Privacy Context Game in Human-Agent Systems

Rokas Gudavičius

4th Year Project Report
Computer Science
School of Informatics
University of Edinburgh

2021

Abstract

Privacy preference knowledge by service providers and individuals themselves is essential for tackling unwanted collection of personal data. This project attempts to improve the standard method of collecting privacy preferences, which are traditionally done through questionnaires, by developing a game that could both collect privacy preference data, and educate users on privacy.

The game presents scenarios in text form and quizzes users on their observations about data collection within them. Standard question formats such as text input are changed to contain parts of scenario text to be improved by the user, and new methods to interact with text are introduced, such as the highlighting feature. These features aim to increase interactivity with the system, and improve engagement in comparison to standard questionnaires. To tackle the educational aspect, a results screen is provided where users are informed of their performance throughout the game with compelling visuals.

Data collected through game playthroughs successfully identified features of scenarios such as disclosed retention time and purpose of collected data having more impact on comfort and consent to data sharing, and generally the nature of privacy concerns prioritised by users.

Acknowledgements

I would like to first and foremost thank my supervisor, Nadin Kökciyan, who kept my spirits up throughout the entire project by remaining consistently positive and calm. I am also very grateful for our weekly meetings that were essential to keep me on track and making consistent progress throughout the year.

Additionally, I would like to thank my flatmates, who were essential in maintaining my mental health throughout the year and were excellent test subjects for early versions of the game.

I thank my parents for tolerating frustrated and tired me in our video calls for the past few months.

Finally, thank you to everyone who played the game and provided their feedback, without you this would not have been possible!

Table of Contents

1 Introduction **1**
1.1 Motivation 1
1.2 Objectives 2

2 Background **3**
2.1 User Privacy Knowledge 3
2.1.1 Privacy Context 4
2.1.2 Context Exploration Through Scenarios 4
2.2 Challenges 5
2.3 Gamification 6

3 Game Design **8**
3.1 Game Features 8
3.2 IUIPC Scale 9
3.3 GUI 10
3.4 Data Model 12
3.5 Ethics 13
3.5.1 User Data 13
3.5.2 Survey Design 13

4 Implementation **15**
4.1 Development Environment 15
4.1.1 Front-End 15
4.1.2 Back-End 15
4.1.3 Database 16
4.2 Game Flow 16
4.2.1 Scenario Pages 17
4.2.2 Privacy Score Questions 23
4.2.3 Results Page 25
4.3 API 27
4.3.1 GET Endpoints 27
4.3.2 POST endpoints 28

5 Evaluation **29**
5.1 Findings on Privacy Preferences 29
5.2 User Experience 32

6 Conclusion **34**
6.1 Future Improvements 35

Bibliography **36**

A Optional Scenarios **39**

B Badges **41**
B.1 Binary Choice Badges 41
B.2 Multiple Choice Badges 43

C Documents **44**
C.1 Participant Information Sheet 44
C.2 Consent Form 47

D Post-Game Questionnaire **51**

Chapter 1

Introduction

Data gathering is getting more frequent, profitable, and ubiquitous with the help of smart and IoT devices. This ever-changing and expanding field makes controlling, or even defining the privacy preferences of individual users more and more challenging. To proceed effectively with helping users manage and enforce their privacy preferences, they first have to be well-defined by the users. The aim of this project is to attempt to find out what users consider as important factors when deciding whether to share their information in distinct situations (or contexts). An additional goal is to make the process of sharing privacy preferences entertaining and educational to the user.

To tackle these goals, a privacy context game will be developed. The game will attempt to improve the participant experience provided by previous studies by introducing new interactive elements and game mechanics to the process. The focus of the game interactions will be centred around scenarios, which are instances of data collection and the circumstances relating to it. A simple example of a scenario would be a CCTV camera taking video of customers entering a store. In this example, the agent is a CCTV camera, and the only description of circumstances is that it is occurring in the vicinity of a store. Scenarios are used in this project as they are an intuitive way to convey threats to privacy to a user. Since the project concerns user preference, it is especially helpful to convey these ideas in a tangible manner that is easy to imagine and place in real-life context.

This report covers the process of developing the privacy-context game chronologically - from high-level design, to implementation, to analysis and discussion of the results.

1.1 Motivation

Data collection is inevitable in the modern age. Owing to the commercialisation of the Internet and the emergence of the attention economy, user information has become a highly valued commodity, and data collection has become a very lucrative affair. As capitalism dictates, an activity will persist as long as it is profitable, and there does not seem to be an end in sight to the ever-growing value of data. Due to the inability to stop this phenomenon outright, it is crucial to pursue privacy not as data restriction, but as control of its flow. To this end, we must understand user privacy preferences more extensively. Finding out what users value most in regards of privacy, and what kinds of data collection they find acceptable, can help us focus our efforts in protecting them from unwanted surveillance, whether through legislation or technology.

However, collecting this data is a separate challenge. Pervasive advertisers have an advantage of collecting data they need by surveilling users through websites, embedded trackers in consumer tech, or even IoT devices in public spaces. Meanwhile, privacy researchers often have to resort to simple questionnaires. This project aims to improve the user experience of these questionnaires.

1.2 Objectives

The main objective of this project was to develop a game that presents players with scenarios and related questions. The game offers new ways to interact (and answer) these questions that provide more convenience to the player. Overall, the game offers a more engaging experience compared to other methods of privacy preference collection available. The implementation is evaluated by players in a post-game survey.

The second objective of the project was collecting privacy preference information from the players. This data was then analysed and presented in the Findings on Privacy Preferences chapter.

The third objective of the project is to provide benefit to the player through the game in the form of entertainment and education on privacy concepts. This was realised through the inclusion of the results page, which shows statistics of their performance in the game, as well as comparing their responses to those of other participants.

Chapter 2

Background

2.1 User Privacy Knowledge

Numerous user surveys were conducted previously regarding privacy preferences. These range from binary choice of private/public in regards of pictures [21], to more extensive surveys covering textual scenarios [25].

However, there are inconsistencies in findings of user preference, and the most important factors vary. For example Naeini *et al.* [25] concludes that the most important parameters that users consider when sharing data is “location” and “type of data”, while parameters such as retention time and who the data is shared with were the least valued parameters. Meanwhile, Mugan, Sharma, Sadeh [24], while agreeing on the type of data being an important factor, also emphasises purpose of the data collection as the major factor. This highlights the complexity of privacy preferences, and their dependence on context, as the former paper focused on IoT data collection, while the latter outlines mobile preferences only.

The increasing prevalence of IoT devices raises unique privacy concerns. Firstly, it is much harder to recognize when data is being collected by these devices. Secondly, the social norms that influence privacy decisions of users are still forming around IoT devices, and they introduce types of data collection which are socially acceptable in public into private spaces, such as audio and video recording. This problem is already being tackled from several angles. Kökciyan, Yolum [19] proposes an automated solution by using a system of interconnected IoT devices that attempts to determine the context, and in turn, whether data sharing and collection is acceptable in a given situation. Another alternative is offered by the IoT Assistant App, which aims to inform the user on the surrounding devices and their capabilities via a mobile app, rather than controlling them directly. The challenge with both of these approaches is that they require cooperation across devices by different manufacturers.

Another challenge in determining clear privacy preferences are the users themselves. There are concerns that users might have some privacy requirements they cannot rationally justify [5]. This can add an additional level of complexity, as we might want to have a distinction between rationalised preferences, and ones that could be classified as “gut feeling”. Additionally, it could be worth attempting to map these vague preferences to certain parameters (that is, finding patterns where unsubstantiated preferences occur more frequently when certain parameters are present).

However, while outlined privacy concerns have been shown not to line up with actual privacy settings in e.g. social networks, this discrepancy has been slowly improving as users get more comfortable with using the technology in question [37]. This tendency will hopefully apply to all data-gathering technology as users find them more intuitive and aware of them.

2.1.1 Privacy Context

The idea of “context” has a wide array of interpretations, as outlined by Nissenbaum [26]. Nissenbaum claims that ethical concerns relating to privacy form and evolve naturally over time, and has based the theory of contextual integrity (CI) on this. CI defines privacy not as secrecy, but as appropriation of flow of (personal) information [27].

To determine whether an instance of data collection satisfies CI, it is split into parameters and each is checked for discrepancies with the established norm.

In a sense, CI formalizes what users usually determine by intuition. A good example of this phenomenon can be seen in Naeini *et al.* [25], where the study found results consistent with established norms. More specifically, data collection in public spaces was more acceptable than at home, which is consistent with the “western tradition of public/private dichotomy” [25]. Preferences were less predictable and consistent when the context involved features for which these norms did not have time to form, e.g. IoT devices.

This project will interpret context as a collection of attributes describing a data collection scenario, which will cover industrial and, both directly and indirectly, social contexts. This decision was made due to the fact that these types of contexts involve the agent and/or user most directly. The attributes pertaining to the industry could include technologies used, parties collecting the data, data sharing and retention policies. The social contexts include time, location (and the user’s relation to it, e.g. workplace), action causing the collection of data, and the social occasion (e.g. party, day at work).

2.1.2 Context Exploration Through Scenarios

Papers analyzing users’ response to contexts do so by formulating specific data collection scenarios. A scenario in this case, is an instance of data collection with defined parameters (such as time, place, use of data, etc.). The way in which these contexts are expressed can range from a text description to image representation.

Notable papers using image datasets were Kurtan & Yolum [21], Ayci & Yolum [4] and Zerr & Siersdorfer [38]. Both [4] and [38] utilized and relied on photo tagging by services such as Clarifai and Flickr. This is an unreliable technique, as the services provide inconsistent tags for images. Namely, the tags do not guarantee to describe consistent features that could be directly compared between images such as the number of people in the photo, time of day, indoors/outdoors, etc. This is a reflection of a bigger issue with image datasets - the depicted scenarios can be incredibly broad leading to inconclusive results, or require careful curation. Despite these limitations, the authors in [38] were able to produce promising results in identifying private and public photos with image processing and learning algorithms. Ayci, Yolum [4] used natural language processing techniques to determine the privacy of an image from its tags. Only one of these papers, [38] collected privacy data from real users, and this was done in a very limited capacity. Namely, a game was used to classify images into categories of "private", "public", and "undecidable". While this served its purpose for the greater ends of using this data to test predictions against, it is trivial for this project, focusing mainly on user feedback. Due to the inconsistencies mentioned above, it is difficult to craft questions that would apply to several images at once, or allow for more specificity than simply picking whether they should be private or public.

Papers focusing more on the human factor, meanwhile, used surveys with textual scenarios [25][31], which were auto-generated. This allowed for a high number of distinct scenarios, while also controlling their attributes to be consistent and describable. Formulating scenarios in text allows us much more freedom in defining them, as opposed to images. The data collected on users can be multifaceted and extend beyond the image of their person, such as location, heart rate, age, shopping history, etc. [24] also introduces the group data is being shared to as a factor, such as family or university, as previously mentioned papers assume data is being kept by the company responsible for the device. Furthermore, both physical and abstract conditions of the data collection can be clearly listed. These scenarios could also be directly transformed to audio for accessibility purposes. The drawback of this format is that these scenarios are even further removed from a real life experience of the situation.

2.2 Challenges

Some contexts do not have definite social norms of privacy yet, due to technology or environment being too new for these norms to naturally form [25]. Selection of the participants. It is essential to have a representative group of participants for the population we are studying. Previous research had distinctly different participant groups, making their results difficult to compare. Namely, these groups were either mostly educated middle-class or volunteers found through Mechanical Turk [25], who by the fact that they work for this program, represent a particular category of participants that may be more tech-savvy than average.

Users are getting used to the fact that websites are collecting their data as they browse, and due to GDPR and Terms and Conditions during registration, services are transparent about this data collection. Furthermore, the very act of signing in or opening a page gives a clear "line" from when data gathering could start [19]. However, by the very nature of IoT devices and their integration into any device, it is much harder to be aware of your data being collected in these scenarios. There are attempts to tackle this, such as IoT Assistant App [8], but, of course, solutions like this are not widespread.

2.3 Gamification

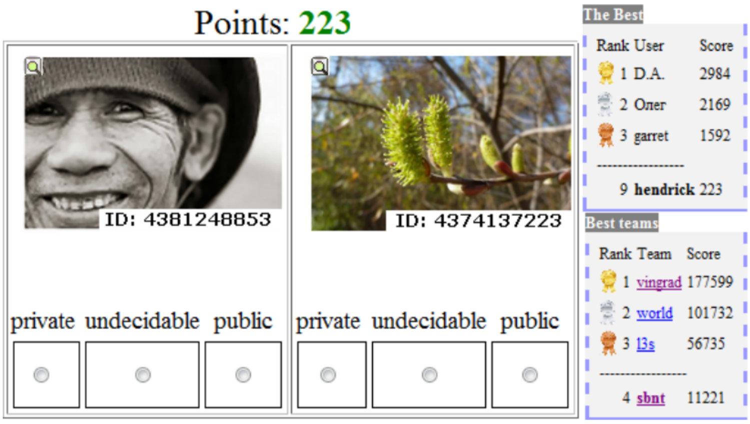
The format of a game has been chosen for this project to both improve engagement with data collection that is usually carried out in a form of a survey, and to hopefully tackle biases present in other approaches (e.g. suggestive/leading questions, assuming knowledge). This approach can also help educate users about privacy, as gamification has been shown to have positive effects on learning [33].

It is also worth exploring the benefits of a game over standard method of gathering opinion data - surveys. Using the extended interactivity allowed by this medium, we are able to tackle apparent issues in surveys, for example:

- Dullness - regular surveys are boring to go through for many participants, which causes problems with finding participants, and even affects the reliability of data, as participants are often tempted to rush through the answers or answer them dispassionately to just be done with the survey [14].
- Engagement - a study by Harms *et al.* [14] found that gamification can potentially lead to a significant increase in engagement with surveys, as participants find them both more enjoyable and spend more time with them compared to a standard survey, without significantly affecting the answers given. As an example of importance of this factor, we can take a previously discussed paper [25], which took completion time into account and found some responses that had to be disregarded outright due to the speed at which they were completed.

A game approach does not necessarily imply a direct improvement, and there are trade-offs. While there was an all-round improvement in participants' enjoyment of the process, and increase in time spent with the survey, there was a decrease in completion rate, as people likely abandoned the gamified survey after finding it too unfamiliar. In other words, it caused the survey to become more polarizing, where those taking it had more positive outlook on it, but a decreased number of participants completed it. While the study found no significant changes in the answers across the gamified and standard survey, there is still a possibility to introduce new biases with this approach [14].

While there are examples of games and gamified interactive systems tackling various educational subjects such as maths, languages [22], and even more informatics-adjacent topics such as security [2] and programming [36]. However, there do not seem to be good, fleshed-out examples of privacy games. The most direct example, used by Zerr, Siersdorfer, Hare [38] only scratches the surface of interactivity with a bare minimum in terms of interface and a singular act of classifying images as public or private (See Figure 2.1). One unique feature this game had was a built-in incentive to play in the form of a high score and a leaderboard. Several of the TULIPS [1] projects present more promising examples with an immersive presentation and more ways to interact with the system. Kuzmiak [6] and Asher [2] closely simulate the real-world interface of firewall management, while Sehl [34] abstracts the concepts with a user-friendly drag-and-drop gameplay. All three projects use visualizations both to make projects pleasant to look at, and to clarify the concepts being simulated. However, these projects lean closer to the topic of security rather than privacy, and are more in-



Points: 223

ID: 4381248853

private undecidable public

○ ○ ○

ID: 4374137223

private undecidable public

○ ○ ○

The Best

Rank	User	Score
1	D.A.	2984
2	Oner	2169
3	garret	1592
9	**hendrick**	223

Best teams

Rank	Team	Score
1	vingrad	177599
2	world	101732
3	l3s	56735
4	**sbnt**	11221

Figure 2.1: User interface of the privacy game by Zerr, Siersdorfer, Hare [38]

terested in exploring static rules rather than socially-defined contexts we are interested in.

Chapter 3

Game Design

This chapter covers the work done and decisions made that informed the final implementation. Firstly, game features are covered, which are intended to tackle the main problem of the project - streamlining and gamifying the process of collecting participant privacy preferences. Afterwards follows the section covering concepts of the graphical user interface that incorporates the outlined features. With features and their presentation decided, we arrive at the data model to structure the data required (or implied) by the design. Finally, the ethics section discusses the ethical concerns of the game, and the structure of the user study.

3.1 Game Features

The features for the game start from the baseline of a standard questionnaire, meaning multi-choice and write-in questions. The goal for the feature list was to improve on this formula by expanding the existing ways of expression in a questionnaire, as well as providing new ones. The secondary goal is to streamline the experience to make it more satisfying to the user. This design led to the following features:

Progress Indicator - showing the user an estimate of the amount of activities, and their progress along them. This adds to the feeling of transparency, as well as providing the users a sense of progress as they go through the game.

Question Display - to maintain focus on the scenario at hand, only one question is provided at a time, with scenario text always in view. This streamlines the process of answering questions, ensuring the user does not have to scroll up and down the page to remember the question or scenario.

Text Selection - a previous study [25] that used textual scenarios did not allow the users to interact with the scenario more directly - they could only express their opinions through predefined answers. A text selection feature is introduced to allow for more interactivity in the response, while still bounding the user within the scenario. This is utilised by asking users questions that can be answered with a snippet of text that they can simply highlight within the scenario. This eliminates possibilities for typos or rephrasing for easier automated processing of the results that may arise by asking users for input in writing.

Pre-Filled Write-In Questions - to supplement multiple choice questions, write-in questions are also intended to be used. These once again can be specialised for this particular survey by pre-filling text boxes with the scenario or part of the scenario and asking users for changes or improvements to the text based on a certain criteria. This has two-fold benefits: firstly, it once again limits the variation in this custom answer, introducing potential for partly automating processing of results, and secondly, it reduces the work required by the user/makes the writing task less daunting.

Help Button - a help button is used to provide elaboration on the intention of a given question. When unconventional features are introduced, this can also explain how to use them. The text provided in the help option could potentially slow down a user that understands the activities initially, which is why it is relegated to a button.

Results Screen - conventional games are known to have a “win state”, or at least a score. In our case, this is implemented by ”rewarding” the user with a results screen at the end of the game. The ideal scenario for this screen is providing valuable information, but at the very least this screen should contextualise and summarise the actions the user took during the game and potentially present them in relation to results of other participants.

3.2 IUIPC Scale

Studies of privacy preference require formalising the often nebulous nature of privacy preferences. For this purpose various scales have been developed that map privacy concerns of individuals to a numerical scale, which can be calculated and compared quantitatively. There are several of these scales, such as GIPC (one-dimensional global information privacy concern scale), CFIP (concern for information privacy), and IUIPC (internet users information privacy concerns).

