Using Interactive Games as a Dark Pattern Teaching Tool

Callum Leask

4th Year Project Report
Artificial Intelligence
School of Informatics
University of Edinburgh

2024

Abstract

This paper researches whether an educational game is an effective tool for teaching users to recognise dark patterns in online shopping websites. It first discusses the current state of dark pattern researches and explains the concept behind dark patterns and online privacy. It then uses a multi-stage quiz game to gauge a baseline for how users perform with no training, gives them a number of training examples and explains the dark patterns they might encounter, and finally quizzes them on a number of classification questions. The pilot study finds that users consistently improve by a considerable amount, with an average improvement in their correct answers of 27.76%. It also discusses places in this style of educational game which may be affected by bias, and finds after analysis that user performance does degrade if there is a long delay between the training and testing as posited by the recency bias, but that users still consistently improve over their baselines score. Overall we present a very promising option for teaching these topics in an effective and interesting way, which would benefit from further study with a larger number of participants.

Research Ethics Approval

This project obtained approval from the Informatics Research Ethics committee.
Ethics application number: [rt #8087]
Date when approval was obtained: 2024-02-18

The participants' information sheet and a consent form are included in the appendix.

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

(*Callum Leask*)

Acknowledgements

I owe a massive thank you to my supervisor for this project Nadin for helping to keep me motivated and for advising how to tackle a project of this size, without the direction given I would certainly have gotten lost along the way!

And thanks to my family and fiancée for putting up with me rambling about manipulation and patterns at any given opportunity, you have the patience of saints.

Table of Contents

Chapter 1

Introduction

With each passing year more and more of our lives move online. From social networks to shopping websites, people give their time, money and information to these sites trusting that they are being completely transparent, but what if that's not the case?

The term "Dark Patterns" was coined in 2010 by Harry Brignull to refer to the manipulation techniques and psychological tricks that online designers were taking advantage of to take advantage of their users. These include simple tricks like pre-selected or highlighted options that users will gravitate towards without thinking, and more involved methods like fake "limited time" countdowns on sales and baskets which force you to check out in a certain time or your saved items are reset.

This paper aims to develop an interactive educational game which can effectively teach users of varying ability to recognise dark patterns where they can appear in online shopping websites, and also to classify the various types to understand the differences between them and how they commonly present themselves.

The report is broken up into 4 sections. First in chapter 2 we review the current state of online privacy and the context surrounding dark patterns, educational games and other relevant research. In chapter 3 we catalogue the methodologies used during the development of the game itself, the dataset used, and the testing regime for the pilot study. We then present the results and discuss them in more detail in chapter 4, alongside analysing some of the biases which could impact the outcome. Lastly in chapter 5 we conclude with the overall results of the study and how they impact the current state of dark pattern research.

By the end of this paper we aim to have analysed the effectiveness of our prototype educational game and to have discerned whether this approach to teaching users about dark patterns is successful and if it warrants further research.

Chapter 2

Background Research

In this chapter, section 2.1 discusses the issue of online privacy in the modern world, and how internet users feel about the way data is treated. In contrast, section 2.2 will introduce the dissonance between the way people consider their online privacy, versus how they act online - also known as the privacy paradox. Section 2.3 will discuss the prevalence of Dark Patterns and how they are used to affect user judgement. In addition section 2.4 will consider the difficulty faced when attempting to quantify privacy risks. Lastly, section 2.5 will examine the use of interactive video games in a learning context, and consider the drawbacks and benefits they offer.

2.1 Online Privacy and Attitudes Towards It

Internet usage over the last 20 years has grown massively, with the number of internet users worldwide placed at just over 1 billion in 2005, rising to well over 5 billion in 2022 [1]. As this use grows, so do the markets surrounding it, especially the market for user's data. The Customer Data Platform market size was valued at 4.67 billion dollars in 2022, and forecasted to reach nearly 50 billion by 2030 [2]. Some have dubbed the rise of this market as proof of 'surveillance capitalism', denoting that surveilling individuals and collecting information on their activities has become a commodity [3] . In this research paper Zuboff argues that people are misdirected to think about data collection as a technological requirement, where in reality data collection is a commercial requirement.

This increase in internet usage has brought with it privacy concerns from many of the people using it most. A 2019 study by the Pew Research Center [4] found that more than 8 out of 10 adults in the U.S. feel that they have little to no control over who is able to see their online search data. This is alongside a similar 81% who disagreed when asked whether the benefits of data collection outweighed the risks.

A report by Cisco from 2021 [5] shows that of the 2600 adults they surveyed, 86% responded that they care about their data privacy, and a subsequent 79% were willing to act to protect their data. By the end however, only 32% were considered to be "Privacy Actives", meaning that they cared about their data but had also actioned to protect it.

A small (and rather entertaining, if not worrying) survey carried out by ProPrivacy.com [6] offered 100 people a dollar if they completed a survey (which consisted only of accepting an agreement). Only 19 people even went as far as the terms and conditions page, and just one participant stopped to read the terms, which included giving away the rights to your firstborn, among other rather egregious agreements. 70% of the participants claimed after the fact to have read the terms and agreements before accepting.

This is a repeating theme throughout the field of online privacy, when asked directly there are few people who would say that they don't care about their online privacy, but they will rarely do anything proactive about protecting it.

2.2 The Privacy Paradox

The results of most privacy-concern related surveys seem to point clearly to the majority of internet users being both aware of their online privacy, and actively engaged in protecting it. You might be surprised then to learn that despite people's concern about their data, they are seemingly willing to sacrifice it at any given point for convenience. This has been dubbed the "Privacy Paradox". [7]

This behaviour has been observed in multiple studies i.e. [8]. For example, a 2019 study by the USA's National Telecommunication and Information Administration found that 73% of internet-using households had major concerns relating to Online Privacy [9]. In that same year however, a survey from Kaspersky found that 93% of people were sharing information digitally [10], and a study from Data and Marketing Association in the UK found that 55% of consumers surveyed would say that they were happy with the amount of personal information they give to organisations [11].

This behaviour is potentially best shown through the lens of social media. Specifically, two good examples are Facebook and more recently, TikTok. Meta reported in 2023 that Facebook now has over 3.03 billion users [12], meaning that over 37% of the world's population use Facebook in some manner. However, Facebook has had a number of large-scale data breaches, including uploading people's email contacts without their permission [13], but while people have continually voiced concerns over this repeated mishandling, the number of Facebook users continues to grow year on year.

