerns & usage intentions in automated bots for finance.
914 In *2020 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)* (IEEE), 190–199
- 915 Norberg, P. A., Horne, D. R., and Horne, D. A. (2007). The privacy paradox: Personal information
916 disclosure intentions versus behaviors. *Journal of consumer affairs* 41, 100–126
- 917 Ohm, P. (2014). Sensitive information. *Southern California Law Review* 88, 1125–1196
- 918 Pang, P. C.-I., McKay, D., Chang, S., Chen, Q., Zhang, X., and Cui, L. (2020). Privacy concerns of
919 the australian my health record: Implications for other large-scale opt-out personal health records.
920 *Information Processing & Management* 57, 102364
- 921 Peer, E., Brandimarte, L., Samat, S., and Acquisti, A. (2017). Beyond the turk: Alternative platforms for
922 crowdsourcing behavioral research. *Journal of Experimental Social Psychology* 70, 153–163
- 923 Rainie, L., Kiesler, S., Kang, R., Madden, M., Duggan, M., Brown, S., et al. (2013). Anonymity, privacy,
924 and security online. *Pew Research Center* 5
- 925 Rindfleisch, T. C. (1997). Privacy, information technology, and health care. *Communications of the ACM*
926 40, 92–100
- 927 Rumbold, J. M. and Pierscioneck, B. K. (2018). What are data? A categorization of the data sensitivity
928 spectrum. *Big data research* 12, 49–59
- 929 Schomakers, E.-M., Lidynia, C., Müllmann, D., and Ziefle, M. (2019). Internet users' perceptions of
930 information sensitivity—insights from germany. *International Journal of Information Management* 46,
931 142–150
- 932 Schomakers, E.-M., Lidynia, C., and Ziefle, M. (2020). All of me? Users' preferences for privacy-
933 preserving data markets and the importance of anonymity. *Electronic Markets* 30, 649–665
- 934 Stiefel, S. (2018). 'The chatbot will see you now': Mental health confidentiality concerns in software
935 therapy. *Science and Technology Law Review* 20

- 936 Thomas, P., Grossman, M., Christmann, K., and Miah, S. (2020). Community reporting on violent
937 extremism by “intimates”: emergent findings from international evidence. *Critical Studies on Terrorism*
938 13, 1–22
- 939 Treiblmaier, H. and Chong, S. (2013). Trust and perceived risk of personal information as antecedents
940 of online information disclosure: Results from three countries. In *Global Diffusion and Adoption of*
941 *Technologies for Knowledge and Information Sharing* (IGI Global). 341–361
- 942 Van den Broeck, E., Poels, K., and Walrave, M. (2015). Older and wiser? Facebook use, privacy concern,
943 and privacy protection in the life stages of emerging, young, and middle adulthood. *Social Media+*
944 *Society* 1, 1–11
- 945 Wadle, L.-M., Martin, N., and Ziegler, D. (2019). Privacy and personalization: The trade-off between data
946 disclosure and personalization benefit. In *Adjunct Publication of the 27th Conference on User Modeling,*
947 *Adaptation and Personalization*. 319–324
- 948 Yu, L., Li, H., He, W., Wang, F.-K., and Jiao, S. (2020). A meta-analysis to explore privacy cognition and
949 information disclosure of internet users. *International Journal of Information Management* 51, 102015
- 950 Zheng, X., Mukkamala, R. R., Vatrappu, R., and Ordieres-Mere, J. (2018). Blockchain-based personal health
951 data sharing system using cloud storage. In *IEEE 20th international conference on e-health networking,*
952 *applications and services (Healthcom)* (IEEE), 1–6. doi:10.1109/HealthCom.2018.8531125