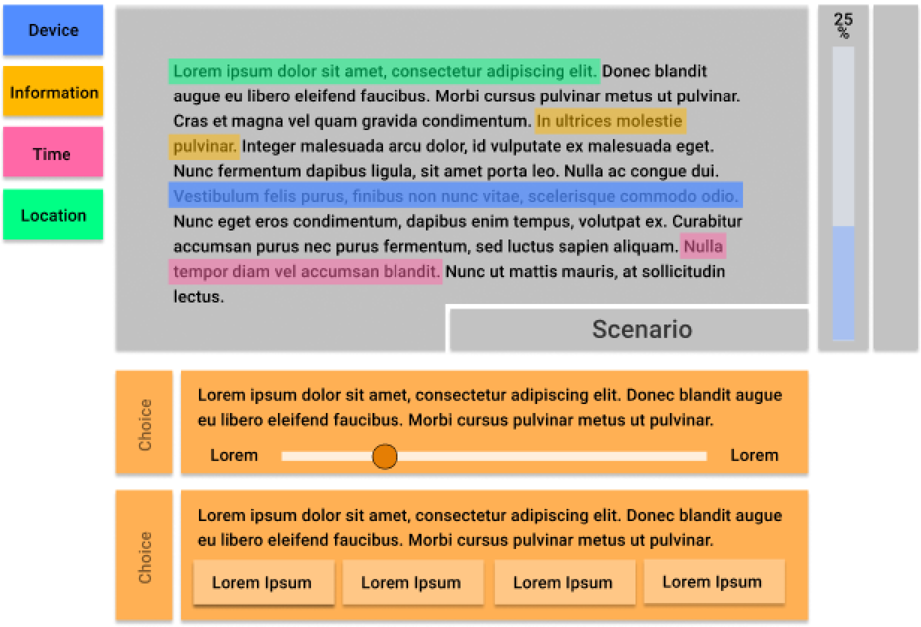
For this project, IUIPC was chosen as a happy medium between the three options. GIPC, as the name implies, is one-dimensional and does not provide information beyond the scale of “privacy concern” in the abstract. CFIP addresses this by providing 4 dimensions and 15 items. The amount of data needed to gain meaningful results from this scale might require numerous participants or larger number of questions per participant, leading to having to ask them for a bigger time commitment. IUIPC, while having a more simplistic scale, still provides dimensionality.

The IUIPC scale can be seen as “the degree to which an Internet user is concerned about online marketers’ collection of personal information, the user’s control over the collected information, and the user’s awareness of how the collected information is used.” [23]. and as such, helps us in understanding the intricacies of user privacy preferences.

The game is intended to contain a collection of questions which would determine the IUIPC score of a player. This score could then be shared with the player and explained at the end of the playthrough.

3.3 GUI

The particular application of the questionnaire style ties each question to a shared text - the scenario, which should be emphasised through the design. Due to this, from the very first concepts, the scenario was placed at the center of the screen, visible at all times.



Lorem ipsum dolor sit amet, consectetur adipiscing elit. Donec blandit augue eu libero eleifend faucibus. Morbi cursus pulvinar metus ut pulvinar. Cras et magna vel quam gravida condimentum. In ultrices molestie pulvinar. Integer malesuada arcu dolor, id vulputate ex malesuada eget. Nunc fermentum dapibus ligula, sit amet porta leo. Nulla ac congue dui. Vestibulum felis purus, finibus non nunc vitae, scelerisque commodo odio. Nunc eget eros condimentum, dapibus enim tempus, volutpat ex. Curabitur accumsan purus nec purus fermentum, sed luctus sapien aliquam. Nulla tempor diam vel accumsan blandit. Nunc ut mattis mauris, at sollicitudin lectus.

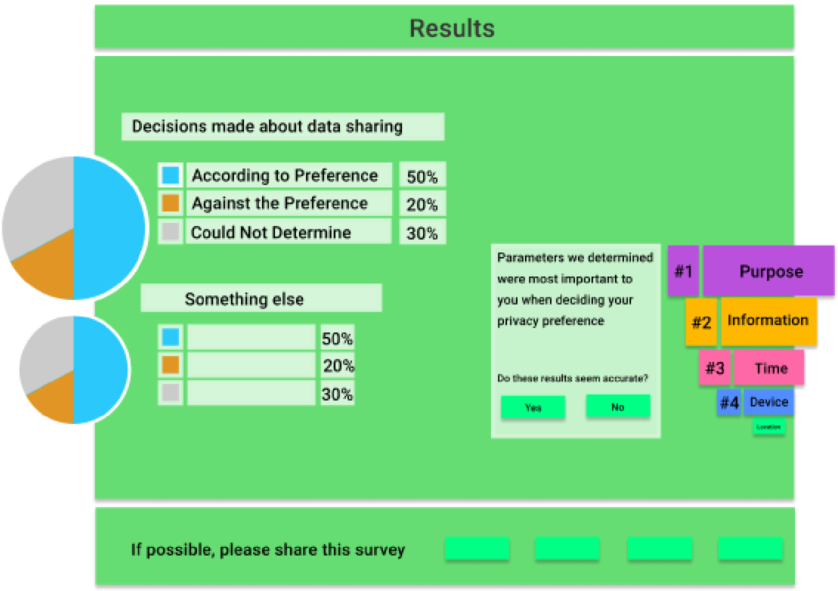
Lorem ipsum dolor sit amet, consectetur adipiscing elit. Donec blandit augue eu libero eleifend faucibus. Morbi cursus pulvinar metus ut pulvinar.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Donec blandit augue eu libero eleifend faucibus. Morbi cursus pulvinar metus ut pulvinar.

Figure 3.1: Concept of the scenario screen with multiple questions.

The first concept (Figure 3.1) attempted to exploit the modular nature of automatically generated scenarios. Easily detectable properties were highlighted and pointed out to the user. However, this was antithetical to the idea of increasing player agency and decreasing external influence. Later iterations completely removed preemptive indicators from scenario text, and highlighting was made into an active mechanic, rather than a passive feature. The need to indicate progress to the user was also addressed by the first design, this has stayed in the implementation and was fleshed out to include several layers of progress (within the scenario and overall). Finally, one of the intentions of the design was to decrease influence on questions caused by previous or future questions. To this end, the concept features a “queue” of questions, where each answered question is removed and replaced by the next. This idea was taken further in the final design and only one question was displayed at a time.

Another feature that was decided early on was the results screen. The intention was to provide the user with interesting and, if possible, educational information about their performance in the game. The mock-up (Figure 3.2) contains data presented in a colorful, visual way. There is an emphasis on comparing player chosen privacy preferences versus inferred privacy preferences. This would be done through two sets of questions - ones related to general preferences, and ones that ask about data sharing in a specific scenario. Lastly, the bottom row indicates an idea to incorporate social media sharing to the game, attempting to replicate the self-advertising nature of online



Results

Decisions made about data sharing

🟦	According to Preference	50%
🟧	Against the Preference	20%
⬜	Could Not Determine	30%

Something else

🟦		50%
🟧		20%
⬜		30%

Parameters we determined were most important to you when deciding your privacy preference

Do these results seem accurate?

Yes No

#1 Purpose

#2 Information

#3 Time

#4 Device

Location

If possible, please share this survey

Figure 3.2: Early concept of the results screen.

quizzes. The sole intention of this feature is to increase the circulation of the survey and in turn, the number of participants. It is worth noting that this design decision was made before ethics considerations and was removed before implementation.

After the initial mock-up focusing on functionality, further concepts focused on exploring possibilities of the visual presentation of the game. The most notable example was the “Papers, Please” concept (Figure 3.3), named after the game it was inspired by [29]. The intention of this aesthetic is to place the player into a mind state of “play”. The medium of games is capable of presenting even mundane tasks in an entertaining way. Koster [20, p. 40] states - “Fun from games arises out of mastery. It arises out of comprehension” and to this end an unconventional design of the web application may add to this sense of exploration and provide a more varied learning experience (both of privacy concepts and the game mechanics).

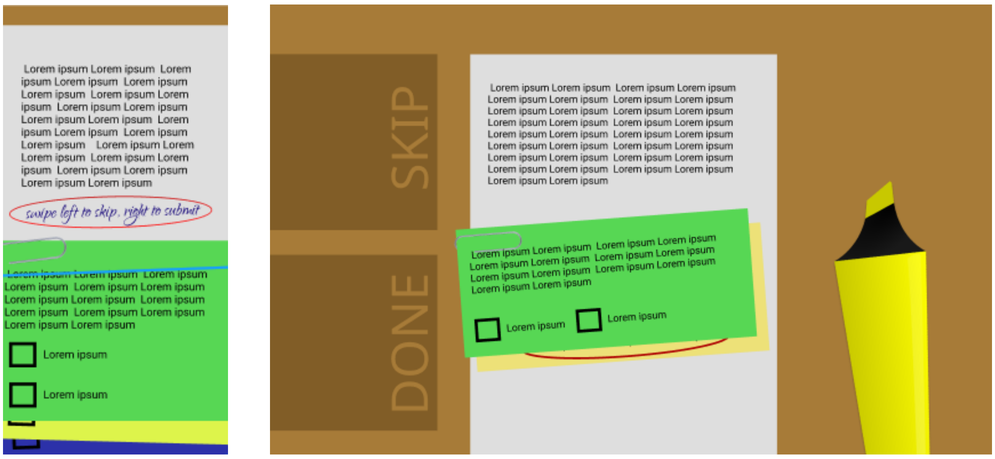

swipe left to skip, right to submit
□ Lorem ipsum
□ Lorem ipsum
SKIP
DONE
□ Lorem ipsum □ Lorem ipsum

Figure 3.3: “Papers, Please” concept. Left - mobile version. Right - desktop version.

3.4 Data Model

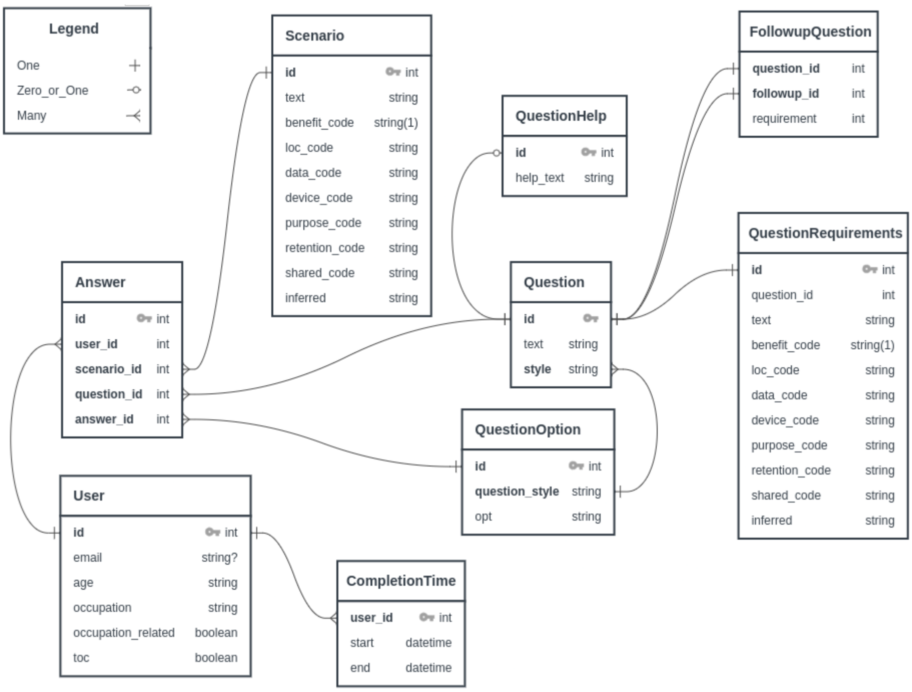


Figure 3.4: Database relationship graph

The data for the project was structured in a relational model. This design decision accommodates the initial scenario dataset, as well as other elements with consistent properties, such as questions.

Questions have significant amount of repetition in the form of multiple choice answer options, especially the Yes/No questions. In order to account for this, questions are given a *style* property which defines the options for answering them. Question options are stored in a separate table. Each option is a separate entry for fluid adjustments, and to account for questions with custom input, such as highlighting or writing. Entries to those types of questions are added as new question options. Each question may have a help prompt, which is linked to the question by ID. The help text is kept in a separate table so that it can be converted to a one-to-many relationship with questions when prompts are reused. Followup relations between questions are kept in a lookup table with IDs of 2 questions (starting question and its followup) and the requirement to be met in order for the followup to appear. Finally, a question requirements table is included to account for questions requiring a specific property from a scenario (e.g. type of device). This data is included as a separate table (related by ID) due to most questions not having special requirements, thus saving computation time and easing the parsing for majority of them.

The most important construct is the answer table, as it connects users, questions, question options, and scenarios together. The scenario table contains the scenario text itself, as well as values describing its properties in two-three letter codes. All of these entities being referenced by ID minimises repetition in the database. Furthermore, it is expected to assist in grouping answers by one of these properties.

3.5 Ethics

The project involved a user study aspect, and thus required ethics considerations and approval. The study was certified according to the Informatics Research Ethics Process under the RT number 5557 on February 26th, 2021. This section describes decisions made in regards of user data and activities involved.

3.5.1 User Data

All decisions regarding user data had to be justified from two aspects - whether it will be useful to the study, and whether it would allow identification of the participant. This eliminated most personal information, such as name, address (both being irrelevant), or sex (no indication whether it would affect the study results, as well as it being a protected characteristic). Information that did satisfy these requirements was the following: age and occupation (with a qualifier). Age was chosen due to potential differences in digital privacy perception [18] between age groups. Age input rather than age range selection was chosen to allow for creation of ranges at a later point, to point out meaningful distinctions if they emerge. Lastly, age could be used to identify limitations of the survey results if the respondents are of very similar age. Occupation was chosen due to its high potential impact on privacy knowledge. Students may have been taught privacy concepts, while professionals may be more wary of data collection by their workplace. There was also a binary qualifier added to this field, asking participants whether their occupation is related to “Computer Science, Privacy, or Cybersecurity” as this detail may influence privacy attitudes the most. Lastly, for transparency, users may optionally provide their email address to be informed about the completion of the study and its results.

3.5.2 Survey Design

To increase the chances of people agreeing to participate in the study, it was decided to streamline the survey process as much as possible. This was done by having most parts of the survey integrated within the game application itself, allowing the survey to be carried out by accessing the link with to the game with no external information required. This meant that the first page had to contain the Participant Information Sheet and require consent through the Consent Form.

The final part of the survey - post-game questionnaire - takes place on the Microsoft Form website. This is tied with the game by providing a hyperlink to the form at the end of it. The game generates an ID for the participants that they are prompted to enter in the survey. This is done to make sure the participants played through the game before accessing the questionnaire.

Finally, the time estimate provided to Ethics Process was “around 30 minutes” and this was taken into account when choosing the content to present to participants. First, all the questions were chosen, then one scenario was play tested and time tracked. Based on that estimate, scenarios were added to fit within the time limit and the whole game played through again. In the end, the developer playthrough took 15 minutes (this includes scenario and privacy score questions) with 5 scenarios. This was deemed the optimal duration, taking into account the variation in time spent on each question by participants that are not familiar with them, as well as the additional post-game questionnaire. To get as much data as possible from a limited participant pool, an option to go through additional scenarios is added. This is mentioned in the Participant Information Sheet and participants were informed the additional scenarios are not included in the 30 minute time estimate.

Chapter 4

Implementation

4.1 Development Environment

4.1.1 Front-End

The project uses vanilla (here meaning no frameworks) JavaScript [16]. Required HTML objects were accessed through the *getElementById* or *getElementsByClass* methods, which remain manageable due to small pages. Most JavaScript code gets invoked through event listeners. The CSS framework *Bulma* [7] was used for its cohesive minimalist style and focus on responsiveness. Using *Bulma* allows the game to be more easily adapted for different screens (e.g. phones, tablets) in the future.

4.1.2 Back-End

Python 3 [32] was chosen for the server implementation as it is one of the fastest languages to develop an application in. Performance was not a concern due to the nature of the application. The only processing-heavy activity that could occur for the game is calculation of numerous statistics at the end. If need be, the highly performant *NumPy* [28] library could be used to optimise these calculations. *NumPy* is known to perform better than standard Python library functions in mathematical tasks.

The game uses the *Flask* [10] web framework for implementing the endpoints. This is one of the most lightweight server frameworks available for Python, but is highly extensible with additional libraries. This framework was chosen due to a relatively low complexity of the game compared to other modern web projects. For the current usage, features such as login capabilities (same account in several sessions) or instance scaling are not required, making more extensive frameworks like *Django* [9] unnecessary. If any advanced features are required, they can be easily added through the numerous extensions. One of them that is being used for this project is *Flask_SQLAlchemy* [35], which is described more in-depth in the next section. It is also worth noting that Flask comes with *Jinja2* [17] support, which is a web templating engine allowing compartmentalisation of HTML code and its reuse in multiple pages, as well as use of variables and conditional statements for page generation.

The service chosen to host the game is *Heroku* [15]. The full functionality of the application does not require exceeding the free tier of this service. *Heroku* handles certificates for an HTTPS connection, as well as provides a human-readable URL to access the game. The deployment process is automatic and allows updates on push to a specified git branch, which allowed for workflow configuration through GitHub [13], which was already being used for version control of the project. The one drawback of this service is that the virtual machine running the application is only kept in an active state for 30 minutes from its last request. Any requests outside of this window require the machine to start again, leading to slightly longer initial load time.

The deployed game was made accessible through URL `https://privacy-context-game.herokuapp.com/` for the purposes of the survey.

4.1.3 Database

*PostgreSQ*L [30] was picked as the database system. The data needed for the project is easily represented in a relational data model, and PostgreSQL offers more advanced database features than MySQL if a need for them arises. Additionally, there are various options for *PostgreSQL* clients on Python, including *asyncpg* [3], which offers a very efficient asynchronous implementation. For this project, *SQLAlchemy* [35] was used, which is an extensive SQL toolkit that includes an ORM (Object-relational mapping) and a client. The ORM was an incredibly useful tool for development, as it allowed modelling of the database from within the same environment as the rest of the server-side features. The object abstraction also allows to save time when writing queries, as all most common SQL commands have respective functions. Updating rows is as convenient as changing the values of the object, rather than writing a full SQL query. Finally, the ORM model simplifies the unwieldy process of requesting the auto-incremented ID from a newly inserted entry.