TikTok has had less time in the spotlight by comparison, with a meteoric rise in popularity over the pandemic, with users rising from 54 million in 2018, to 1.8 billion in 2023 [14]. Despite the app's relative age, it has had no shortage of data privacy incidents, from being fined by the UK government for collecting children's info without parent's consent, to news that Chinese TikTok employees could access American user's data despite the company claiming otherwise [15].

These trends also aren't tied to any particular method of accessing the internet. In another survey into people's perceptions of privacy in regards to mobile app use [16] participants were asked questions on their thoughts on data privacy and collection, before later being asked to download specified apps from the app store. Broadly, people were all spoke of their interest in privacy, but many went on to ignore the EULA entirely

when downloading the apps.

A paper by Choi et al. [17] draws a connection between so-called ”Privacy Fatigue” and people’s willingness to give up information online. They define fatigue as being a combination of feelings of exhaustion from constantly having to consider their privacy choices, and futility or cynicism due to the constant news of data breaches and data mishandling from companies and bodies that claim they can be trusted to handle data. Their data analysis shows a purported link wherein people who claim to be security conscious generally are, but the more that fatigue is taken into consideration the more likely they are to let slide their privacy standards.

Another interpretation is an article by C. Hallam and G.Zanella [18], which proposes that the privacy paradox can be explained as a risk vs reward exercise undertaken by the user. They describe a cognitive research theory known as ”Construal Level Theory” that classifies behaviours into abstract or concrete, representing distant-future intentions and near-future intentions respectively. They posit that the social benefits of online interactions are near-future and quickly realised whereas the risks of a data breach or privacy invasion are distant-future possibilities and people are biased towards taking the short term gratification.

2.3 User Interfaces and Dark Patterns

People’s willingness to sacrifice their information online for the sake of convenience has not gone unnoticed by those looking to design websites and services that gain access to as much user data as possible. These tactical design choices aimed at deceiving users has been coined ”Dark Patterns”, by Dr Harry Brignull in 2010 [19]. He determined a number of patterns used online to trick or convince people into parting with information they wouldn’t normally, or to get them to sign up for services they didn’t necessarily need [20]. Some of these include practices such as Preselection (having a default option already selected), Fake Urgency (pretending there is a time limit on a product/service) and ”Confirmshaming” (describing some options in a manner that paints them in a negative light). [21].

Much of this study is inspired by the research of [22] which aimed to study how prevalent dark patterns were becoming on online shopping websites. They used a webscraper across over 11,000 of the most popular retail sites, finding evidence of dark patterns on 11.1% of them. Given how much traffic these sites receive this is a massive number of websites employing dark patterns to manipulate their users. While this is revealing, the study doesn’t go further than presenting how often these appear. This leaves a gap when it comes to discussing what to do about this uncomfortable trend in online shopping spaces.

So far the trend has become the development of tools which inexperienced users can use to recognise dark patterns for them to act as a warning when one or more are present. A 2023 paper by Ryan Matthew Wood [23] develops a chrome extension which detects specific dark pattern types and informs the user. They then performed a study where users interact with a number of sites while using the extension and provide feedback as to how having this tool affected their experience, with most users

reporting positively about their experience. While this does show the potential for an early warning system like the one they designed, it doesn't address the underlying issue that the users themselves for the large part don't understand dark patterns, and can't recognise types that the system doesn't pick up. If we were to give users the necessary skills to pick up on these themselves and pair it with a system that can recognise patterns then individuals would have a much better chance of being able to pick up on all the dark patterns that they encounter.

But why is this responsibility to learn falling on users, what is the law doing to combat these dark patterns? GDPR laws which have been in place in Europe since 2018 introduced rules that require companies to gain explicit consent from users when collecting their data, and also requirements for the user's ability to easily accept and reject cookies (without pre-ticked default selections). Since these were introduced, a number of companies began to offer templates for GDPR consent pop-up windows for use on websites and apps. A study taken in 2020 [24] found that very few of these templates that were in wide use actually followed the requirements that had been laid out. As many as 32.5% of the websites scraped used implicit consent (implying that by using the service in a number of ways, the user accepts the cookies - without explicitly stating it - clearly violating the GDPR guidelines). They also found that only 12.6% of the sites had an option to reject all cookies at once that was as accessible as the accept all option.

These patterns aren't only found on smaller websites, with research conducted by Merchant Machine found that out of 72 online retailers they visited, Amazon was the site which featured the most dark patterns with 11 different tactics counted [25]. The company was also contacted by the European Commission in 2022 after the European Consumer Organisation and other national consumer bodies complained that Amazon were deliberately utilising dark patterns to make unsubscribing from their Prime subscription more difficult than signing up, after which Amazon committed to improve it's cancellation policy. Despite this statement, dark patterns remain a common feature of the site and many others like it.

2.4 Problems with Quantifying Online Privacy

As research into online privacy continues, one of the consistent issues that arises is measuring and quantifying privacy and people's attitudes towards it. Many of the different studies and research papers that have so far looked into topics such as the privacy paradox have created their own scales or measurement methods, which are often difficult to compare to each other directly and can cause confusion.

One of the earlier attempts to create a standardised measuring tool for privacy concerns was the Concern for Information Privacy instrument. This was a 15 question survey, in which each question fit into one of four categories, either Collection, Errors, Unauthorized Secondary Use or Improper Access. Each question was answered on a scale of 1 through 7, with 1 being "Strongly Disagree" and 7 being "Strongly Agree". [26] This scale is thorough in the context it was designed for, namely for managers and companies researching what concerned consumers about their information collection and privacy

practices. It struggles more however in being generalized to an online context, largely due to it's age and domain.

A more modern attempt at a privacy scale is the Internet User's Information Privacy Concerns model. As the name implies, it was designed to be used in the analysis of consumer's concerns in specifically online information privacy. It attempts to extend the existing CFIP scale by adding new dimensions, namely control and awareness. Control refers to the amount of control a user has over their data once it has been provided (i.e. the ability to modify data, the ability to exit, etc) which tends to directly correlate with their trust. Awareness refers to a users knowledge of how their data will be used and how fair they believe the relationship is. These additions aide in making the scale easier to generalize, and more applicable to use when studying online privacy concerns.

While the IUIPC model is becoming the most used model when discussing online privacy, it would benefit from continued research to ensure it remains applicable to the current online world, and to help to improve it's overall accuracy.

2.5 Video Games as a Teaching Tool

Play based learning has been employed for far longer than most realise, with games such as Chess and Xiangqi being used historically to teach strategy [27], the latter of which has been played at least since the 1st century AD [28]. In rather more modern times, games have been taking a larger role in the classroom and have been becoming a promoted learning tool. The pandemic also caused a large uptake in interactive games such as Kahoot during the period that schooling was largely done online.

Research into games as a teaching tool have proved that by engaging children (or indeed adults) and getting them to interact with their learning more directly can help to speed up learning, and make topics easier to digest. [27]

Some large gaming IP's have begun to release editions targeted specifically at the teaching market, a good example being Minecraft Education Edition, which launched in 2016. This too saw a large uptick in use during the pandemic, with 35 million registered users, and at the same time Mojang released a collection of online-learning themed packs for the base edition of Minecraft (the Minecraft Education Collection) which received more than 63 million downloads in it's first 6 months of release [29]. The game aims to help children work on a variety of topics, from exploring topics like history, computing and coding, and social skills [30].

Many of the games being developed as educational tools are designed to be as broad as possible to make them applicable to many different subjects (such as the minecraft and kahoot examples above). While these platforms offer a novel way to make different subjects more interactive, far fewer choose to make a game which is dedicated to one topic exclusively. This is what we chose to expand on for this project, to judge the viability of making a game that can explore one single theme in more depth.

There are a few hurdles to be avoided with game based learning as a whole, including the technological learning curve they can require (both for the children and potentially the teachers/classroom assistants), and also the time spent interacting with a screen as

for many schools limiting screen time is slowly becoming a priority. Games also have to be aligned with the learning goals in the classroom, and as such need to be either designed or tailored to fit individual settings.

Chapter 3

Methodologies

This section will discuss in section 3.1 the game project and it's development process, in section 3.2 the methodologies employed during the acquisition of the dataset and the accompanying difficulties, and lastly in section3.3 how the final user study was carried out.

3.1 Game Development

3.1.1 Game Engine Choice

The Godot engine [31] was chosen as the best candidate for the development of the training game as it offered a few distinct advantages, including being completely free to use, open-source, having a large number of resources available to speed up learning, and making exporting the final game project to different operating systems and platforms easy.

3.1.2 Game Type

Many different types of game were considered initially, including making a top-down Role Playing Game, or a 3D third-person platformer. All of these came with their own problems, either in the amount of time/pre-requisite knowledge required to make and implement assets and gameplay, or in the ease of adding in an educational element. These are also games that require a lot of user input and which require time to learn how to play, making the game inaccessible to those less experienced with games in general. Someone who has never interacted with a 3D game might have to spend much longer learning how to control their character and interact with the world as opposed to taking in the content being shown to them. In a regular game this isn't much of a problem, people are generally happy to spend 5 - 10 minutes learning the controls when it leads into hours of gameplay, however in the context of a short educational game the ratio of tutorial to game could put players off playing at all, or make it feel like a slog that they have to commit too much time to. Instead, the final game type was chosen to be a quiz style game. This would allow for a very quick to learn intuitive game that can be accessed by people of different experience without requiring any time spent on

learning how to play, and also for educational content to be implemented easily without seeming out of place like it can other genres.

3.1.3 Gameplay Structure

The game is split into three distinct sections which are completed in a linear fashion. The player first answers 10 questions to establish a baseline for how well they perform before training. These first questions are True/False only as we assume that the player won't have the knowledge required to identify any dark pattern types. They are then asked to complete the training, which consists of a brief explanation of each of the four dark patterns and three visual examples with red highlighting of the main features. Finally the player classifies more examples, with the added difficulty of being required to select a dark pattern type each time they indicate a positive example. The length of this final quiz varies slightly each playthrough, described in more detail in 3.1.5

3.1.4 Visual Design

To ensure that the "game" aspect wasn't sacrificed for a generic quiz, efforts were made during the design process to give the project an interesting visual style. Rather than presenting information straight to the screen, the player interacts with a 2D world containing a desk and virtual computer monitor which gives them the necessary prompt. The images are displayed next to the monitor by a "projector", and the player can use the arrow keys to look left and right in this world. The aim of having controls to change the view inside the world was to make it more immersive and interesting for the player, and to create space to add more objects. Unfortunately due to time constraints the planned feature of interactable objects around the room had to be scrapped, but the movement was left in.

Figure 3.1: The virtual monitor for player inputs

3.1.5 Gameplay Features

A key aspect when designing the game was the intention for it to be replayable, meaning that it shouldn't just repeat the same questions at the player every run. To achieve this the game is built to randomly choose either a positive or negative example for each question, and if a positive example is chosen, to choose a random unseen image. To do this it keeps track of the number of images shown for each category of dark pattern, and selects a random pattern type that still has unseen images remaining. Once all positive examples have been shown, the game terminates.

To ensure that there was no possibility of many positive or negative examples back to back, the question system also has a "fairness counter" which keeps track of how many times in a row the same type has been chosen, forcing it to flip to the opposite type after a given counter threshold is reached.

At the end of game once all questions are completed, the player is shown a feedback screen with a breakdown of their performance in the final quiz, alongside how much of an improvement this represents over their initial baseline score. They are also shown their accuracy in identifying dark pattern types, which is the percentage of correct classifications they made when selecting a positive example.