4.2 Game Flow

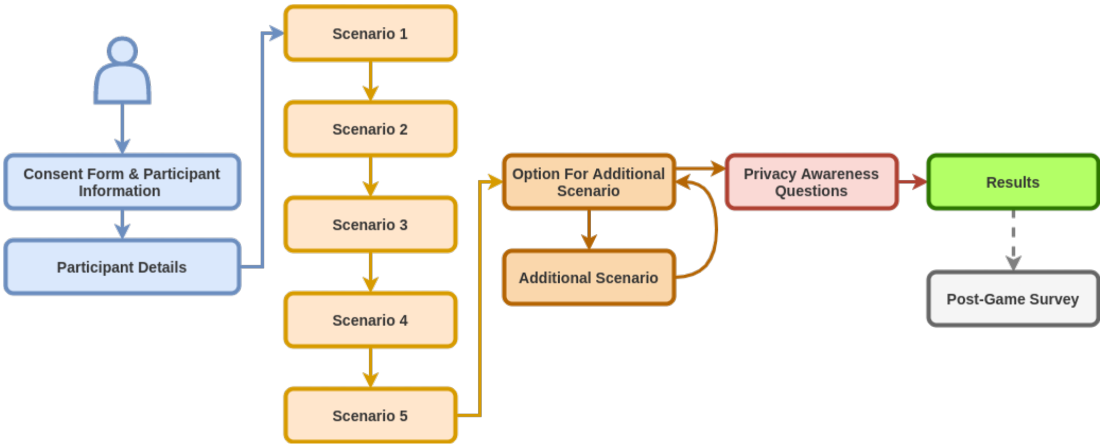


Figure 4.1: Game Flow Diagram

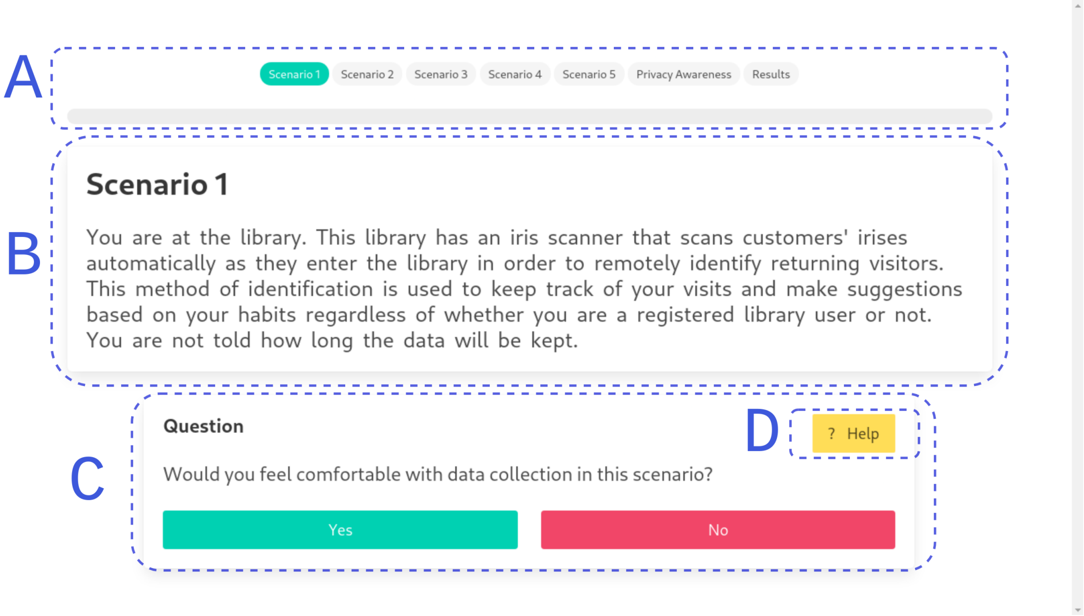
This section describes the user-facing implementation of the game in the order it would be played through.

The game flow in the abstract is succinctly illustrated in Figure 4.1 - the players are first presented with study-related documents, then fill out their details, before starting the game proper (illustrated by blue nodes). The game consists of 5 mandatory scenarios (light orange nodes), and then moves into a loop that allows to complete additional scenarios up to a certain limit (dark orange node). The next section contains IUIPC privacy awareness questions (red node), after which results are displayed. The results lead to an external page which contains the post-game survey.

The game is designed to act as a self-contained study, thus it begins with a page displaying University study Participant Information Sheet and Participant Consent Form, and includes buttons to indicate consent. The game cannot be started until consent is given. The next page asks for some basic information about the participant, namely their age, category of occupation, whether it is related to any fields that may inform them about privacy or technology, and an optional field for their email in case they wish to be informed about the findings of the study and the finished report. The age field includes input validation that checks that the age entered is realistic and that the participant is over 18 (the acceptable value range is 18-110).

After providing the information, the user starts the game. All pages beyond this point (except for results) contain a progress indicator at the top, displaying progress within the current page, and the remaining sections. The user starts by entering a scenario page.

4.2.1 Scenario Pages



Scenario 1

You are at the library. This library has an iris scanner that scans customers' irises automatically as they enter the library in order to remotely identify returning visitors. This method of identification is used to keep track of your visits and make suggestions based on your habits regardless of whether you are a registered library user or not. You are not told how long the data will be kept.

Question

Would you feel comfortable with data collection in this scenario?

Figure 4.2: Scenario Page with marked sections. A - progress indicators. B - Scenario box. C - Question box. D - Help button.

The scenario page is the core component of the game. It features a scenario at the center of the screen (Area B in Figure 4.4). The study used a set of 14 scenarios, 5 of which were presented to all users, while the remaining 9 were optional (optional scenarios can be found in Appendix A). The mandatory scenarios are as follows:

Scenario ID: 177
You are at the library. This library has an iris scanner that scans customers' irises automatically as they enter the library in order to remotely identify returning visitors. This method of identification is used to keep track of your visits and make suggestions based on your habits regardless of whether you are a registered library user or not. You are not told how long the data will be kept.

Scenario ID: 354
You are in a public restroom. This restroom has cameras that are recording video of the entire room. The video is shared with law enforcement and they will keep it for one year. You are not told what the data is used for.

Scenario ID: 338
You are in a public restroom and your smartwatch is keeping track of your specific location. Your location is shared with the device manufacturer. You are not told what the data is used for or how long it will be kept.

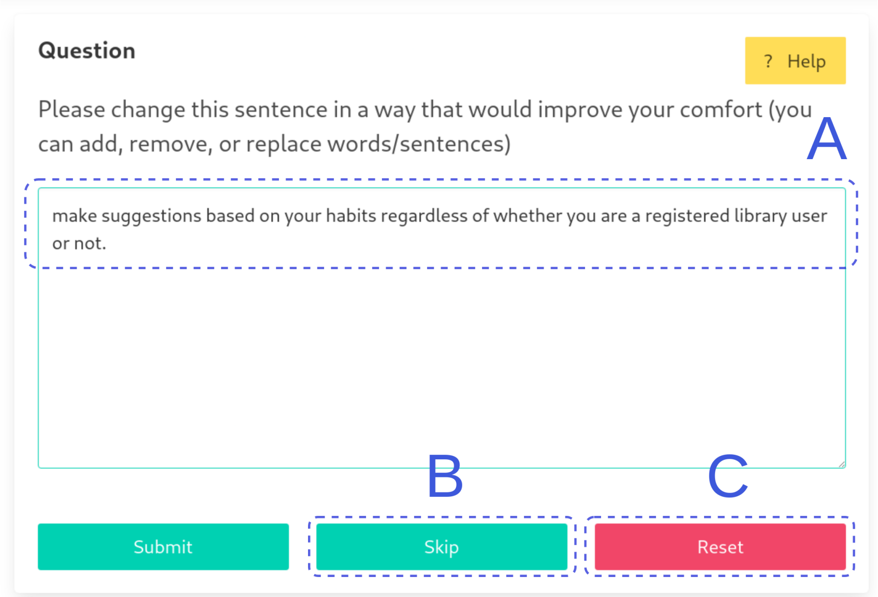
Scenario ID: 234
You are at a friend's house. All rooms have temperature sensors that check for abnormal temperatures, which indicate potential hazards, e.g., fire. This data is managed by their security company. You are not told how long the data will be kept.

Scenario ID: 260
You are at work and your smartphone is keeping track of your specific position in the building. Your position is used by the device to determine possible escape routes in the case of an emergency or a hazard. This data will be kept on your phone until you leave for the day.

Progress indicators can be found above the scenario, displaying the total game progress with a list of items, and current scenario questions progress with a bar (area A in 4.4). One question is displayed under the scenario at a time (area C in 4.4. The questions vary both in their text and the type of interaction. The types of interaction can be split into these categories:

Multiple Choice - most generic questions, presenting list of predefined options in the form of buttons. In some cases (e.g. Yes/No questions), these buttons are color-coded for easier navigation.

Write-In - most questions that use user text input have been streamlined to provide user with some initial text that they only have to modify. These questions act as followups and depending on the type of question that came earlier - If the previous question asked to highlight an area, the write-in followup may ask to improve that particular selection, while questions that follow up on multi-choice questions ask to modify the entire scenario. Players are provided a "Skip" button (B in 4.3) to indicate they are free to not make changes in cases where they cannot think of ways to improve a given text, or do not believe there are ways to improve it. The "Reset" button can be used to



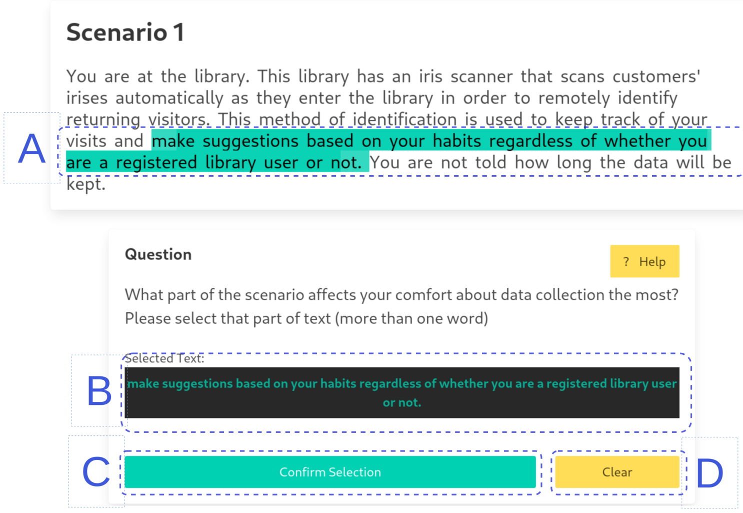
Question

Please change this sentence in a way that would improve your comfort (you can add, remove, or replace words/sentences)

make suggestions based on your habits regardless of whether you are a registered library user or not.

Figure 4.3: Write-In question with initial text. A - text received from a previous highlight question. B - “Skip” button. C - “Reset” button

set the text within the text box to its initial content (C in 4.3).



Scenario 1

You are at the library. This library has an iris scanner that scans customers' irises automatically as they enter the library in order to remotely identify returning visitors. This method of identification is used to keep track of your visits and make suggestions based on your habits regardless of whether you are a registered library user or not. You are not told how long the data will be kept.

Question

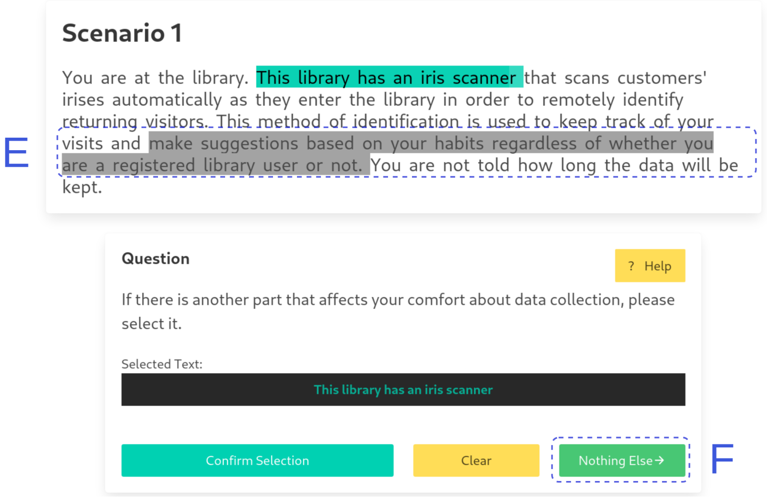
What part of the scenario affects your comfort about data collection the most? Please select that part of text (more than one word)

Selected Text:

make suggestions based on your habits regardless of whether you are a registered library user or not.

Figure 4.4: Highlight Question with marked sections. A - text highlighted by user. B - highlighted text indicator. C - Submit highlighted text. D - Clear current selection.

Highlight - these questions offer most direct interaction with the scenario, as the users can choose part of the text within the scenario display box. There is a box that displays selected text at a given time to confirm to the user selection has worked. The selected part of text is also highlighted even when the user clicks away, and is cleared only when a new selection is started or the “Clear” button is pressed. When followup questions of the same type are asked, previously submitted parts of text remain highlighted in grey. Due to this method of interaction not being used frequently in web applications, precautions were taken to accept varied user input. The text segment can be selected from start to end, or vice versa, and selection is completed if the user leaves the scenario window (accounting for the common mouse movement used to mark everything until the end of text).



Scenario 1

You are at the library. This library has an iris scanner that scans customers' irises automatically as they enter the library in order to remotely identify returning visitors. This method of identification is used to keep track of your visits and make suggestions based on your habits regardless of whether you are a registered library user or not. You are not told how long the data will be kept.

Question

If there is another part that affects your comfort about data collection, please select it.

Selected Text:

This library has an iris scanner

Figure 4.5: Secondary Priority Highlight Question. Players can submit an arbitrary number of text snippets. E - Previously submitted text. F - "Nothing Else" button, progressing to next question.

There is a subcategory of highlight question, which can be answered indefinitely (until the player chooses to stop). These questions are designed to follow up on the initial highlight question, which asks to select part of the text that has the most significant effect on their evaluation of the scenario (e.g. their comfort). The additional highlight question asks to select parts that had *any* influence in that particular aspect. The "Confirm Selection", instead of switching to the next question, clears the highlight indicator area and replaces the text there with "Submitted!". "Nothing Else" (F in 4.5) button is added to progress from this question. It has a unique color to draw attention, as well as an arrow indicating movement to make it clearer to the player it can be used to move on. The first and all subsequent highlight submissions are marked in grey within the scenario text to help players keep track of their choices (E in 4.5). This colouring is cleared once a different type of question is entered.

Some of the questions are conditional followups - they only appear if the user answered the previous question in a particular way. The followup questions may also have followups of their own, leading to chains of additional questions. This allows to ask more specifics from users on particular choices, while not burdening the rest of the participants with irrelevant questions.

The following questions were used for this study:

ID:0
Question: Would you feel comfortable with data collection in this scenario?
Style (Options): Binary (Yes, No)
Help Text: The question asks about discomfort. You may allow your data collection, but if you feel any reservations about it, you should still choose 'No'.

ID:1
Question: Do you believe this data collection is beneficial to you?
Style (Options): Binary (Yes, No)
Help Text: Do you feel that your data will be used to have a positive impact on your life? For example, do you think this data can be used to improve your safety or improve convenience?

ID:2
Question: Would you expect to be informed of data collection in this scenario?
Style (Options): Binary (Yes, No)
Help Text: If you were experiencing this scenario, rather than reading about it, do you believe you would be aware of the data collection through a pop-up or staff member telling you?

ID:3
Question: Do you believe this data collection is essential to the service being provided?
Style (Options): Binary (Yes, No)
Help Text: Do you think the service could not function without collecting this data? E.g. a delivery service would need to know your address to carry out their work, but not your full schedule (even though that could be used to decide delivery time).

ID:4
Question: Do you believe this data collection improves safety?
Style (Options): Binary (Yes, No)
Help Text: Would collection of this data directly or indirectly improve your, or overall safety?

ID:5
Question: Do you believe this scenario is occurring today? (Does this scenario seem realistic?)
Style (Options): Binary (Yes, No)
Help Text: Can you see yourself or anyone experiencing this scenario in this current day and age? **Followup: If Answer "No":** Question 7

ID:6
Question: Scenarios like this will occur in X years.
Style (Options): Multiple Choice - Years (2, 5, 10, 15+, Will Never Occur)
Help Text: If this scenario seems futuristic, how soon do you believe we could reach the point where it could be happening? Pick 'Will Never Occur' if you believe the scenario is completely unrealistic.

ID:7
Question: Would you like to be notified of data collection in this scenario?
Style (Options): Binary (Yes, No)
Help Text: Would you like to be explicitly informed when this data collection is occurring? In other words, would you be fine with this data being collected without your knowledge?

ID:8

Question: How often would you like to be notified of this data collection by your mobile phone?
Style (Options): Multiple Choice - Years (2, 5, 10, 15+, Will Never Occur)
Help Text: This presupposes that there was a method to detect when your data is being collected in this scenario, and inform you on your phone.

ID:9
Question: If you had the choice, would you allow or deny this data collection?
Style (Options): Binary (Allow, Deny)
Help Text: Assuming you could still access the service/location without agreeing for your data to be collected.

ID:10
Question: What part of the scenario affects your comfort about data collection the most? Please select that part of text (more than one word)
Style (Options): Highlight (Clear, Confirm Selection)
Additional Features: Allows selecting part of the scenario text that is submitted with ”Confirm Selection” option.
Help Text: You are now able to select a part of text by dragging and releasing your pointer. If you click away, the highlighted area should remain, and the selected text is indicated in the area below the question text. Please make sure the correct text is being displayed in that box before submitting. You can press ’Clear’ to remove any current selection.

ID:11
Question: If there is another part that affects your comfort about data collection, please select it.)
Style (Options): Highlight Loop (Clear, Confirm Selection, Nothing Else)
Additional Features: Allows selecting part of the scenario text that is submitted with ”Confirm Selection” option. Submission can be repeated indefinitely until ”Nothing Else” is pressed.
Help Text: You can repeatedly pick any part of the scenario and press ’Confirm Selection’ to select more aspects that influence your comfort. Please select everything that influences your decision to any extent. You can press ’Clear’ to remove your current selection. ’Nothing Else’ will move on to the next question.

ID:12
Question: What benefit do you see from this data collection?
Style (Options): Write In (Submit, Skip)
Additional Features: ”Submit” submits any text present in the text box as the answer.
Help Text: Write in any reasons why this data collection may benefit you directly or indirectly.

ID:14
Question: Move to next scenario
Style (Options): Continue (Continue)
Additional Features: Used to transition to next scenario
Help Text: Move to next scenario

ID:15
Question: How would you change the scenario to be beneficial to you?(you can add, remove, or replace words/sentences)
Style (Options): Change Scenario (Submit, Skip, Reset)
Additional Features: ”Submit” submits any text present in the text box as the answer. Text box is pre-filled with scenario text. ”Reset” sets contents of the text box to the scenario text again.
Help Text: You can change the scenario as much as you want to add a positive outcome you can come up with for it.

ID:16
Question: Please change this sentence in a way that would improve your comfort (you can add, remove, or replace words/sentences)
Style (Options): Change Sentence (Submit, Skip, Reset)
Additional Features: ”Submit” submits any text present in the text box as the answer. Text box is pre-filled with previously highlighted text. ”Reset” sets contents of the text box to the highlighted text again.
Help Text: This question contains the text snippet you selected as most affecting your comfort level. You can change it up in a way that would improve your comfort with data collection in the scenario above. For example, if it was most worrying that the scenario was occurring at ’work’, this could be changed to ’library’. You can select ’None’ if you cannot think of an improvement that would make you more comfortable with this data collection.