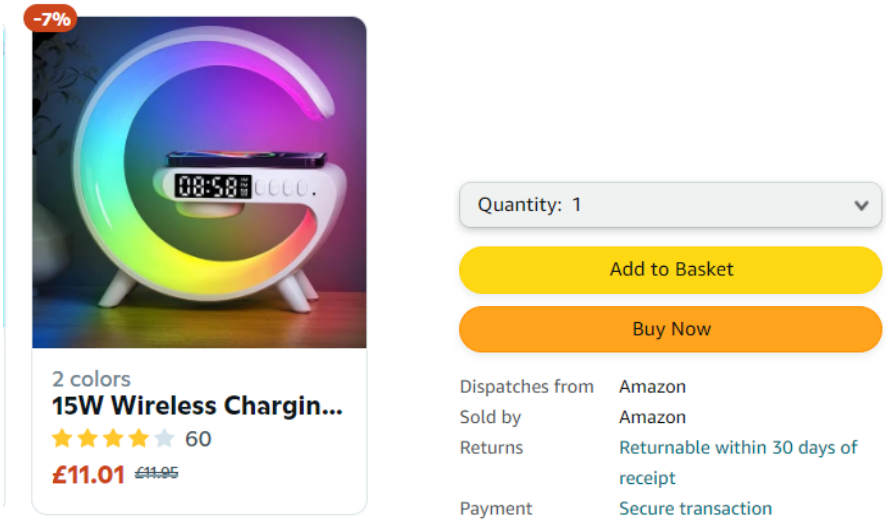
3.2 Dataset Gathering

3.2.1 Question Data

One of the key challenges encountered during the process of designing the game was acquiring the data to use for the question creation. The 2019 study by Mathur et al. [22] was one of the largest and most complete analyses of dark pattern prevalence online as of the writing of this paper. It observes roughly 11,000 of the most popular shopping websites using a webcrawler, which saved snippets of dark pattern text and the URLs where it was found. The paper itself made a specific note to having taken a screenshot of the pages and how they appeared when the patterns were detected, however we found out only later that this dataset of screenshots was too large to be included in the Github repository and so was never published. Attempts to contact the original paper's authors initially were unsuccessful, so it was decided to limit the number of dark patterns covered in the game and to find examples that either still existed on the websites who's urls were collected, or which were found manually on other shopping websites.

An effort had to be made to manually collect a dataset of screenshots which displayed non-dark pattern shopping websites. This also proved to be more difficult, as each screenshot had to vary in scale (i.e. just a checkout button, an item description, or larger sections of the page) to ensure that they were as varied as the dark pattern examples and weren't identifiable without closer analysis by players. These screenshots also had to be closely examined before being included, as it was very easy to unintentionally include stock counts or small pop-ups that could potentially be considered dark patterns, and could confuse players. When selecting non-pattern images we aimed to have screenshots that contained enough small text and other information that players have to

take their time and analyse each, aiming to not make them too easy, or alternatively too ambiguous.



(a) (b)

Figure 3.2: Examples of Non-Dark Pattern images

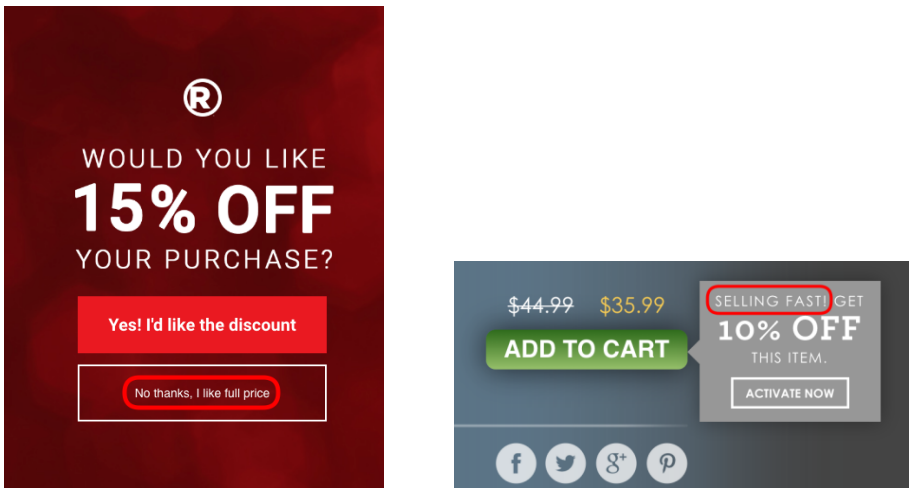
In the end the game's scope was limited to only test players on the 4 most common types of dark pattern found during Mathur's study of shopping websites, as some of those patterns identified in the paper were only present on a very small subset of the pages scraped (i.e. forced action, which appeared only 6 times in the 11k websites). This decision served to keep the game a simple proof of concept with room for expansion, and also to minimise the amount of work required in finding examples. This does also limit the scope to specifically shopping websites and the dark patterns that commonly appear on them, but they serve as effective examples because of the prevalence and the visibility of the patterns. There is also no reason that the scope couldn't be expanded in future given the availability of alternative datasets.

The homepage for the study [1] features example images where the given dark pattern is highlighted with a red circle. These were used as the examples during the training portion of the game, giving the player 3 examples for each pattern, with each being a different type of that category.

In the end we had 12 example pattern images, 12 dark pattern images and 20 non-pattern images. This amount was enough to ensure a number of variations on each pattern is shown during the final test, and that there are enough non-patterns for there to be a good mix for the random selection.

Due to the manual nature of creating the dataset this also marked a convenient point to stop, as some patterns such as misdirection were proving difficult to find examples of. This was mostly because of how old the initial study's list of URLs was, most of the pages had long updated and no longer contained the examples listed. This meant testing examples had to be found by manually searching websites.

[1] https://webtransparency.cs.princeton.edu/dark-patterns/



(a) Misdirection Example (b) Scarcity Example

Figure 3.3: Examples of the training images used

3.3 User Testing

3.3.1 Testing Methodology

The intention when beginning this study was for every participant to play the game in-person, on the same machine, which was supposed to simplify the process and allow for more people to take part. This quickly proved to be folly, and a massive limiter both to how many people could take part and the variety in participant's computing experience. Most of the people who were able to take part in-person were school of informatics students who have thousands of hours working online and on computing projects, whereas this project is aimed to be widely applicable and to give people without much technical knowledge the information needed to be able to identify dark patterns. In lieu of this, the testing methodology pivoted to instead make use of the Godot engine's ability to easily export games into executable files. A folder for the project was created on a Google Drive, containing only the two files required to run the game, the executable and the data pack. This meant that participants could be supplied with a relatively simple set of instructions for downloading and setting up the game, and it could be run on any modern Windows machine. On completion of the game, the participant is shown their results, and at the same time a full breakdown is saved into a separate results file. The instructions indicated to candidates how to send their results back to us, and through this method we were able to acquire much more varied data, even in this limited study.

Chapter 4

Results and Discussion

This chapter breaks down the numerical results in section 4.1 before discussing them in more detail in section 4.2. Finally in section 4.3 we delve into some of the potential biases which may be at play in the results and use player tests to analyse their impact.

4.1 Results

The final study was conducted as described in section 3.3.1. Participants were sent a link to download the game files to play on their local machine, alongside a set of instructions for how to launch the game, what to expect, and how to return the output.

Once playing, players first completed a short introductory section, asking them to self-report two values, how confident they believe they are using computers and information technology, and how attentive they think they are while shopping online. These will later be referred to as the users' confidence and attentiveness values.