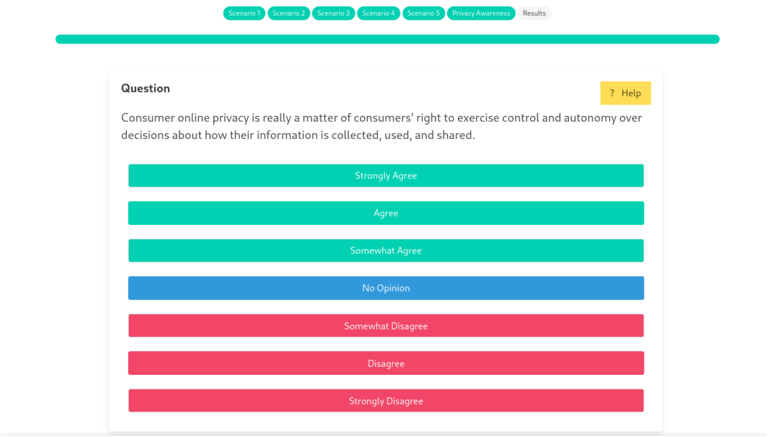
All the questions are pre-loaded and hidden/displayed in a sequence. Each question completion immediately sends out a request to the server and saves the answer in order to preserve results in case of an early exit by the participant.

Players go through 5 scenarios with the same questions. This ideally gets them comfortable with the format and allows for faster completion of subsequent scenarios. After the 5 scenarios, the users are prompted with whether they would have time to complete an additional scenario. If they choose “Yes”, an additional scenario is provided from a pool of 14 scenarios used in one of the surveys from Naeini *et al.* [25]. Once these are exhausted, the user is directed to the results screen.

4.2.2 Privacy Score Questions

Privacy Score questions are presented after all the scenario questions are done (including the optional). They are not tied to a specific scenario but rather ask users to identify their general feelings towards privacy. The questions are identical to the ones used in Naeini *et al.* [25] in order to evaluate whether the experience with the previous part of the game influences responses to them. The options are in a scale and range from “Strongly Agree” to “Strongly Disagree”.

There are a total of 10 of these questions, and they test the users in three aspects described more in depth in the previous section 3.2 - control (Question IDs 17-19), awareness (Question IDs 20-22), and collection (Question IDs 23-26). The questions are taken from Naeini *et al.* [25] All of these questions have the following answer op-



Consumer online privacy is really a matter of consumers' right to exercise control and autonomy over decisions about how their information is collected, used, and shared.

Figure 4.6: Privacy Score Question example

tions: “Strongly Agree”, “Agree”, “Somewhat Agree”, “No Opinion”, “Somewhat Disagree”, “Disagree”, “Strongly Disagree”. They also share the same Help text: “Please select how much you agree with this statement. If you are not sure or do not have strong feelings one way or another, please pick the middle option”

ID:17
Question: Consumer online privacy is really a matter of consumers’ right to exercise control and autonomy over decisions about how their information is collected, used, and shared.

ID:18
Question: Consumer control of personal information lies at the heart of consumer privacy.

ID:19
Question:I believe that online privacy is invaded when control is lost or unwillingly reduced as a result of a marketing transaction.

ID:20
Question: Companies seeking information online should disclose the way the data are collected, processed and used.

ID:21
Question:A good consumer online privacy policy should have a clear and conspicuous disclosure.

ID:22
Question: It is very important to me that I am aware and knowledgeable about how my personal information will be used.

ID:23
Question: It usually bothers me when online companies ask me for personal information.

ID:24

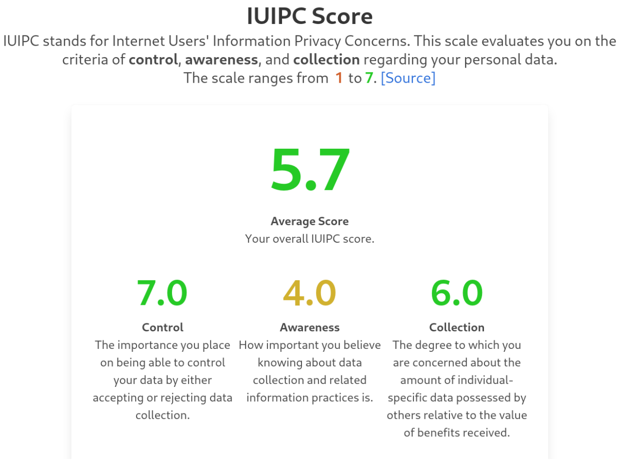
Question: When online companies ask me for personal information, I sometimes think twice before providing it.

ID:25
Question: It bothers me to give personal information to so many online companies.

ID:26
Question:I'm concerned that online companies are collecting too much personal information about me.

4.2.3 Results Page

The results page is the last screen of the game. This screen is intended to present the user analysis of their performance over the game in an engaging way. It also contains a notification at the top of the page that leads to the post-game questionnaire and provides a randomly generated player ID to use there. Information presented in the results page is split into three distinct categories: IUIPC Score, Badges, and Statistics.



IUIPC Score

IUIPC stands for Internet Users' Information Privacy Concerns. This scale evaluates you on the criteria of **control**, **awareness**, and **collection** regarding your personal data.
The scale ranges from 1 to 7. [Source]

5.7
Average Score
Your overall IUIPC score.

7.0
Control
The importance you place on being able to control your data by either accepting or rejecting data collection.

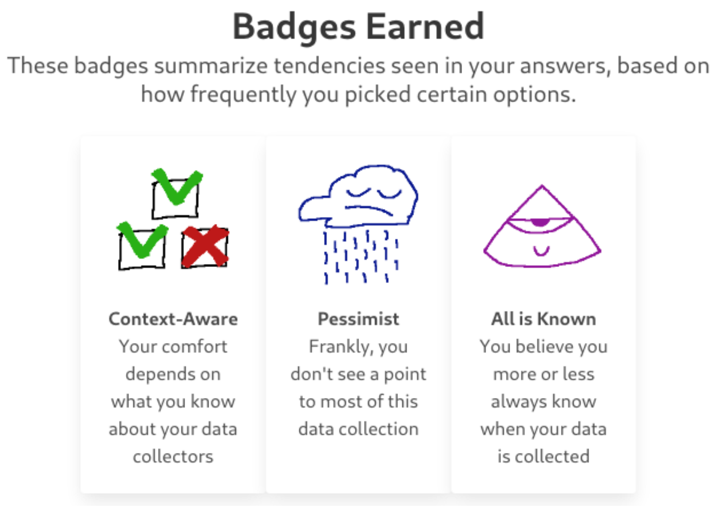
4.0
Awareness
How important you believe knowing about data collection and related information practices is.

6.0
Collection
The degree to which you are concerned about the amount of individual-specific data possessed by others relative to the value of benefits received.

Figure 4.7: IUIPC Score Display

IUIPC score is displayed in four categories: control, awareness, collection, and the overall average. The section itself is introduced with an explanation, its range, and a source for more information (which leads to Malhotra, Kim, Agarwal [23] paper). Each category also has a short description under it, and the score numbers in all categories are color-coded based on the value to intuitively express their meaning. The score received in this section is determined purely by the answers to the Privacy Score questions.

The badges represent player performance in the scenario-based questions. Each badge corresponds to a particular question across scenarios, and describes the emerging tendencies across all the scenarios (or lack thereof). Number of badges may vary depending on the answers, and up to three are displayed in one row. The graphics, along with playful titles and descriptions were included to present data in a more engaging way. Besides entertainment value, more engaging display of this information is intended to raise the chances of contemplation. The descriptions, rather than informing the user they answered X number of questions in Y way, provide guesses about player tendencies when faced with privacy questions. A full list of badges can be seen in Appendix B.



Badges Earned

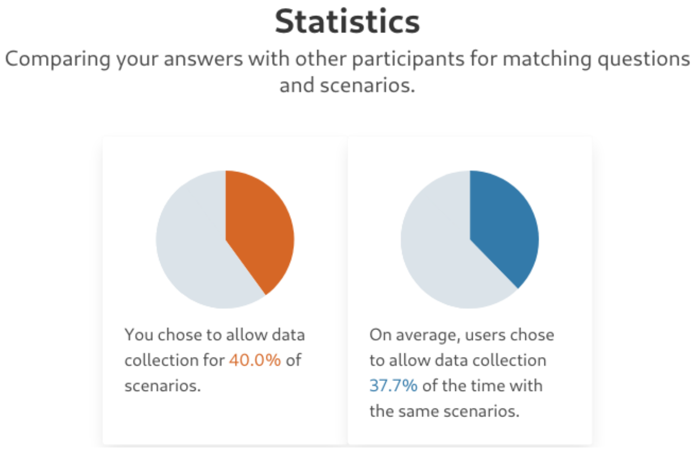
These badges summarize tendencies seen in your answers, based on how frequently you picked certain options.

Context-Aware
Your comfort depends on what you know about your data collectors

Pessimist
Frankly, you don't see a point to most of this data collection

All is Known
You believe you more or less always know when your data is collected

Figure 4.8: Badges display with 3 badges



Statistics

Comparing your answers with other participants for matching questions and scenarios.

You chose to allow data collection for 40.0% of scenarios.

On average, users chose to allow data collection 37.7% of the time with the same scenarios.

Figure 4.9: Statistics display with one comparison

The final part of the results is the Statistics section (Figure 4.9). Here the answering tendencies of the current player are compared with answers from previous participants. Answers are compared on the same questions *and* same scenarios (e.g. if there is data for question X with scenarios A, B, C, but the player only played A, only that scenario is taken into account) to provide a more accurate picture. There are 3 questions being compared in this section - control, awareness, and collection. Pie charts are used for visual aid. To have the overall statistics populated before any playthroughs are carried out, participant data from Naeini *et al.* [25] were used.

4.3 API

This section describes the various endpoints implemented in the Flask API that were used throughout the game. More detail is provided for parts of the implementation that were not revealed through its front-end descriptions in section (link previous section). Some endpoints are left out (e.g. */favicon*) due to their trivial nature.

4.3.1 GET Endpoints

/ (Default Endpoint) - Leads to Participant Information Sheet & Consent Form page.

/register - Leads to the Registration page. Reads the value of a consent variable which gets set if entering from the consent page.

/survey - Serves the scenario with a set of corresponding questions. It starts by receiving user ID from session variable that is meant to be set at the point of registration. If it is not set, the user is redirected to registration. A survey number is also tracked through a session variable, which indicates which scenario is needed. The scenarios are picked in order from a list of IDs. The database is then queried for the chosen scenario, as well as all relevant questions. Queries are also made to fetch followup relations between questions, and their corresponding Help tooltips. Each question is modeled into a JSON object for convenient access on the front-end. Followups and all other questions are split into two lists, so that the general questions can be iterated through in order. Each question object has a property indicating a followup question ID (if it exists) and its requirement.

/survey/next - Invoked on clicking “Continue” after finishing all questions within a given scenario. Increments the survey number counter and redirects to **/survey** endpoint.

/general - Serves a collection of IUIPC scale questions in a format similar to one used in **/survey**. Includes a registration check.

/completion - Serves all the data for the results screen, as well as updates the time table for the corresponding user with completion time. The database is queried for all relevant answers of the current user. For each Yes/No question that has corresponding badges a ratio of positive answers is calculated and checked against two thresholds. This splits the range of values into three sections, which map to (up to) 3 possible badges for each these questions (Example of badges for one question, along with their requirements can be seen in Table4.1). Multiple choice questions have a similar process, which is presented in more detail in Appendix B. Statistics that are presented in pie charts reuse the ratios of positive answers. In addition to this, it calculates this ratio for answers to the same (slightly differing in wording) questions in Naeini *et al.* [25] survey. The game uses a processed set of data from the previous survey, which is a table of positive answer percentages for each question in relation to the scenario. IUIPC score calculations map integer values from 1 to 7 to answers ranging from “Strongly Disagree” to “Strongly Agree” (e.g. “No Opinion” maps to 0. For a full list of options in IUIPC questions see section 4.2.2). Once the answers are converted to point values,

Question Text

If you had the choice, would you allow or deny this data collection?

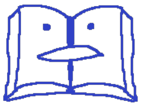
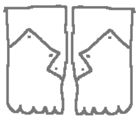
Requirement	Icon	Title	Description
$U \leq$		Open Book	You do not worry too much about your data, and usually accept the collection.
$L < x < U$		Situational	You are cautious about your privacy, and would rather disallow data collection altogether than risk it being compromised or misused.
$L \geq$		Iron Curtain	You decide to allow or deny data collection based on the information known, rather than having a predisposition

Table 4.1: The requirement field describes the required percentage of positive answers to receive the badge L stands for lower threshold, U stands for upper threshold. For the study, thresholds were set to 0.2 and 0.7, respectively.

medians of the questions in the three categories (control, awareness, collection) are calculated, as well as overall.

/onemore - Presents the user with an option to play an additional scenario. If the currently set collection of scenario IDs contains less items than the index an additional scenario would have, redirects to **/completion** .

4.3.2 POST endpoints

/submit_question - Used for submitting all answers. For questions with predefined answers, an entry is added to the answer table with appropriate IDs for: user, scenario, question, answer option. Due to the answer table only storing IDs, questions with custom answers (Highlight, Write-in) also cause an entry to be added to the options table. The ID of that new option is then used for the answer entry. This endpoint is invoked at the submission of each answer so that data is still collected in case of early termination.

/register - Used for saving user information.

Chapter 5

Evaluation

This chapter presents the data gathered from participation in the study and offers an interpretation of what tendencies can be observed. Section 5.1 covers data gathered from playing the game and revolves around user privacy preference findings, while Section 5.2 presents the results from the post-game survey and discusses the user experience aspect of the game.

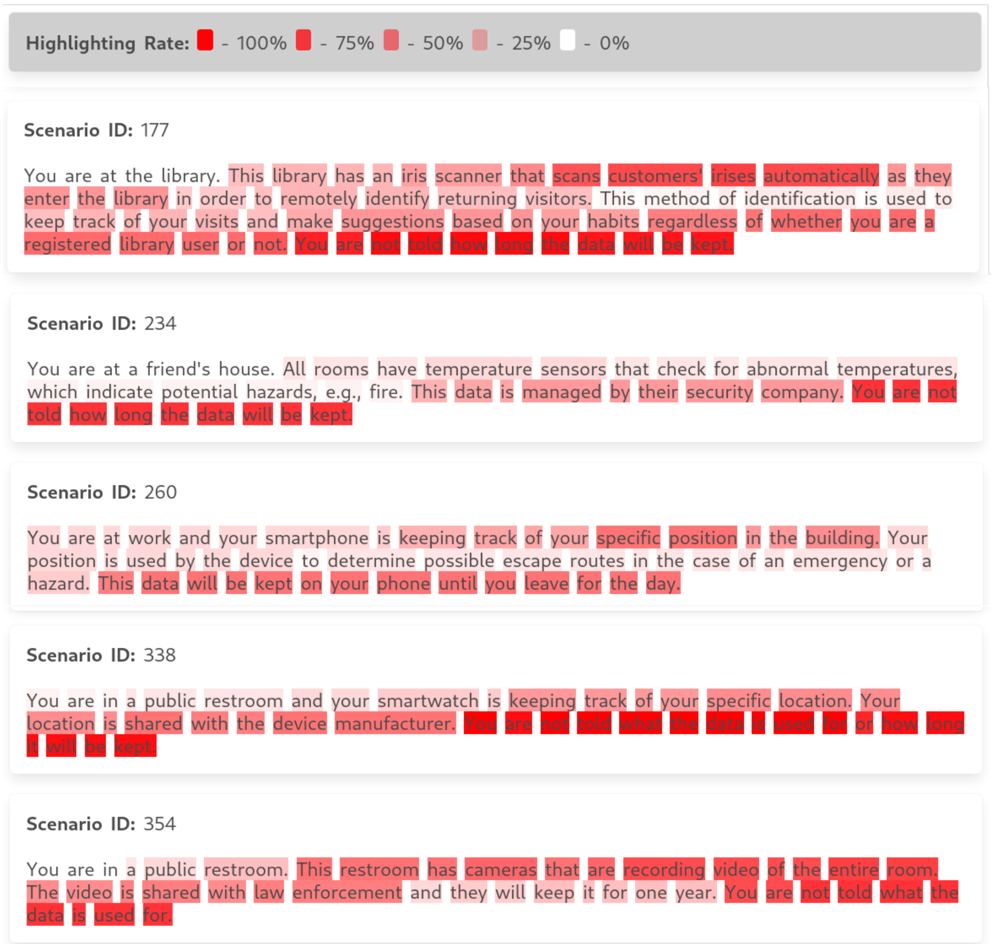
5.1 Findings on Privacy Preferences

The project aims to assess the privacy preference data collected through playing of the game. Overall, 24 playthroughs were recorded, and data from 19 were used for analysis in this section, as playthroughs that were incomplete (the player did not reach the results screen) were discarded. On average, it took 15 minutes to complete the game. All players fall in the age range of 19-25, and majority of them were undergraduate students. 63.2% of these players were pursuing a degree that was related to Computer Science, Privacy, or Cybersecurity.

Highlighting results were analysed with a heatmap generated based on highlighting frequency of each word across all participants (Figure 5.1). The intensity of the color is mapped to the percentage of participants that highlighted the area.

The most frequent source of discomfort seems to be an explicit lack of information - most participants marked sentences describing no information about data retention time or its purpose. This property was prioritised even when pervasive data collection methods were present, such as location trackers or cameras in bathrooms.

Yes/No questions were analysed in several ways. Firstly, answer counts were aggregated for each scenario. The results were plotted in stacked bar chart displaying percentage of all participants that chose that answer (Figure 5.2). The horizontal axis lists one-word descriptors and IDs of the questions the bars represent. The charts expose a correlation between willingness to allow data collection and these features: comfort, benefit, and essentiality to a service. Although in most scenarios safety impact of the data collection also exhibited this correlation, it is notable that safety improvements



Highlighting Rate: 🟥 - 100% 🟥 - 75% 🟥 - 50% 🟥 - 25% ⬜ - 0%

Scenario ID: 177

You are at the library. This library has an iris scanner that scans customers' irises automatically as they enter the library in order to remotely identify returning visitors. This method of identification is used to keep track of your visits and make suggestions based on your habits regardless of whether you are a registered library user or not. You are not told how long the data will be kept.

Scenario ID: 234

You are at a friend's house. All rooms have temperature sensors that check for abnormal temperatures, which indicate potential hazards, e.g., fire. This data is managed by their security company. You are not told how long the data will be kept.

Scenario ID: 260

You are at work and your smartphone is keeping track of your specific position in the building. Your position is used by the device to determine possible escape routes in the case of an emergency or a hazard. This data will be kept on your phone until you leave for the day.

Scenario ID: 338

You are in a public restroom and your smartwatch is keeping track of your specific location. Your location is shared with the device manufacturer. You are not told what the data is used for or how long it will be kept.

Scenario ID: 354

You are in a public restroom. This restroom has cameras that are recording video of the entire room. The video is shared with law enforcement and they will keep it for one year. You are not told what the data is used for.