The player then completes an initial set of 10 questions which simply consist of stating whether they agree or disagree with the statement "I believe this website is attempting to manipulate me" while being shown an image of a section or snippet from a shopping site. This baseline is used to establish how they perform without any training. It forgoes asking the user to classify the types of patterns when they agree as we assume that they have no knowledge on what the different types would be, and the data would largely be noise. The percentage scored in this section is saved as the initial score.

The next section consists of a set of training examples. The user is shown a short block of text explaining what the dark pattern type aims to do, and common features to look out for while classifying. They are shown alongside this three images for each pattern type, with the common feature circled in red. These training examples are not used again anywhere in the baseline/testing.

Finally the user completes a quiz consisting of the 12 dark pattern examples which are mixed with a randomly sized selection of examples with no dark patterns present. The random selection is limited so that no more of two of one type will be shown back to back, but the player is unaware of this. On top of selecting agree/disagree to the same

statement as the baseline quiz, when a player indicates an example is a dark pattern they must then select which of the 4 types it is. From this section we save the percentage that the user gets correct when selecting agree/disagree, alongside an indicator of how many of each pattern the user classifies correctly.

The initial study consisted of 6 users who completed the full quiz. Table 4.1 contains the numerical results, with each row corresponding to one tester's results after playing the game.

Confidence	Attentiveness	Init. Score (%)	Final Score (%)	Improvement (%)	Accuracy (%)
5	5	60	92.59	32.59	66.67
5	4	80	91.3	11.3	83.33
4	4	40	81.25	41.25	66.67
4	5	70	85	15	58.33
3	4	50	83.33	33.33	66.67
1	3	40	73.08	33.08	50

Table 4.1: Study Results

From this initial testing there is a significant positive result, with every participant improving between the baseline testing and the final quiz. The mean overall improvement of 27.76% is a strong indication that users have begun to recognise dark patterns that they didn't prior to training. There is a considerable amount of variation in the amount improved by different users however, with a standard deviation across all improvement scores of 11.81%.

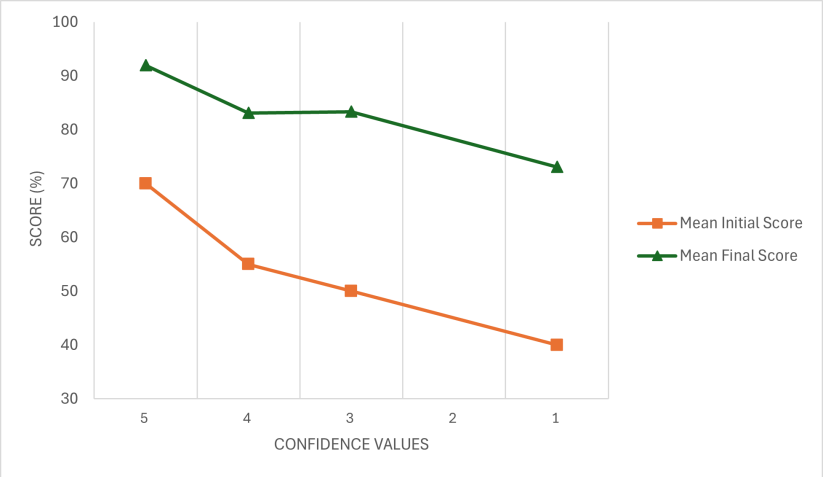


Figure 4.1: Initial and Final Scores vs Reported Confidence Values

A more surprising result is a clear trend between the self-reported confidence value, which indicates how confident users believe they are with computer technology, and the initial and final scores (shown in Figure 4.1) When taking the mean value of each confidence level, testers that claimed to be more confident with using computers in general scored higher in both the initial and final quiz questions.

Each tester was shown a total of 12 patterns, 3 from each category. Figure 4.2 shows how many of each pattern were identified as a dark pattern, and categorised correctly.

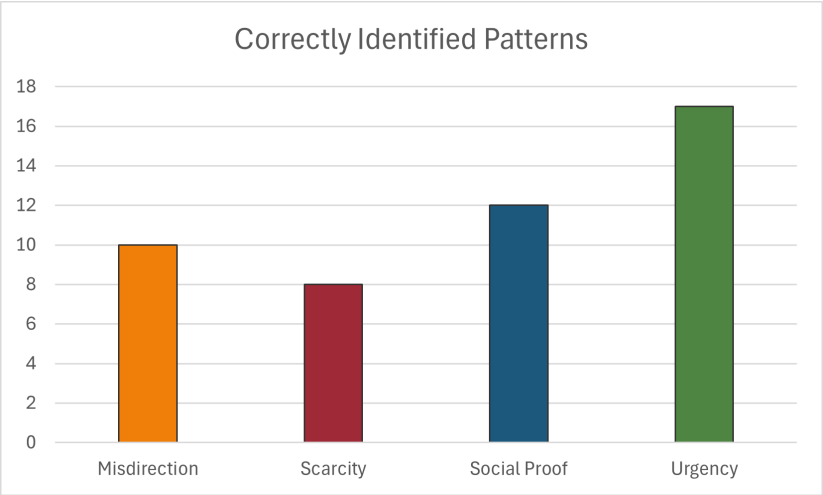


Figure 4.2: Number of Correctly Identified Patterns

Urgency was almost universally correctly identified with 17/18 examples, social proof had 12, misdirection had 10, and scarcity was the least correctly recognised, with only 8/18 examples recognised and classified.

Figure 4.3 shows the market share of each pattern, i.e. it's percentage of the correctly identified types. We can more clearly see here that users were better at identifying the urgency patterns than they were the other types.