Figure 5.1: Heatmap of highlights for part of the scenario that most affects user comfort

alone do not meaningfully increase the chances of willingness to share data, as exemplified by chart for scenario 354. Majority of participants believed that all scenarios except for 177 are likely to occur in present time. Almost all participants believed they would be informed of the data collection and would like to be notified about it in every scenario.

Another approach groups scenarios by their properties. It was observed that disclosure of purpose and retention time had the most significant effect on whether participants chose to allow data collection. This is shown by the pie charts in Figure 5.3. Note that having information about both of these properties greatly increases the likelihood of participants sharing data.

Overall Score	Control	Awareness	Collection
6.406	6.526	6.667	6.026

IUIPC score medians for each category across all participants were calculated. The overall score is close to the maximum value, indicating a high level of concern in all aspects of privacy within the participants. The scores indicate collection as the lesser concern out of the three, while awareness is seen as the highest priority. This lines up

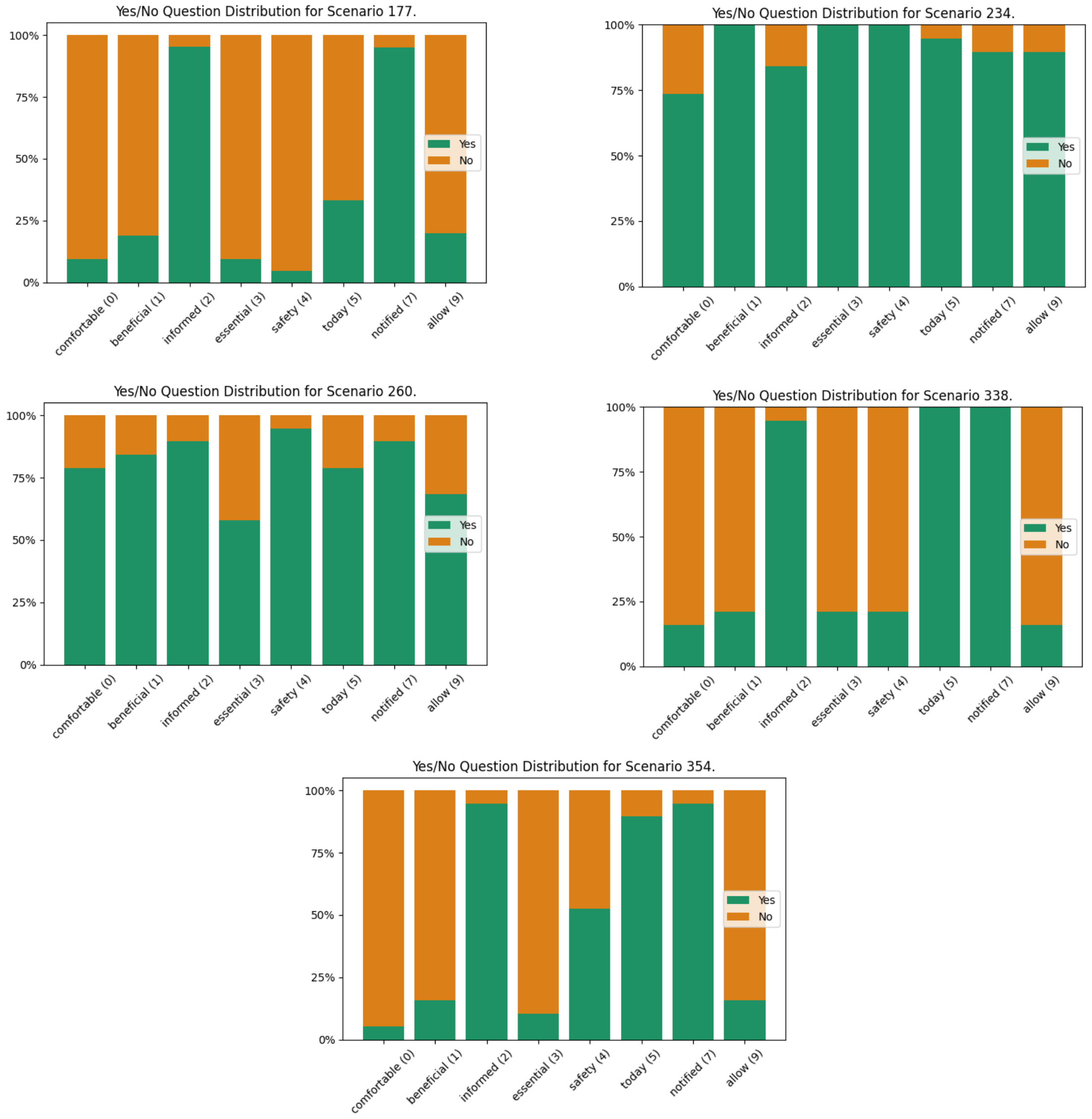


Figure 5.2: Responses to Yes/No questions for each of the 5 scenarios

with what we have seen in yes/no question analysis (where users universally chose that they wish to be informed of the data collection).

Observing IUIPC score median for each question (5.4) shows that participants agreed with the following statements the most:

- (Control) Consumer online privacy is really a matter of consumers' right to exercise control and autonomy over decisions about how their information is collected, used, and shared.
- (Awareness) I believe that online privacy is invaded when control is lost or unwillingly reduced as a result of a marketing transaction.
- (Awareness) Companies seeking information online should disclose the way the data are collected, processed and used.

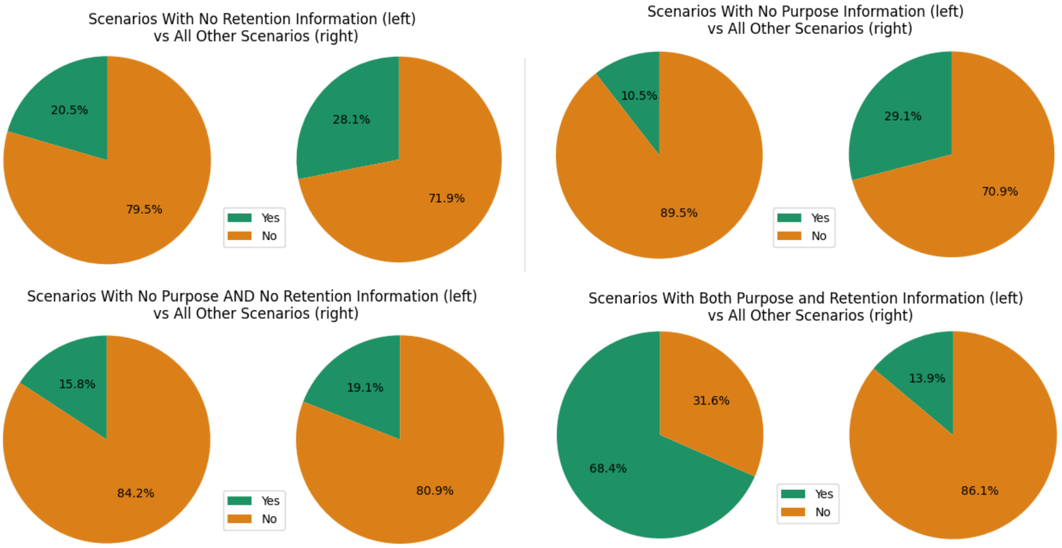


Figure 5.3: Choices to allow data collection grouped by retention and purpose information provided in scenario

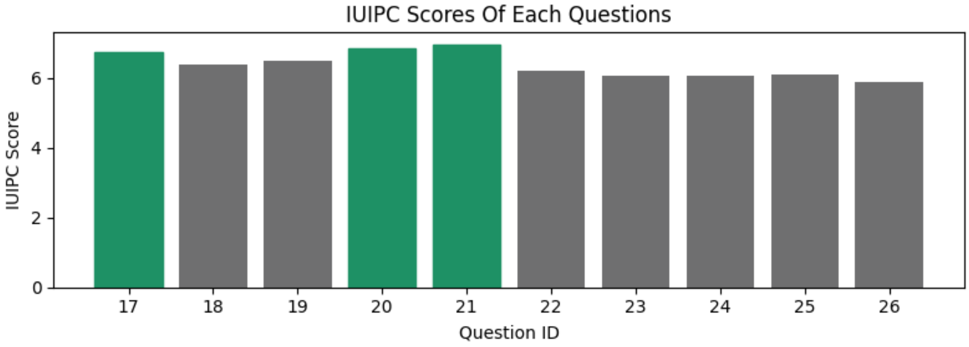


Figure 5.4: IUIPC Score for each question across all participants

5.2 User Experience

The data about user experience with the application was gathered through a post-game survey questionnaire hosted on Office 365 Forms [11]. The printout of the questionnaire can be found in Appendix D.

There were a total of 16 responses. Participants took 4 minutes on average to complete the questionnaire. There were 3 players that finished the game but did not complete the questionnaire. The estimate of the post-game survey taking up to 5 minutes was accurate, as the average completion time was 4 minutes and 2 seconds.

First, users were quizzed on their preference of question presentation. Questions in-game were compared to closest possible implementation of the same question in Office 365 Forms. Highlighting question was compared to writing the answer into a text box. Pre-filled write-in questions were compared to write-in questions that have initially empty text boxes. Participants overwhelmingly preferred the game implementation of each of these questions. Surprisingly, this applied even to the Yes/No question. Users quoted easier distinction between options at a glance, as well as aesthetic preference as their reasoning. As expected, there was highest variation of preference for this style of question. Participants preferred Highlighting and Write-In questions from the game largely for the same reason - convenience.

Question Type	Preferred Game	Preferred Form	No Preference
Yes/No	81.3%	6.2%	12.5%
Highlight (vs Form Write-In)	93.8%	0%	6.2%
Pre-Filled Write-In (vs Write-In)	87.5%	12.5%	0%

All participants thought they understood what badges represented, and 81.3% found them useful. In contrast, 81.3% believed they understood what was meant by IUIPC score, and thus in turn only 62.5% found it useful. While the badges were expected to be intuitive and easy to understand, the number of participants that found them useful exceeded expectations, as their function was partly entertainment value.

Only half of the participants used the Help button, and out of them, 75% found the help text useful. It was expected for only fraction of users to use the help button, however the clarity of the help text seems lacking.

The free-form section of the questionnaire revealed that the most requested feature was a back/undo button. While the application was intentionally designed to disallow navigation between questions, functionality that would allow recovery from errors could be implemented without interfering with this goal. The lack of this feature breaks a fundamental design principle and was independently noticed by 25% of respondents. Other recurring observations feature confusion regarding the wording of questions, most notably one asking whether the data collection is “essential to the service provided”. Some scenarios chosen for the study do not have a clear “service”. Most participants enjoyed using the highlighting feature, although some experienced frustration as well with bugs and unintended behaviour. Any responses mentioning the results screen were positive.

Chapter 6

Conclusion

The game implementation was successful. Users overall reacted favourably to the new features introduced to the standard questionnaire model. The highlighting feature worked well in most cases and was found to be more convenient than the alternatives. Users also appreciated the inclusion of the results screen and found the information presented within it clear and useful. Although it was not prioritised, even the presentation of Yes/No questions was appreciated and widely preferred within the game implementation rather than standard questionnaires, showing that user experience considerations are noticed even in familiar features. The help feature was found and used by participants that needed it, and while it helped in majority of cases, it was found that the text could be even clearer. The most glaring flaw of the implementation was the lack of an undo button. This, combined with no secondary confirmation when providing input proved frustrating for multiple users.

The game allowed us to collect player privacy preferences. Analysis of these preferences revealed that users surveyed first and foremost value being informed about data collection that is occurring, irregardless of context. Users were found to be significantly more accepting of data collection when both retention time and purpose for the data was disclosed. This finding was reinforced by both scenario question results, as well as IUIPC scores, as users were shown to be most concerned about the “Awareness” aspect.

The highlighting feature confirmed the importance of disclosure, but also provided insight into features found concerning by users that were not covered by other questions explicitly - the method and nature of the data collection. Namely, users comfort was affected by pervasive devices such as iris scanners, and the tracking of their specific location within a building. These concerns likely stem from current rarity of the former device, and both may be seen as gathering information that is deemed too personal. Finally, it was found that data being used for safety is not enough to increase the acceptance rate of collection, while data being not needed for the purposes of the service provided are highly correlated with disallowing collection.

Overall, it can be said that the project achieved the objectives set at the beginning, but there is space for improvement.

6.1 Future Improvements

One of the ideas informing the design was reducing the influence previous questions had on further answers. This is diminished by the current implementation, with questions being repeated for multiple scenarios, however an alternative approach would have required significantly more participants to get a meaningful quantity of data to evaluate.

While the game was successful in the function of collecting data (this aspect is explored further in the Results chapter), its features as a game may be lacking. This is mainly due to the challenge of trying to consolidate meaningful data collection with creative interactivity. The most notable interactive feature, scenario highlighting, is underutilised in the current question set, being utilised in only two questions in the final survey. On the other hand, the questions that used this feature did utilise it to its fullest extent, having a clear question that benefits from this method of answering. Furthermore, it does not overstay its welcome in case users find it unusable or annoying. With confirmation that the feature was well received, it is more justifiable to user it more prevalently in future iterations.

Users responded favourably to imagery in badges, this idea could be expanded to the game loop itself, providing visual information generated from scenario context. This could be as simple as displaying an image of a device that is collecting the data, to a system that populates a picture with objects based on the properties of the scenario. To expand even further, this idea could be used to create an interactive visual environment, allowing to simply click on an object (or a series of them) to describe privacy concerns.

The results screen could have included more analysis of responses and comparisons across all players, including some of the data representations seen in this paper. This was not included purely due to time constraints. A decision had to be made between releasing the survey earlier in order to get a bigger participant pool, or include more features risking lower turnout. The former option was chosen and further analysis code was written while the game was open to the public. This was done in order to have more meaningful results, but there is no reason to not include more statistics for users in a future iteration, as participants expressed interest in them. Comparisons seen in Statistics section 4.9 could be applied to other parts of the results screen by e.g. displaying the frequency any particular badge is earned overall. Showing where the user lands in comparison to other participants with their IUIPC scores could prove beneficial in understanding the meaning behind the numbers better as well.

Finally, a big feature that could contribute to the secondary goal of educating users would be personalised educational information about privacy at the results screen. For example, participants that neglect data retention policies could be informed about the frequency of data breaches even at large established companies [12] and even directed to user-friendly hands-on illustrations of this such as *haveibeenpwned.com*. It would be interesting to see a pivot of the game that focuses on education as a primary goal, and preference collection as secondary, allowing for more learning opportunities that would otherwise be left out in order to not influence the answers within the survey.

Bibliography

[1] Tulips. `https://groups.inf.ed.ac.uk/tulips/student_projects.html`. Accessed 2020-08-07.

[2] Rusab Abrez Asher. Vulcan - the video game to teach informatics students about firewalls. 2019.

[3] asyncpg. `https://github.com/MagicStack/asyncpg`. Accessed 2020-09-22.

[4] Yolum Ayci. Recommending privacy labels for online content. 2020.

[5] Igor Bilogrevic, Kévin Huguenin, Berker Agir, Murtuza Jadliwala, Maria Gazaki, and Jean-Pierre Hubaux. A machine-learning based approach to privacy-aware information-sharing in mobile social networks. *Pervasive and Mobile Computing*, 25:125–142, January 2016.

[6] Blue team : A firewall setup game. `https://github.com/karel12/FirewallInternship`. Accessed 2020-08-07.

[7] Bulma. `https://bulma.io/`. Accessed 2020-10-11.

[8] Jessica Colnago, Yuanyuan Feng, Tharangini Palanivel, Sarah Pearman, Megan Ung, Alessandro Acquisti, Lorrie Faith Cranor, and Norman Sadeh. Informing the design of a personalized privacy assistant for the internet of things. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, page 1–13, New York, NY, USA, 2020. Association for Computing Machinery.

[9] Django. `https://www.djangoproject.com/`. Accessed 2021-01-12.

[10] Flask. `https://flask.palletsprojects.com/en/1.1.x/`. Accessed 2020-09-20.

[11] Office 365 forms. `https://www.microsoft.com/en-gb/microsoft-365/online-surveys-polls-quizzes`. Accessed 2021-03-15.

[12] Adam Forrest. Facebook data scandal: Social network fined $5bn over 'inappropriate' sharing of users' personal information. *The Guardian*.

[13] Github. `https://github.com/`.

[14] Johannes Harms, Stefan Biegler, Christoph Wimmer, Karin Kappel, and Thomas Grechenig. Gamification of online surveys: Design process, case study, and eval-

uation. volume 9296 of *Lecture Notes in Computer Science*, pages 219–236, Cham, 2015. Springer International Publishing.

[15] Heroku. `https://www.heroku.com/`. Accessed 2021-02-17.

[16] Javascript. `https://www.javascript.com/`. Accessed 2020-09-10.

[17] Jinja2. `https://jinja2docs.readthedocs.io/en/stable/`. Accessed 2020-09-23.

[18] Murat Kezer, Barış Sevi, Zeynep Cemalcilar, and Lemi Baruh. Age differences in privacy attitudes, literacy and privacy management on facebook. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, 10, 2016.

[19] Nadin Kökciyan and Pinar Yolum. Context-based reasoning on privacy in internet of things. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence*, IJCAI'17, page 4738–4744. AAAI Press, 2017.

[20] Raph Koster. *A theory of fun for game design*. 2005.

[21] Abdurrahman Can Kurtan and Pinar Yolum. Pelte: Privacy estimation of images from tags. In *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems*, AAMAS '18, page 1989–1991, Richland, SC, 2018. International Foundation for Autonomous Agents and Multiagent Systems.

[22] S. Loewen, D. Crowther, Daniel R. Isbell, K. Kim, J. Maloney, Zachary F. Miller, and Hima Rawal. Mobile-assisted language learning: A duolingo case study. *ReCALL*, 31:293–311, 2019.

[23] N. Malhotra, S. Kim, and J. Agarwal. Internet users' information privacy concerns (iuipc): The construct, the scale, and a causal model. *Inf. Syst. Res.*, 15:336–355, 2004.

[24] J. Mugan, T. Sharma, and N. Sadeh. Understandable learning of privacy preferences through default personas and suggestions. 2011.

[25] Pardis Emami Naeini, Sruti Bhagavatula, Hana Habib, Martin Degeling, Lujo Bauer, Lorrie Faith Cranor, and Norman Sadeh. Privacy expectations and preferences in an iot world. In *Thirteenth Symposium on Usable Privacy and Security (SOUPS 2017)*, pages 399–412, Santa Clara, CA, July 2017. USENIX Association.

[26] Helen Nissenbaum. Respecting Context to Protect Privacy: Why Meaning Matters. *Science and Engineering Ethics*, 24(3):831–852, June 2018.

[27] Helen Fay. Nissenbaum. *Privacy in context: technology, policy, and the integrity of social life*. Stanford Law Books, 2010.

[28] Numpy. `https://numpy.org/`. Accessed 2020-11-12.

[29] Papers, please. `https://store.steampowered.com/app/239030/Papers_Please/`. Accessed 2021-04-04.

[30] Postgresql. `https://www.postgresql.org/`. Accessed 2021-02-17.