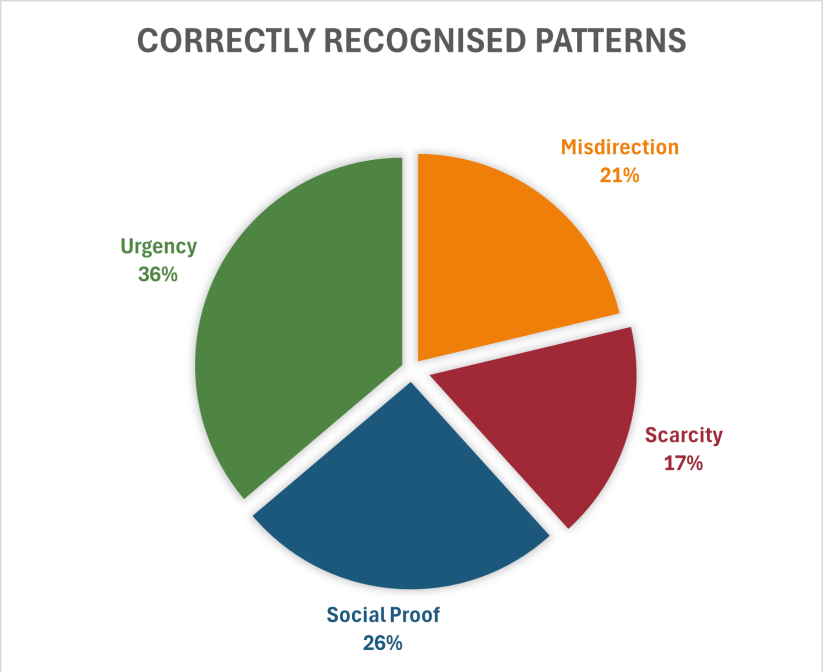


Figure 4.3: Percentage of Correct Answers Attributed to Each Pattern

4.2 Discussion

These results show us a substantial improvement in players ability to recognise dark patterns from their initial baseline to their post-training results. Across all self-reported confidences, players had an average initial score of 56.67%, representing only a slightly better than 50/50 chance to recognise a dark pattern when interacting with a website. In stark contrast, the average score for the final quiz was 84.43% (to 2 decimal places).

The accuracy of dark pattern classification seems to indicate that players have more difficulty identifying some patterns than others. From the initial study it appears that Scarcity was either difficult for players to spot, or was regularly being misclassified. This is discussed further in section 4.3.

The improvement between the initial score and final score seems to correlate to how technically confident players believe that they are. Players with a high confidence value have an average improvement of 21.945%, players with a confidence of 4 had an average of 28.125%, and players at 3 and 1 had 33.33% and 33.08% respectively. This implies that the game is working as intended, and those who are less confident using computers and working online are able to improve and close the knowledge gap with those who are more familiar with the technology. To be confident in this result however it would be useful to acquire more test results to see if this trend holds.

After completion, users were asked to give a brief piece of verbal or written feedback that discussed overall how they found the game, whether they found it interesting or entertaining, and whether they found that there was any features that would make it more usable for them. This feedback was useful to indicate where there could be potential future improvements, for example a large portion of feedback (66%) contained mention of including a back button during the quiz section which could allow a player to return and edit their answers if they realised they made a mistake or mis-clicked an answer. 50% of the comments also mentioned that they believed the left and right movement discussed in section 3.1.5 was more confusing for them than it was immersive, despite it being mentioned in the brief there were some who forgot when they were playing that this was an option. They state that it would be an improvement to simply lock the screen in one place and have everything clearly visible to ensure that nothing can be missed.

Almost all players (5/6) commented afterward that they found what they learned to be interesting, if not surprising. They reported that they had seen patterns such as the countdown timers from the urgency pattern before while browsing, but had always taken them at face value and assumed that they would have to be quick about their purchase. Interestingly this same comment was made by players of all self-reported attentiveness values, meaning that even people who believed they paid a lot of attention while shopping online hadn't considered these before. This goes

4.3 Biases and Mitigations

While the results indicate that there is a large potential for this to be a viable method to teach people about dark patterns, there are a few potential biases which need to be

addressed.

The main bias considered during development was the recency bias. This is defined in [32] as being ”a cognitive bias in which those item’s, ideas, or arguments that came last are remembered more clearly than those that came first.”. This is relevant to the project in two ways, how well the users perform in the final quiz, and the order in which the training information is presented to the user before the quiz.

If the recency bias is present in the training-testing setup as a whole, it would cause players to perform well when tested immediately after being trained, but to perform worse and struggle to retain the information if a length of time is present between the training and the testing.

Ideally all users would complete the training and first round of testing before having an arbitrary wait and completing the testing phase again to provide results to compare to. This was largely infeasible during this initial study due to both the time limit, and the difficulty in getting participants to commit to retesting. However, to gain an insight into whether this warrants further research we asked two testers to complete only the final quiz again, 9 days after they had initially completed the full game.

Initial Score	1st Attempt Score	Retest Score	1st Attempt Accuracy	Retest Accuracy
40	81.25	76.47	66.67	50
60	92.59	78.94	66.67	75

Table 4.2: Change in results after 9 day break

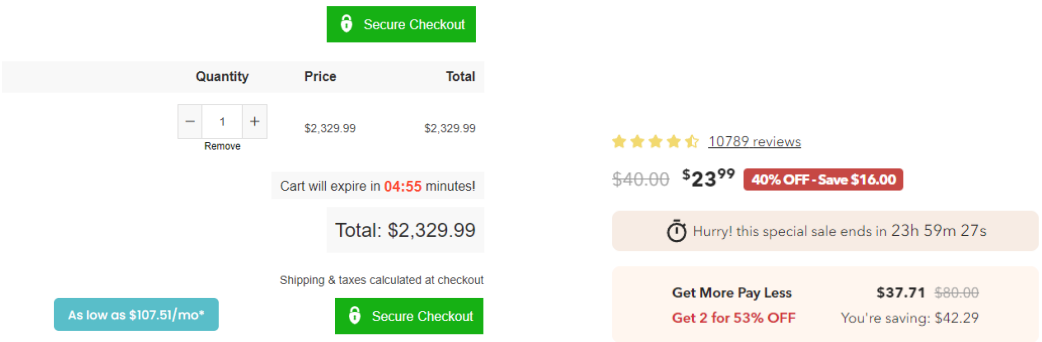
As shown in Table 4.2, there is a noticeable decrease in the performance score of both users after having spent time away from the testing. The change is more pronounced in the second user test with a drop of 13.65%, whereas the first only fell by 4.78%. Interestingly, while the first user’s accuracy decreased slightly (from 8 correctly classified to 6) the second user actually improved slightly, going from 8 to 9 classified correctly.

All of this goes to show that the recency bias of completing the quiz right after training does impact results, with the average of both tests being a drop of 9.215% after the 9 day gap. It is worth noting however that this far from negates all improvements made. There is quite a large variation between the two users tested, but on average they still improved 27.71% from their baseline to the 9-Day retest result. From this we can determine that a game like this can indeed be effective, but for proper application and long-term learning users would require some upkeep or reminders to keep the knowledge fresh. During this extra testing users were again given their score feedback at the end of the quiz, but it would also be worthwhile to (on the second attempt onwards) show users specifically which questions they got right and wrong to act as a reminder and to provide reinforcement learning.

The other area of the project with a risk of recency bias is the training section. The quiz was programmed so that users would see the training information in a set order, specifically Misdirection, Scarcity, Social Proof and Urgency.

As we saw in Figure 4.2, Urgency also happens to be the pattern which is most succesfully identified. The concern is that this was mostly due to the proximity of the

training examples to the quiz questions, but it's difficult to blame this entirely on the recency bias as the Urgency pattern also happens to have one of the most consistent visual appearances in the form of countdown timers. Some examples are shown in Figure 4.4. While the timers do vary in size and in location in the snippets, once they are spotted the classification is fairly simple.



(a) (b)

Figure 4.4: Urgency Questions from the Final Quiz

To get a better indication as to which aspect was creating this outcome, two more users who hadn't previously completed the quiz were asked to complete the full game, with the only change being the order of the training examples (now in the order of social proof, urgency, scarcity and misdirection). If the amount of correctly identified patterns remains roughly the same this would indicate it's a feature of the pattern itself, whereas if the distribution changes to weight the last training example more we can identify it as a bias.

Confidence	Attentiveness	Init. Score(%)	Final Score(%)	Improvement (%)	Accuracy(%)
4	4	60	82.35	22.35	50
4	4	70	88.24	18.24	83.33

Table 4.3: Results from Game with different training order

Table 4.3 shows the results of the additional two tests. The final scores and the improvement percentages are consistent with all the other results thus far despite the change of training order. More interesting however is which examples they classified correctly. (Table 4.4.)

Misdirection	Scarcity	Social Proof	Urgency
$1/3$	$2/3$	$3/3$	$0/3$
$2/3$	$2/3$	$3/3$	$3/3$

Table 4.4: Num of correctly identified pattern

Unfortunately from these additional results it is difficult to draw any global conclusion. For the first time in this project we see a result of 0 correctly identified urgency patterns

in one user, however the other user correctly identifies them all, meaning that this first result may more likely be an outlier as opposed to a meaningful result.

It is our belief that the results from the 9-day retesting also support the conclusion that the high rate of urgency correct classifications is due to the appearance of the pattern rather than the training being biased. While these users were shown the training examples in the original order with urgency last, they had long enough of a gap between the training and the retesting that we would expect this to normalise the results somewhat due to the results being affected by the forgetting of the training material. Their results in Table 4.5 show that the only category to remain consistently high across both results is the urgency classification, with both users getting a full 3/3, whereas their results in the other categories have changed. This would indicate that the pattern itself has so little ambiguity and is recognisable enough that it is still recognisable even when others have been forgotten.

Misdirection	Scarcity	Social Proof	Urgency
$1_{/3}$	$0_{/3}$	$2_{/3}$	$3_{/3}$
$1_{/3}$	$2_{/3}$	$3_{/3}$	$3_{/3}$

Table 4.5: Results of 9-Day delayed testing classifications

This section warrants more research, as with a larger set of results both in training with different orders and with delayed testing we could more confidently say that these trends are caused by bias or not. However, with the results present from this pilot study it would suggest that the recency bias does indeed affect people's overall performance, but they have still improved after completion. It also indicates that the type of dark pattern is more important to how well it is recognised than the order that the user is trained.

Chapter 5

Conclusion

In this chapter we discuss in section 5.1 the limitations that affected the study and how they could be mitigated in future research. The paper then concludes with an overall summary and final thoughts in section 5.2.

5.1 Limitations

One of the most obvious is the limitation in the size of the study. To be certain that the results observed here can be replicated it would be necessary to undertake a much larger study across as wider testing group to ensure that the trends observed hold, especially those relating to confidence when different people from more varied backgrounds are trialled. The current results are certainly promising but without further research cannot be generalised to everyone.

Another of the initial concerns was that the study participants would be limited to informatics students who would all be very confident and who may even have heard of dark patterns before. In an attempt to correct for this the opposite problem arose, while the study did feature people who were confident with computing and technology the final results had nobody who currently works or studies in the field of computing. This could be considered a problem as it means there is no data to prove that people that familiar could improve with a game like this. They only represent a small subsection of the target audience however, and it is likely that in a larger study, especially one outside of the bubble of informatics, this wouldn't be enough to make a meaningful difference to the results.

As with any educational tool, this game also requires users to engage with the process to replicate the results shown in this paper. One result was omitted from the final results table as the participant didn't engage with the training material, moving swiftly past without reading it, and instead classified the images based on their opinions on what could be considered manipulation from an ethical standpoint. This ended with their results not changing at all from the baseline to the final quiz output, hence the data being considered uninformative. This does however emphasise the need for the game to be interesting and enjoyable to interact with, as problems like this can arise when

using it in a learning setting such as a school or workplace where participants haven't necessarily chosen themselves to learn about the topic.

5.2 Final Thoughts

This section will conclude the paper and summarise the results found from the pilot study. It will also consider the wider implications of the results and how they could provide an opportunity for future research.

With this project we aimed to prove that interactive games could be an effective teaching tool to inform users about dark patterns in online shopping and to improve their ability to recognise when a website is attempting to manipulate them. The results of the pilot study are very promising, with improvements in dark pattern recognition being seen across the large majority of users. Additional tests also confirmed that even with a short game, users can retain the knowledge past completion.

While other papers in the field of dark patterns aim to create tools to help people, these results help to fill the gap in research into how the users themselves can be effectively taught about dark patterns to better understand the risks of shopping online. By combining these two fields of research there is the potential to have online users who are both informed and who have the tools to combat dark patterns as and when they appear, limiting how many people will be negatively affected by these increasingly prevalent tricks.

As discussed in the previous Limitations section, this study is limited in how far the results can be extrapolated due to the size of the results dataset. Future research has a lot of space to expand on the findings of this paper by increasing the sample size of user testing to ensure that the increase in ability to spot dark patterns observed hold when applied on a broad scale. It would also be beneficial to study how this improvement translates into spotting dark patterns during browsing, or how it impacts user behaviour while shopping. There is also the potential to use this format to compare different types of games, or even different teaching methods to see which are the most effective at informing people of dark patterns.

In conclusion, using games as a teaching tool to inform users about dark patterns has the potential to be an effective and engaging method which can encourage people to take full control of their online shopping experience.

Bibliography

[1] "Measuring digital development: Facts and figures 2022," 2022.