[31] Ismini Psychoula, Deepika Singh, Liming Chen, Feng Chen, Andreas Holzinger, and Huansheng Ning. Users' privacy concerns in iot based applications. 09 2018.

[32] Python. `https://www.python.org/`. Accessed 2020-09-20.

[33] Michael Sailer and Lisa Homner. The gamification of learning: a meta-analysis. *Educational Psychology Review*, 32:77–112, 03 2020.

[34] Sibylle Sehl. Permission impossible - the design and evaluation of a video game that teaches beginners about firewalls. 2017.

[35] Sqlalchemy. `https://flask-sqlalchemy.palletsprojects.com/en/2.x/`. Accessed 2020-09-22.

[36] Tis-100. `http://www.zachtronics.com/tis-100/`. Accessed 2020-08-05.

[37] Yang Wang, Gregory Norcie, Saranga Komanduri, Alessandro Acquisti, Pedro Giovanni Leon, and Lorrie Faith Cranor. "i regretted the minute i pressed share": A qualitative study of regrets on facebook. In *Proceedings of the Seventh Symposium on Usable Privacy and Security*, SOUPS '11, New York, NY, USA, 2011. Association for Computing Machinery.

[38] Sergej Zerr, Stefan Siersdorfer, Jonathon Hare, and Elena Demidova. Privacy-aware image classification and search. In *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval - SIGIR '12*, page 35, Portland, Oregon, USA, 2012. ACM Press.

Appendix A

Optional Scenarios

Scenario ID: 4
You are at a department store and your smartwatch is keeping track of your specific position in the department store. Your position is used by the device to determine possible escape routes in the case of an emergency or a hazard. You are not told how long the data will be kept.

Scenario ID: 95
You are at the library and your smartwatch is keeping track of your specific position. This data will be kept on your watch for one week. You are not told what the data is used for.

Scenario ID: 139
You are at the library. This library has cameras that are recording video of the entire library. The video is shared with law enforcement and they will not delete it. You are not told what the data is used for.

Scenario ID: 187
You are at the library. This library has a facial recognition system that scans customers' faces automatically as they enter the library. Your picture will not be deleted. You are not told what the data is used for.

Scenario ID: 189
You are at home. All rooms have presence sensors that are used to determine when to switch on and off the lights to reduce costs and save energy. This data will be kept for one week.

Scenario ID: 200
You are at home. All rooms have cameras that are recording video of the entire room you're in. The video is shared with law enforcement to improve public safety. They will also use it to infer your movement patterns, e.g., where and how you spend your time. You are not told how long the data will be kept.

Scenario ID: 294
You are at work. This building has presence sensors that are used to determine how

busy it is in order to optimize heating and cooling to make the employees most comfortable. This data will be kept for one year.

Scenario ID: 320
You are at work. This building has a fingerprint scanner. Your fingerprint will not be deleted. You are not told what the data is used for.

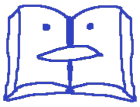
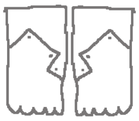
Scenario ID: 376
You are in a public restroom. This restroom has temperature sensors that check for abnormal temperatures, which indicate potential hazards, e.g., fire. This data will not be deleted.

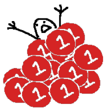
Appendix B

Badges

B.1 Binary Choice Badges

This is the full list of badges for questions that present two possible answers. The requirement field describes the required percentage of positive answers to receive the badge L stands for lower threshold, U stands for upper threshold.

Question	Requirement	Icon	Title	Description
If you had the choice, would you allow or deny this data collection?	$U \leq$		Open Book	You do not worry too much about your data collection, and usually accept the collection.
	$L < x < U$		Situational	You are cautious about your privacy, and would rather disallow data collection altogether than risk it being compromised or misused.
	$L \geq$		Iron Curtain	You decide to allow or deny data collection based on the information known, rather than having a predisposition
Would you feel comfortable with data collection in this scenario?	$U \leq$		Chill Pill	You don't worry too much about the potential use of your data.
	$L < x < U$		Context-Aware	Allowing your data to be used usually leads to discomfort.
	$L \geq$		Sus	Your comfort depends on what you know about your data collectors

Do you believe this data collection is beneficial to you?	$U \leq$		Optimist	You see benefit in most data collection scenarios.
	$L < x < U$		Economist	Frankly, you don't see a point to most of this data collection.
	$L \geq$		Pessimist	You evaluate the benefits of data collection on a case-by-case basis
Would you expect to be informed of data collection in this scenario?	$U \leq$		All is Known	You believe you more or less always know when your data is collected.
	$L < x < U$	-	-	-
	$L \geq$		Cautious	You don't think you would be told about data collection most of the time.
Do you believe this data collection is essential to the service being provided?	$U \leq$		Essential Worker	You believe most of the time the data being collected is essential to core functionality.
	$L < x < U$	-	-	-
	$L \geq$		Bloated Software	Most of this data collection is just stitched on to stuff that could work fine without it!
Do you believe this data collection improves safety?	$U \leq$		Patriot Act	You believe most of this data collection improves safety.
	$L < x < U$	-	-	-
	$L \geq$	-	-	-
Do you believe this scenario is occuring today?	$U \leq$		They Live	You think most of this data collection is occuring right here and now!
	$L < x < U$	-	-	-
	$L \geq$		Sci-fi	You think most of these scenarios are long way from being realized
Would you like to be notified of data collection in this scenario?	$U \leq$		Open The Flood Gates	You'd like to be notified almost every time your data is taken
	$L < x < U$		Priority Queue	You like getting informed about important data collection, as long as it's not too annoying
	$L \geq$		Too Many !@&* Notifications!!	You usually prefer little to no notifications, even if they are about privacy

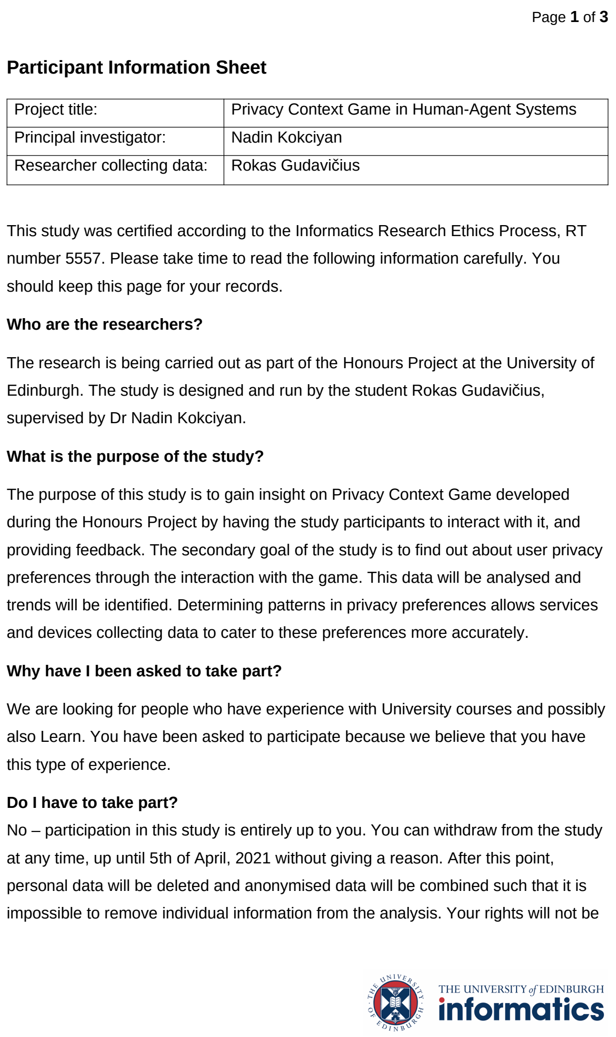
B.2 Multiple Choice Badges

Question	Requirement	Icon	Title	Description
How often would you like to be notified of this data collection by your mobile phone?	”Every time this occurs” $\geq U$		Ping! Ping! Ping!	You would prefer to always know when data is being collected.
	”First time this occurs” $\geq U$		Today I Learned	Being notified once is enough for you to keep in mind when your data is collected.
	”Occassionally” $\geq U$		Remind Me Later	You’d like to get reminders about data collection more than once.
	”Never” $\geq U$		See No Evil	YYou receive too many notifications already.
Scenarios like this will occur in X years.	n(”2”) + n(”5”) $\geq U$		Right Around The Corner	You would prefer to always know when data is being collected.
	n(”10”) + n(”15+”) $\geq U$		The Year Is 2525...	Being notified once is enough for you to keep in mind when your data is collected.
	”Will Never Occur” $\geq U$		Preposterous!	YYou receive too many notifications already.

Appendix C

Documents

C.1 Participant Information Sheet



Page **1** of **3**

Participant Information Sheet

Project title:	Privacy Context Game in Human-Agent Systems
Principal investigator:	Nadin Kokciyan
Researcher collecting data:	Rokas Gudavičius

This study was certified according to the Informatics Research Ethics Process, RT number 5557. Please take time to read the following information carefully. You should keep this page for your records.

Who are the researchers?

The research is being carried out as part of the Honours Project at the University of Edinburgh. The study is designed and run by the student Rokas Gudavičius, supervised by Dr Nadin Kokciyan.

What is the purpose of the study?

The purpose of this study is to gain insight on Privacy Context Game developed during the Honours Project by having the study participants to interact with it, and providing feedback. The secondary goal of the study is to find out about user privacy preferences through the interaction with the game. This data will be analysed and trends will be identified. Determining patterns in privacy preferences allows services and devices collecting data to cater to these preferences more accurately.

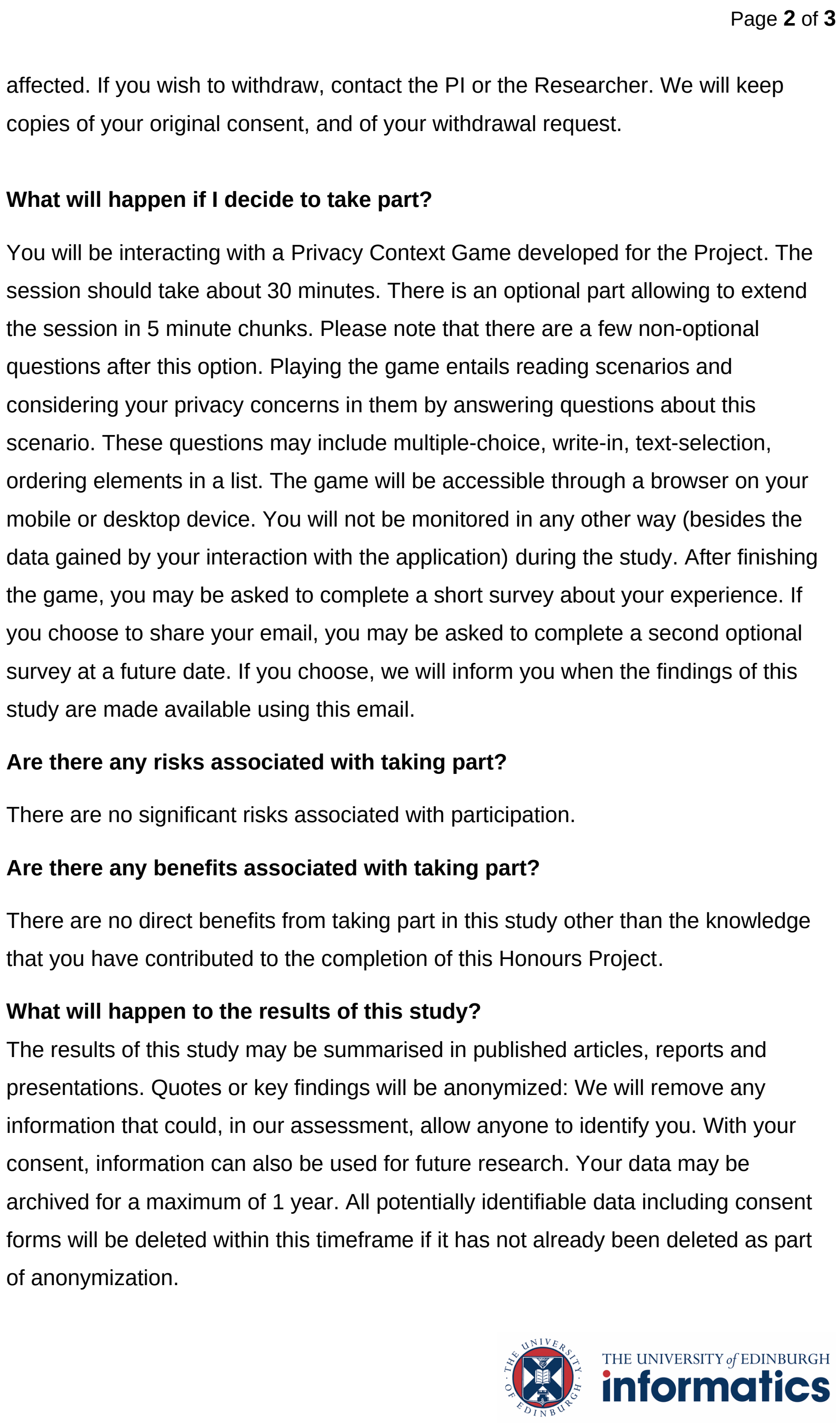
Why have I been asked to take part?

We are looking for people who have experience with University courses and possibly also Learn. You have been asked to participate because we believe that you have this type of experience.

Do I have to take part?

No – participation in this study is entirely up to you. You can withdraw from the study at any time, up until 5th of April, 2021 without giving a reason. After this point, personal data will be deleted and anonymised data will be combined such that it is impossible to remove individual information from the analysis. Your rights will not be

THE UNIVERSITY *of* EDINBURGH
informatics



Page **2** of **3**

affected. If you wish to withdraw, contact the PI or the Researcher. We will keep copies of your original consent, and of your withdrawal request.

What will happen if I decide to take part?

You will be interacting with a Privacy Context Game developed for the Project. The session should take about 30 minutes. There is an optional part allowing to extend the session in 5 minute chunks. Please note that there are a few non-optional questions after this option. Playing the game entails reading scenarios and considering your privacy concerns in them by answering questions about this scenario. These questions may include multiple-choice, write-in, text-selection, ordering elements in a list. The game will be accessible through a browser on your mobile or desktop device. You will not be monitored in any other way (besides the data gained by your interaction with the application) during the study. After finishing the game, you may be asked to complete a short survey about your experience. If you choose to share your email, you may be asked to complete a second optional survey at a future date. If you choose, we will inform you when the findings of this study are made available using this email.

Are there any risks associated with taking part?

There are no significant risks associated with participation.

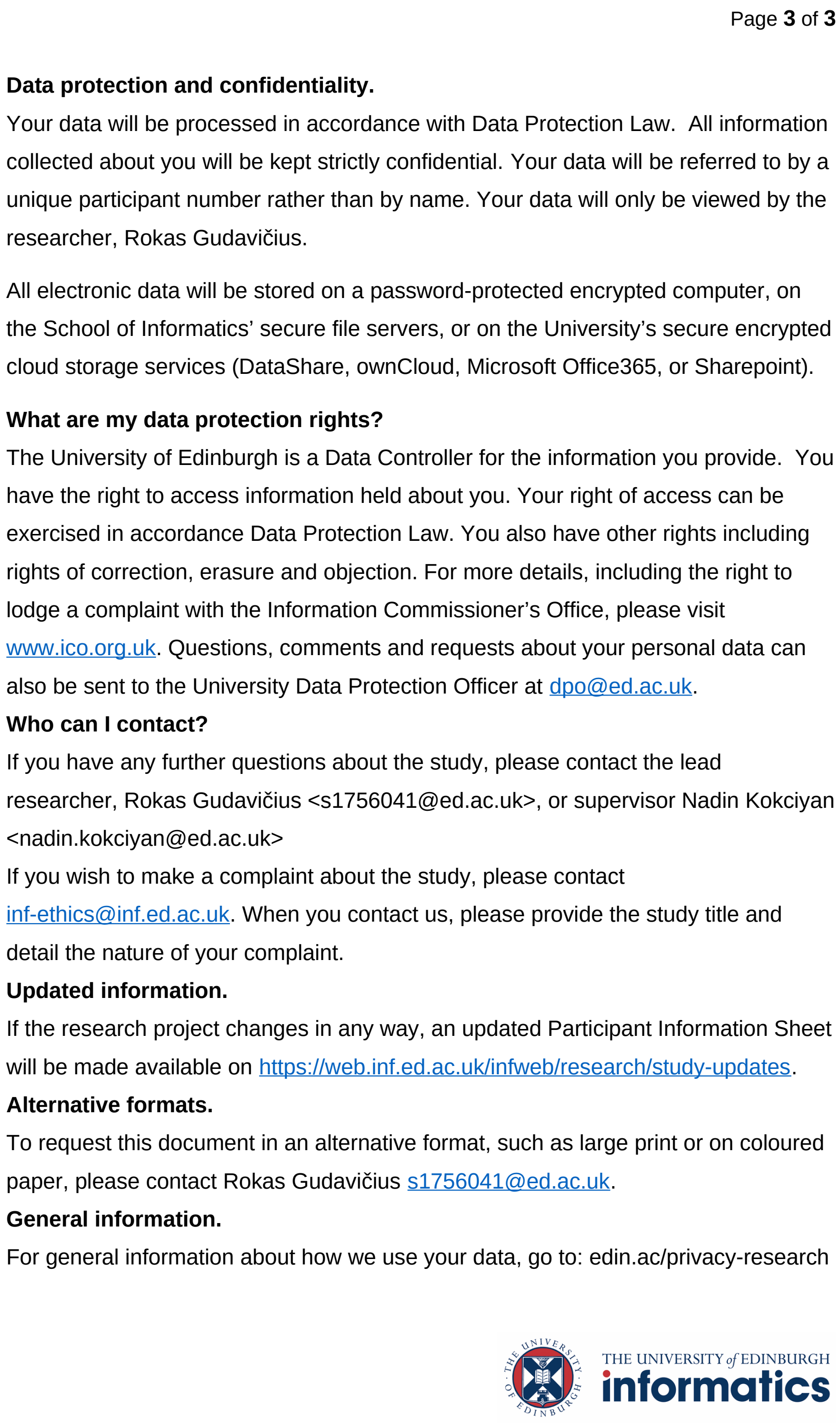
Are there any benefits associated with taking part?

There are no direct benefits from taking part in this study other than the knowledge that you have contributed to the completion of this Honours Project.

What will happen to the results of this study?

The results of this study may be summarised in published articles, reports and presentations. Quotes or key findings will be anonymized: We will remove any information that could, in our assessment, allow anyone to identify you. With your consent, information can also be used for future research. Your data may be archived for a maximum of 1 year. All potentially identifiable data including consent forms will be deleted within this timeframe if it has not already been deleted as part of anonymization.

THE UNIVERSITY of EDINBURGH informatics



Page **3** of **3**

Data protection and confidentiality.

Your data will be processed in accordance with Data Protection Law. All information collected about you will be kept strictly confidential. Your data will be referred to by a unique participant number rather than by name. Your data will only be viewed by the researcher, Rokas Gudavičius.

All electronic data will be stored on a password-protected encrypted computer, on the School of Informatics' secure file servers, or on the University's secure encrypted cloud storage services (DataShare, ownCloud, Microsoft Office365, or Sharepoint).

What are my data protection rights?

The University of Edinburgh is a Data Controller for the information you provide. You have the right to access information held about you. Your right of access can be exercised in accordance Data Protection Law. You also have other rights including rights of correction, erasure and objection. For more details, including the right to lodge a complaint with the Information Commissioner's Office, please visit www.ico.org.uk. Questions, comments and requests about your personal data can also be sent to the University Data Protection Officer at dpo@ed.ac.uk.

Who can I contact?

If you have any further questions about the study, please contact the lead researcher, Rokas Gudavičius <s1756041@ed.ac.uk>, or supervisor Nadin Kokciyan <nadin.kokciyan@ed.ac.uk>

If you wish to make a complaint about the study, please contact inf-ethics@inf.ed.ac.uk. When you contact us, please provide the study title and detail the nature of your complaint.

Updated information.

If the research project changes in any way, an updated Participant Information Sheet will be made available on https://web.inf.ed.ac.uk/infweb/research/study-updates.

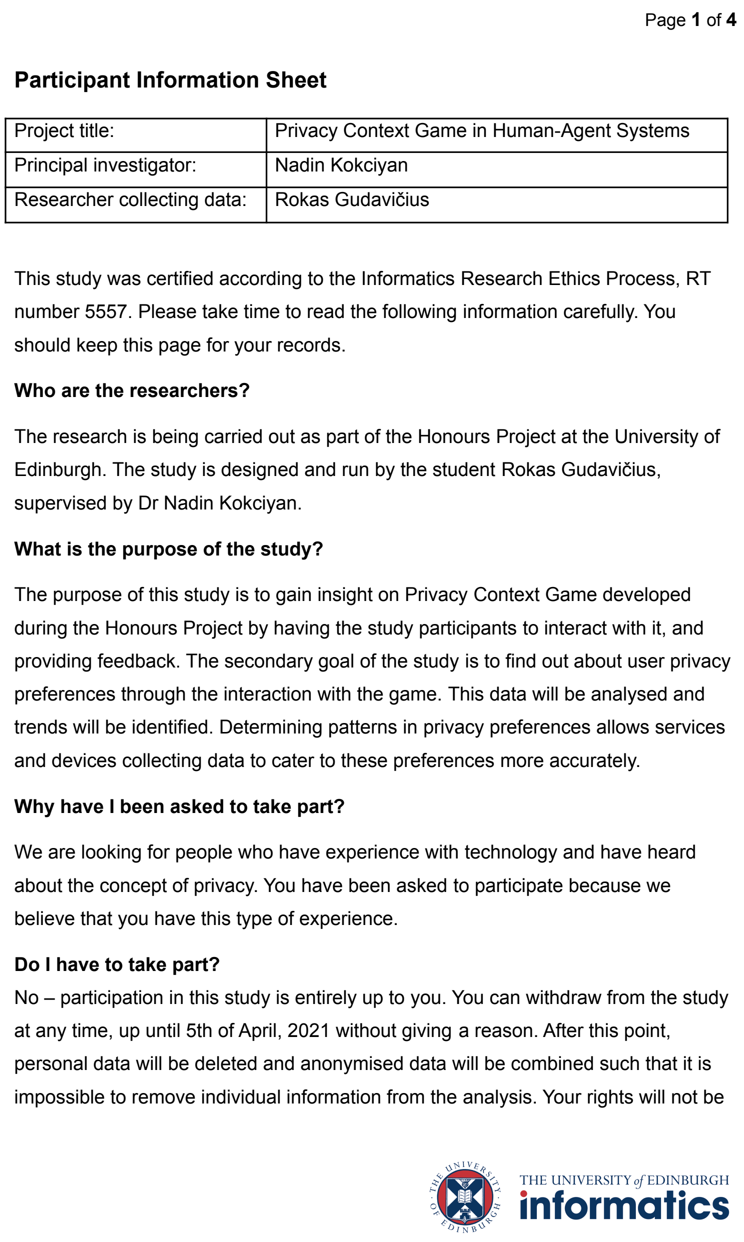
Alternative formats.

To request this document in an alternative format, such as large print or on coloured paper, please contact Rokas Gudavičius s1756041@ed.ac.uk.

General information.

For general information about how we use your data, go to: edin.ac/privacy-research

C.2 Consent Form



Page **1** of **4**

Participant Information Sheet

Project title:	Privacy Context Game in Human-Agent Systems
Principal investigator:	Nadin Kokciyan
Researcher collecting data:	Rokas Gudavičius

This study was certified according to the Informatics Research Ethics Process, RT number 5557. Please take time to read the following information carefully. You should keep this page for your records.

Who are the researchers?

The research is being carried out as part of the Honours Project at the University of Edinburgh. The study is designed and run by the student Rokas Gudavičius, supervised by Dr Nadin Kokciyan.

What is the purpose of the study?

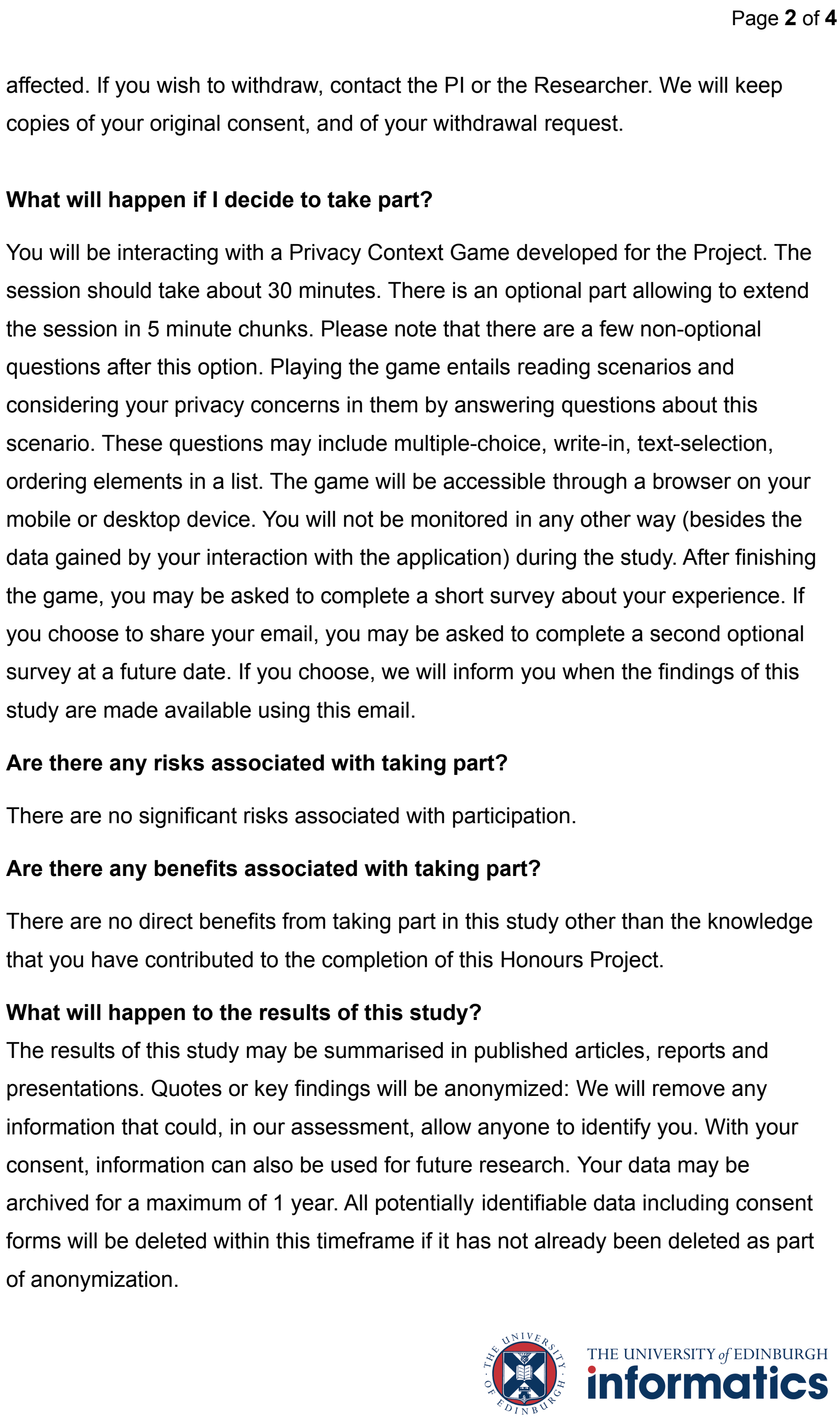
The purpose of this study is to gain insight on Privacy Context Game developed during the Honours Project by having the study participants to interact with it, and providing feedback. The secondary goal of the study is to find out about user privacy preferences through the interaction with the game. This data will be analysed and trends will be identified. Determining patterns in privacy preferences allows services and devices collecting data to cater to these preferences more accurately.

Why have I been asked to take part?

We are looking for people who have experience with technology and have heard about the concept of privacy. You have been asked to participate because we believe that you have this type of experience.

Do I have to take part?

No – participation in this study is entirely up to you. You can withdraw from the study at any time, up until 5th of April, 2021 without giving a reason. After this point, personal data will be deleted and anonymised data will be combined such that it is impossible to remove individual information from the analysis. Your rights will not be



affected. If you wish to withdraw, contact the PI or the Researcher. We will keep copies of your original consent, and of your withdrawal request.

What will happen if I decide to take part?

You will be interacting with a Privacy Context Game developed for the Project. The session should take about 30 minutes. There is an optional part allowing to extend the session in 5 minute chunks. Please note that there are a few non-optional questions after this option. Playing the game entails reading scenarios and considering your privacy concerns in them by answering questions about this scenario. These questions may include multiple-choice, write-in, text-selection, ordering elements in a list. The game will be accessible through a browser on your mobile or desktop device. You will not be monitored in any other way (besides the data gained by your interaction with the application) during the study. After finishing the game, you may be asked to complete a short survey about your experience. If you choose to share your email, you may be asked to complete a second optional survey at a future date. If you choose, we will inform you when the findings of this study are made available using this email.

Are there any risks associated with taking part?

There are no significant risks associated with participation.

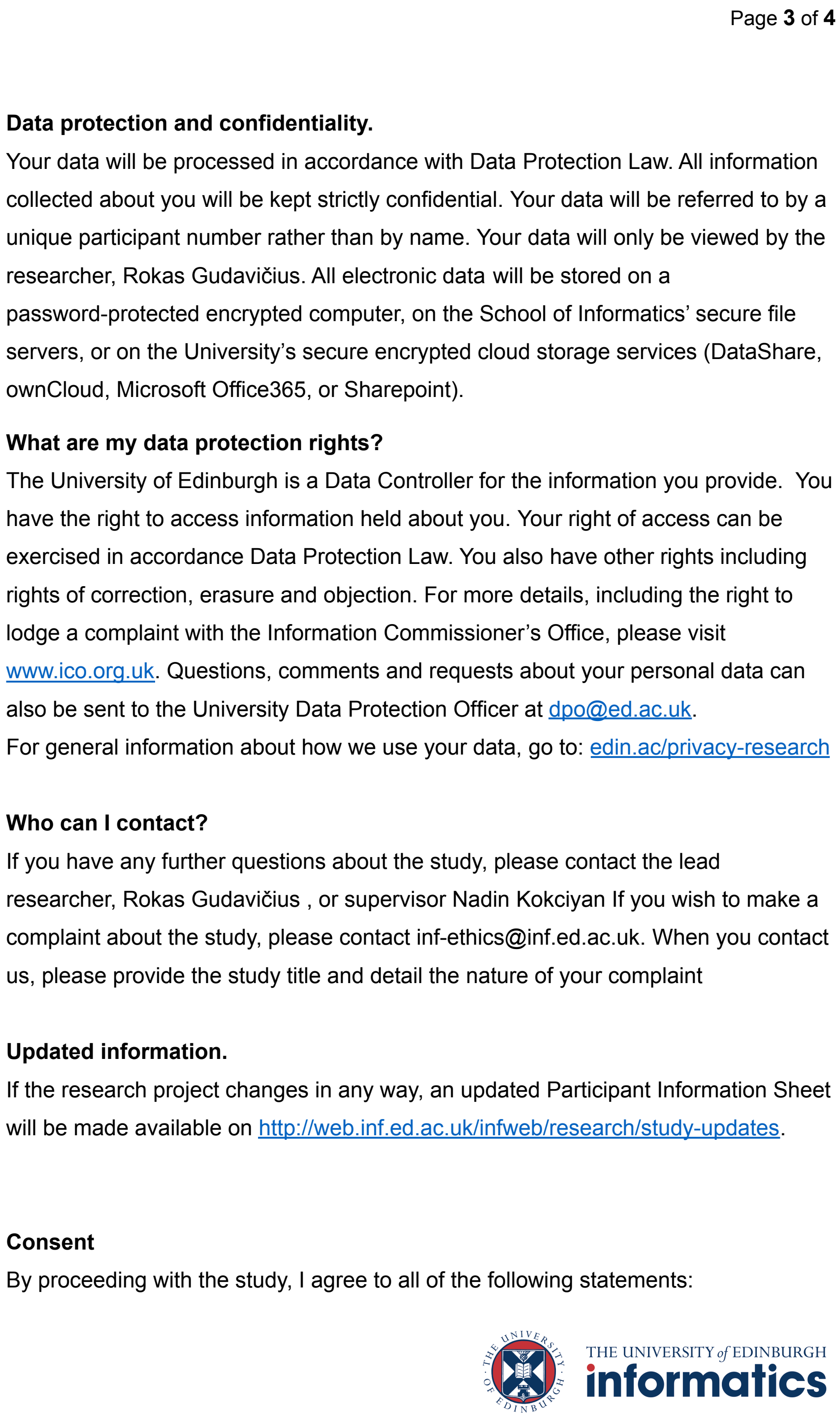
Are there any benefits associated with taking part?

There are no direct benefits from taking part in this study other than the knowledge that you have contributed to the completion of this Honours Project.

What will happen to the results of this study?

The results of this study may be summarised in published articles, reports and presentations. Quotes or key findings will be anonymized: We will remove any information that could, in our assessment, allow anyone to identify you. With your consent, information can also be used for future research. Your data may be archived for a maximum of 1 year. All potentially identifiable data including consent forms will be deleted within this timeframe if it has not already been deleted as part of anonymization.

THE UNIVERSITY *of* EDINBURGH
informatics



Data protection and confidentiality.

Your data will be processed in accordance with Data Protection Law. All information collected about you will be kept strictly confidential. Your data will be referred to by a unique participant number rather than by name. Your data will only be viewed by the researcher, Rokas Gudavičius. All electronic data will be stored on a password-protected encrypted computer, on the School of Informatics' secure file servers, or on the University's secure encrypted cloud storage services (DataShare, ownCloud, Microsoft Office365, or Sharepoint).

What are my data protection rights?

The University of Edinburgh is a Data Controller for the information you provide. You have the right to access information held about you. Your right of access can be exercised in accordance Data Protection Law. You also have other rights including rights of correction, erasure and objection. For more details, including the right to lodge a complaint with the Information Commissioner's Office, please visit www.ico.org.uk. Questions, comments and requests about your personal data can also be sent to the University Data Protection Officer at dpo@ed.ac.uk.

For general information about how we use your data, go to: edin.ac/privacy-research

Who can I contact?

If you have any further questions about the study, please contact the lead researcher, Rokas Gudavičius , or supervisor Nadin Kokciyan If you wish to make a complaint about the study, please contact inf-ethics@inf.ed.ac.uk. When you contact us, please provide the study title and detail the nature of your complaint

Updated information.

If the research project changes in any way, an updated Participant Information Sheet will be made available on http://web.inf.ed.ac.uk/infweb/research/study-updates.

Consent

By proceeding with the study, I agree to all of the following statements:

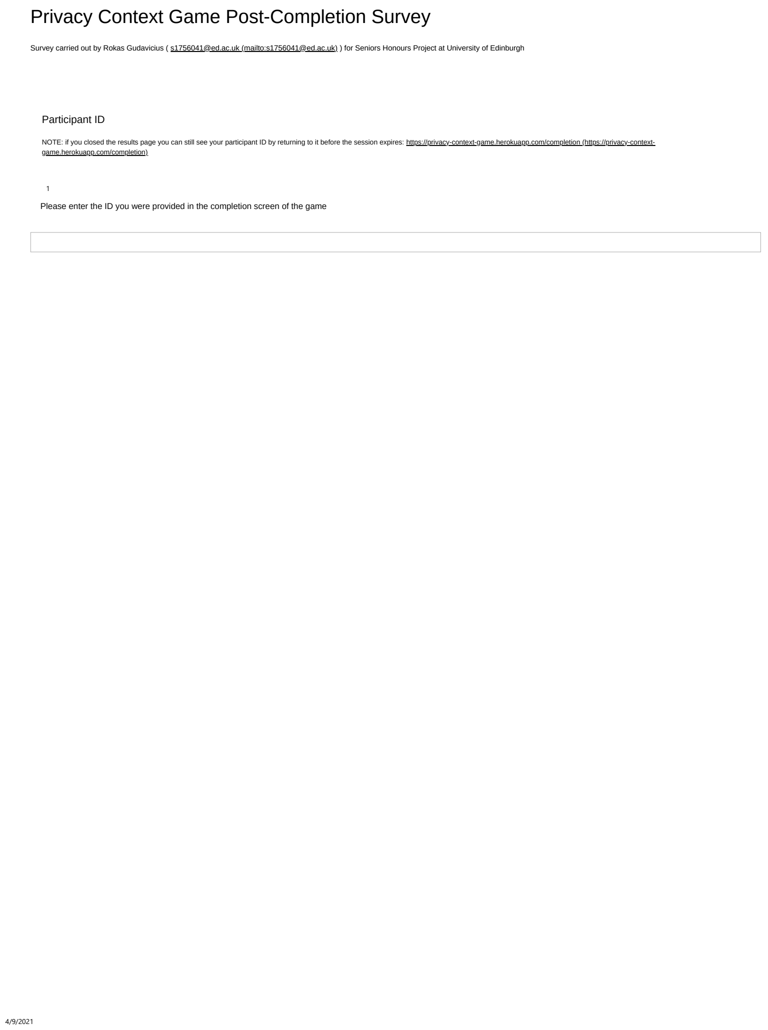
Page **4** of **4**

- I have read and understood the above information.
- I understand that my participation is voluntary, and I can withdraw at any time.
- I consent to my anonymised data being used in academic publications and presentations.
- I allow my data to be used in future ethically approved research.