[2] "Customer data platform market size to surpass 47.68 billion with an excellent cagr of 33.70

[3] S. Zuboff, "Surveillance capitalism and the challenge of collective action.," 2019.

[4] B. A. et al, "Americans and privacy: Concerned, confused and feeling lack of control over their personal information," 2019.

[5] "Building consumer confidence through transparency and control," 2021.

[6] T. Sandle, "Report finds only 1 percent reads 'terms and conditions'," 2020.

[7] S. Barth and M. D. de Jong, "The privacy paradox – investigating discrepancies between expressed privacy concerns and actual online behavior – a systematic literature review," 2017.

[8] A. Acquisti, L. Brandimarte, and G. Loewenstein, "Privacy and human behavior in the age of information," 2015.

[9] M. Cao, "Nearly three-fourths of online households continue to have digital privacy and security concerns," 2017.

[10] "The connection between sharing online and losing the data we love," 2017.

[11] "Uk data privacy: What the consumer really thinks," 2022.

[12] "Meta reports second quarter 2023 results," 2023.

[13] M. X. Heiligenstein, "Facebook data breach timeline," 2023.

[14] S. Aslam, "Tiktok statistics," 2023.

[15] M. X. Heiligenstein, "Tiktok data breach timeline," 2023.

[16] I. Shklovski, S. D. Mainwaring, H. H. Skúladóttir, and H. Borgthorsson, "Leakiness and creepiness in app space: Perceptions of privacy and mobile app use," 2014.

[17] H. Choi, J. Park, and Y. Jung, "The role of privacy fatigue in online privacy behavior," *Computers in Human Behavior*, vol. 81, pp. 42–51, 2018.

[18] C. Hallam and G. Zanella, "Online self-disclosure: The privacy paradox explained as a temporally discounted balance between concerns and rewards," *Computers in Human Behavior*, vol. 68, pp. 217–227, 2017.

[19] D. H. Brignull, "Deceptive design - our mission," 2023.

[20] "Dark patterns - how consumer choices are manipulated online," 2021.

[21] D. H. Brignull, "Types of deceptive pattern," 2023.

[22] A. Mathur, G. Acar, M. Friedman, E. Lucherini, J. Mayer, M. Chetty, and A. Narayanan, "Dark patterns at scale: Findings from a crawl of 11k shopping websites," *Proc. ACM Hum.-Comput. Interact.*, vol. 1, no. CSCW, 2019.

[23] R. M. Wood, "Understanding the impact of dark pattern detection on online users," 2023.

[24] M. Nouwens, I. Liccardi, M. Veale, D. Karger, and L. Kagal, "Dark patterns after the GDPR: Scraping consent pop-ups and demonstrating their influence," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, ACM, apr 2020.

[25] I. Wright, "Which retail websites have the most dark patterns?," 2023.

[26] H. J. Smith, S. J. Milberg, and S. J. Burke, "Information privacy: Measuring individuals' concerns about organizational practices," 1996.

[27] A. Hellerstedt and P. Mozelius, "Game-based learning - a long history," 06 2019.

[28] "Chinese chess," 2023.

[29] "Minecraft franchise fact sheet," April 2021.

[30] "Game-based learning with minecraft," 2023.

[31] 2024.

[32] B. Turvey and J. Freeman, "Jury psychology," in *Encyclopedia of Human Behavior (Second Edition)* (V. Ramachandran, ed.), pp. 495–502, San Diego: Academic Press, second edition ed., 2012.

Appendix A

Participants' information sheet

Participants were supplied with the following information sheet prior to taking part in the study to inform them of what data would be collected and how it would be handled.

Participant Information Sheet

Project title:	Privacy Context Game in Human-Agent Sytems
Principal investigator:	Nadin Kokciyan
Researcher collecting data:	Callum Leask
Funder (if applicable):	n/a

This study was certified according to the Informatics Research Ethics Process, reference number XXXXX [edit accordingly]. Please take time to read the following information carefully. You should keep this page for your records.

Who are the researchers?

This project is being researched by Callum Leask, a 4th year Artificial Intelligence student at Edinburgh University, and is being supervised by Nadin Kokciyan, a lecturer in Artificial Intelligence at Edinburgh University.

What is the purpose of the study?

The aim of this study is to research whether interactive video games can be useful as a teaching tool to make people more aware of their online privacy, and ways which they may be being mislead into giving away information online that they wouldn't otherwise.

Why have I been asked to take part?

This study aims to include people from all backgrounds with varying age, occupation and familiarity with the internet and computers as a whole.

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 without giving a reason, up until you complete the interactive game and choose to save your data, at which point is saved anonymously with no identifiable information attached. 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. 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 given access to an online game, which consists of a small teaching stage and an interactive quiz section to test your knowledge. You would be required to play through the game in one sitting, which should not take more than 15-20 minutes. Playing through the game will only be required once, and it can be completed using a standard internet browser wherever or whenever is convenient.

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 significant personal benefits associated with taking part.

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 4 years. All potentially identifiable data will be deleted within this timeframe if it has not already been deleted as part of anonymization.

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/research team, Callum Leask and Nadin Kokciyan (named above).

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, or Sharepoint) and all paper records will be stored in a locked filing cabinet in the PI's office. Your consent information will be kept separately from your responses in order to minimise risk.

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, Callum Leask, at s2083848@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 http://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 s2083848@ed.ac.uk .

General information.

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

Appendix B

Participants' consent form

Participants were asked to sign this consent form to indicate they had read and understood that their anonymised results would be used in this paper.

Participant number:_____

Participant Consent Form

Project title:	Privacy Context Game in Human-Agent Sytems
Principal investigator (PI):	Callum Leask
Researcher:	Nadin Kokciyan
PI contact details:	s2083848@ed.ac.uk

By participating in the study you agree that:

- I have read and understood the Participant Information Sheet for the above study, that I have had the opportunity to ask questions, and that any questions I had were answered to my satisfaction.
- My participation is voluntary, and that I can withdraw at any time without giving a reason. Withdrawing will not affect any of my rights.
- I consent to my anonymised data being used in academic publications and presentations.
- I understand that my anonymised data will be stored for the duration outlined in the Participant Information Sheet.

Please tick yes or no for each of these statements.

		Yes	**No**
1.	I allow my data to be used in future ethically approved research.	☐	☐
2.	I agree to take part in this study.	☐	☐

Name of person giving consent	Date dd/mm/yy	Signature

Name of person taking consent	Date dd/mm/yy	Signature

THE UNIVERSITY of EDINBURGH informatics