THE UNIVERSITY *of* EDINBURGH
informatics

Appendix D

Post-Game Questionnaire



Privacy Context Game Post-Completion Survey

Survey carried out by Rokas Gudavicius (s1756041@ed.ac.uk (mailto:s1756041@ed.ac.uk)) for Seniors Honours Project at University of Edinburgh

Participant ID

NOTE: if you closed the results page you can still see your participant ID by returning to it before the session expires: https://privacy-context-game.herokuapp.com/completion (https://privacy-context-game.herokuapp.com/completion)

1

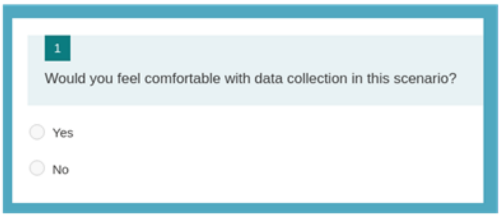
Please enter the ID you were provided in the completion screen of the game

4/9/2021

Yes/No Question Evaluation

NOTE: Please assume a scenario is provided at the top of the page along with the question for either option

A.



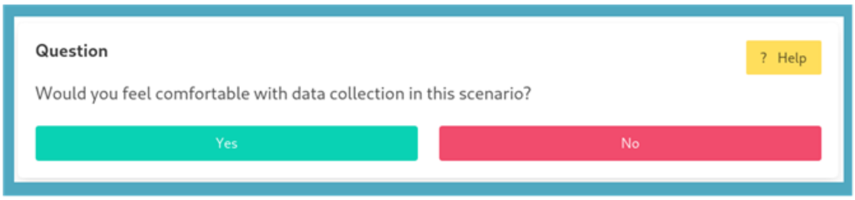
1

Would you feel comfortable with data collection in this scenario?

○ Yes

○ No

B.



Question

Would you feel comfortable with data collection in this scenario?

2

Which form of a Yes/No question do you prefer?

○ A

○ B

○ No Preference

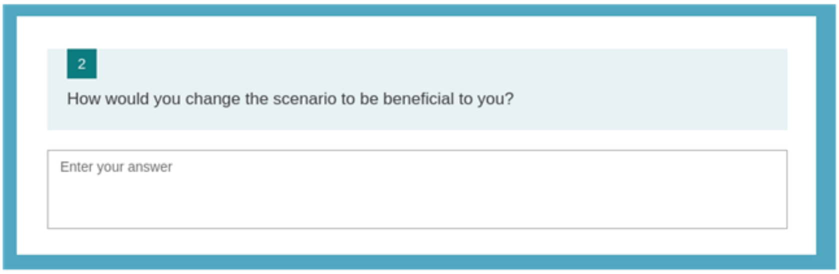
3

(Optional) Can you explain your choice?

Change Scenario Question Evaluation

NOTE: Please assume a scenario is provided at the top of the page along with the question for either option

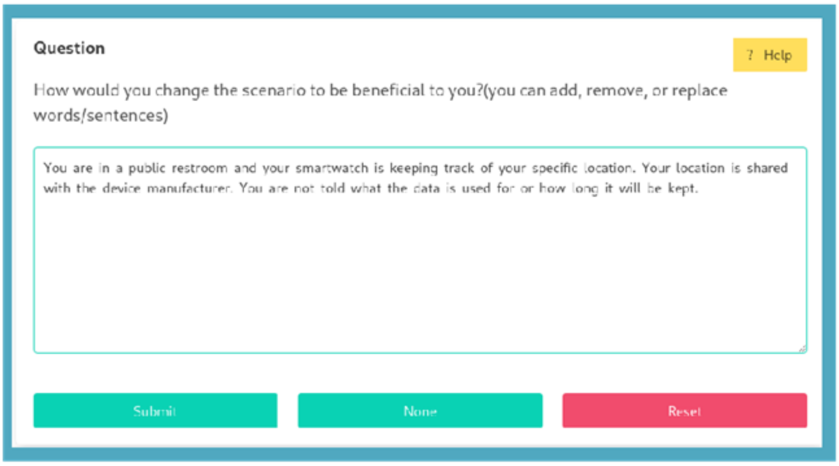
A.



2

How would you change the scenario to be beneficial to you?

B.



Question

How would you change the scenario to be beneficial to you?(you can add, remove, or replace words/sentences)

You are in a public restroom and your smartwatch is keeping track of your specific location. Your location is shared with the device manufacturer. You are not told what the data is used for or how long it will be kept.

4

Which form of a Scenario Change question do you prefer?

○ A

○ B

○ No Preference

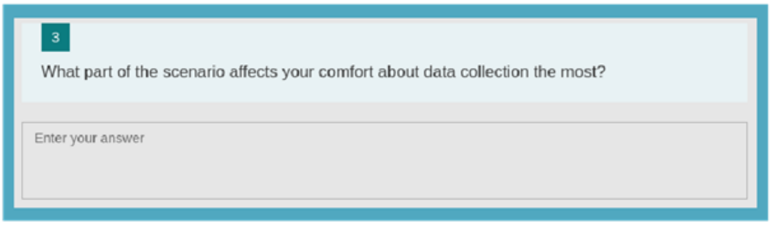
5

(Optional) Can you explain your choice?

Highlight Question Evaluation

NOTE: Please assume a scenario is provided at the top of the page along with the question for either option

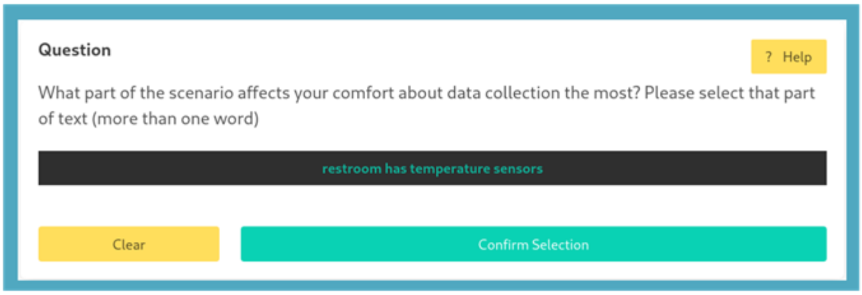
A.



3

What part of the scenario affects your comfort about data collection the most?

B.



Question

What part of the scenario affects your comfort about data collection the most? Please select that part of text (more than one word)

restroom has temperature sensors

6

Would you prefer to answer the question by writing or by highlighting?

○ Writing (A)

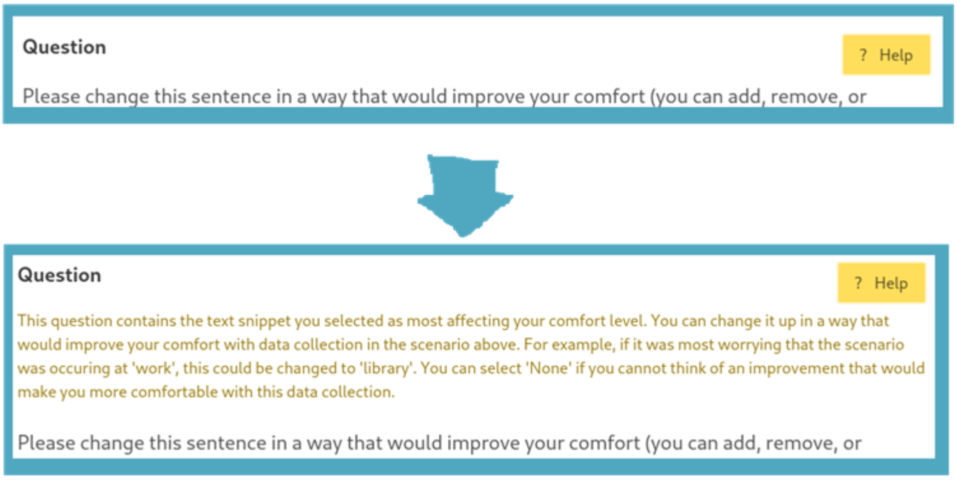
○ Highlighting (B)

○ No Preference

7

(Optional) Can you explain your choice?

Help Feature



Question

Please change this sentence in a way that would improve your comfort (you can add, remove, or

Question

This question contains the text snippet you selected as most affecting your comfort level. You can change it up in a way that would improve your comfort with data collection in the scenario above. For example, if it was most worrying that the scenario was occuring at 'work', this could be changed to 'library'. You can select 'None' if you cannot think of an improvement that would make you more comfortable with this data collection.

Please change this sentence in a way that would improve your comfort (you can add, remove, or

8

Have you used the Help button?

○ Yes

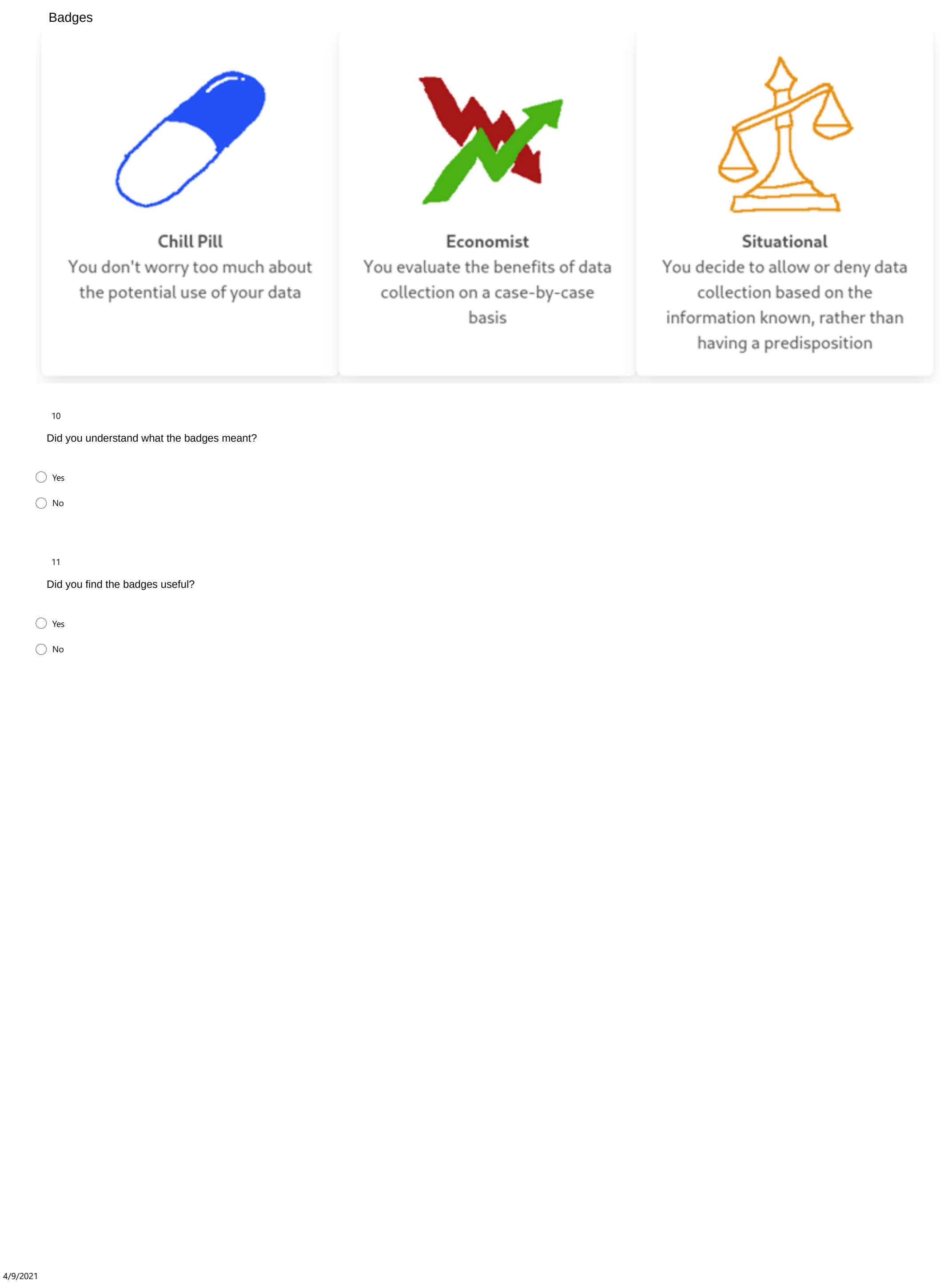
○ No

9

(if you answered "Yes" to previous question) Did you find the Help text useful?

○ Yes

○ No



Badges

Chill Pill
You don't worry too much about the potential use of your data

Economist
You evaluate the benefits of data collection on a case-by-case basis

Situational
You decide to allow or deny data collection based on the information known, rather than having a predisposition

10

Did you understand what the badges meant?

○ Yes

○ No

11

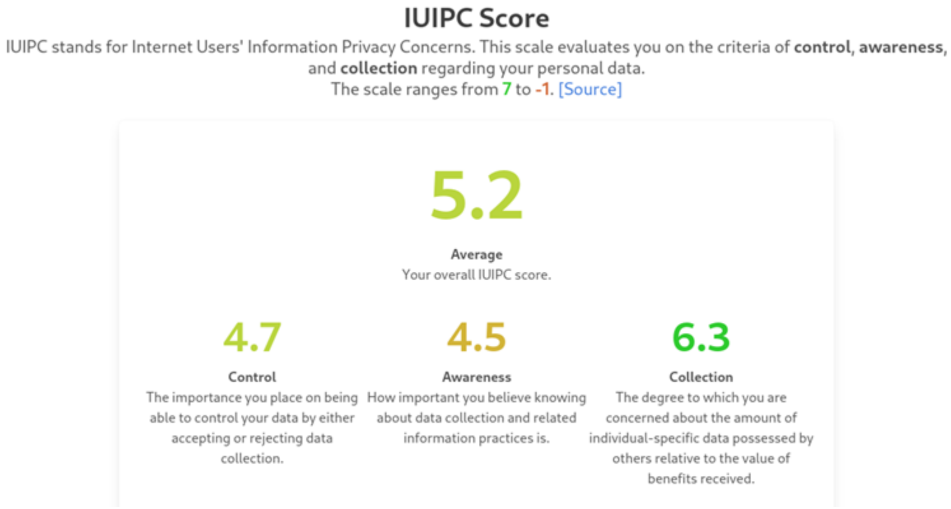
Did you find the badges useful?

○ Yes

○ No

4/9/2021

Privacy Scale



IUIPC Score

IUIPC stands for Internet Users' Information Privacy Concerns. This scale evaluates you on the criteria of **control**, **awareness**, and **collection** regarding your personal data.
The scale ranges from 7 to -1. [Source]

5.2

Average
Your overall IUIPC score.

4.7

Control
The importance you place on being able to control your data by either accepting or rejecting data collection.

4.5

Awareness
How important you believe knowing about data collection and related information practices is.

6.3

Collection
The degree to which you are concerned about the amount of individual-specific data possessed by others relative to the value of benefits received.

12

Did you understand what the Privacy Scale score meant?

○ Yes

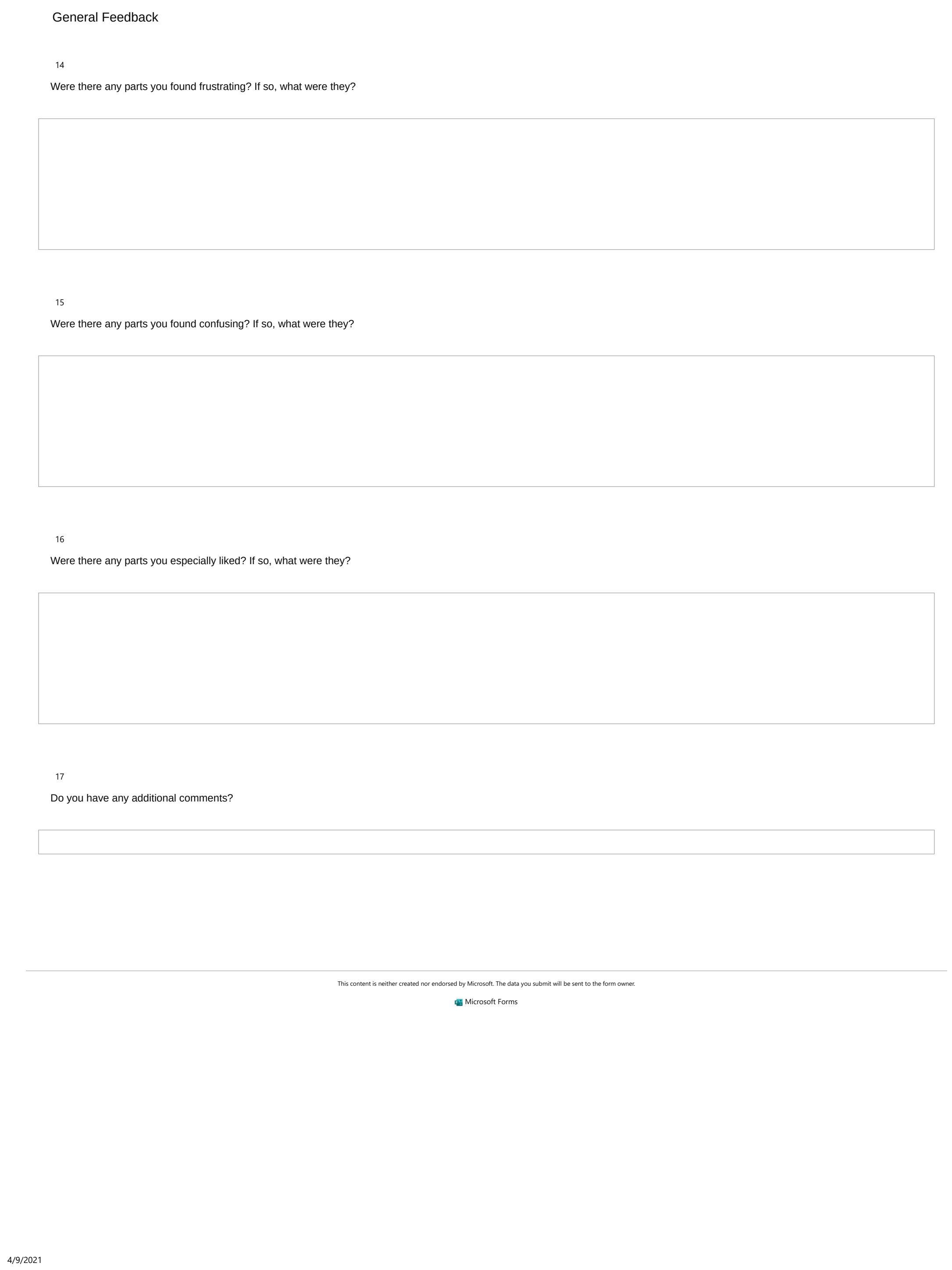
○ No

13

Did you find the Privacy Scale score useful?

○ Yes

○ No



General Feedback

14

Were there any parts you found frustrating? If so, what were they?

15

Were there any parts you found confusing? If so, what were they?

16

Were there any parts you especially liked? If so, what were they?

17

Do you have any additional comments?

This content is neither created nor endorsed by Microsoft. The data you submit will be sent to the form owner.

Microsoft Forms

4/9/2021