Evaluating privacy: determining user privacy expectations on the web.

PILTON, C., FAILY, S., and HENRIKSEN-BULMER, J.

2021

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/cose

Computers & Security

Evaluating privacy - determining user privacy expectations on the web

*Callum Pilton**, *Shamal Faily*, *Jane Henriksen-Bulmer*

Department of Computing & Informatics, Bournemouth University, Fern Barrow, Poole, United Kingdom

ARTICLE INFO

Article history:
Received 14 July 2020
Revised 1 February 2021
Accepted 18 February 2021
Available online 24 February 2021

Keywords:
Privacy paradox
Chrome extension
Case study
Privacy
Privacy settings
Privacy policy
GDPR
Web tracking

ABSTRACT

Individuals don't often have privacy expectations. When asked to consider them, privacy realities were frequently perceived not to meet these expectations. Some websites exploit the trust of individuals by selling, sharing, or analysing their data. Without intervention, individuals do not often understand privacy implications, nor do anything to address it. This study has identified that many users do not have privacy expectations. An extension developed for this study improved privacy awareness, privacy behaviour, and created privacy expectations in participants. The extension also demonstrated that privacy-focused behavioural changes occur when individuals consider the implications of privacy policies, and are exposed to the ways in which their data is being used.

1. Introduction

Poor privacy decisions may lead to undesirable consequences, such as the sale of personal data to unknown third-parties, or unexpected personal data collection for use in newsletters, personalisation, analytics, or even phishing attempts. Data is increasingly valuable and recent changes in law - specifically the principles in Art. 5, section 2 of the General Data Protection Act (GDPR) - have been designed to restrict how data is managed (European Commission, 2018).

Has trust in large online companies become an expectation rather than a consideration, and would perceptions and expectations of trust change if the value of personal data and the ways it could be used were clearer? For some companies, it may be more profitable to accept privacy issues - and potential fines - rather than address them. As people continue to give their data away, are they considering whether the trust they place in websites is warranted?

The privacy paradox claims that people are concerned about their privacy, but usually give it away for relatively small rewards (Pötzsch, 2009). Many people may have already decided to hand over their personal details to several websites, but this decision was based on personal judgement; it is hard to put your trust in someone you have never met, and whom you may never meet, where the website acts as a mediator. Given this, many websites provide transparency of usage through a privacy policy. Privacy policies are used to disclose the ways in which data is gathered, disclosed, and managed, in a legal document; but do people read these, and do they understand their implications? Privacy policies are designed to make privacy decision making transparent, yet

* Corresponding author.
E-mail addresses: callum@callumpilton.co.uk (C. Pilton), sfaily@bournemouth.ac.uk (S. Faily), jhenriksenbulmer@bournemouth.ac.uk (J. Henriksen-Bulmer).
https://doi.org/10.1016/j.cose.2021.102241

74% of web users do not read them (Obar and Oeldorf-Hirsch, 2018).

To date, there has been no resolution of the privacy paradox. Rather than attempting to resolve it, the aim of our work is to educate users on it. We achieve this by providing a platform where users can make more informed privacy decisions on disclosing personal data, along with a better understanding of how that data might be used. Additionally, this study explores whether there are common discrepancies between the content of privacy policies and the actions of websites, and – if there are discrepancies – to determine why these discrepancies may occur and therefore whether it is justifiable to always trust the content of a privacy policy.

Existing work in this area has yielded no clear method for raising awareness to improve user understanding and providing a basis for making better privacy-based decisions to ensure users believe that the trust they place in websites is warranted and meets their expectations. Privacy awareness has influenced how users interact with websites (Paramarta et al., 2019), while privacy extensions can raise awareness (Schaub et al., 2016). And while most users do not read privacy policies, many have an increased trust in a website if it has one (Wu et al., 2012).

Existing tools can block tracking cookies, but provide no context on what it is they are blocking, and how this impacts the user. By using such cookie blocking tools, it can be easy to forget how untrustworthy a website is with some of their worst intensions shielded from view. However, these types of blocking tools are useful for gauging and improving privacy awareness, purely through demonstrating how much privacy-invasive content is being blocked on each website.

In contrast, we present the Privacy Paradox extension: a browser extension capable of interpreting privacy policies, and displaying them to users in a simple, readable, summarised format whereby users can easily assess their perceived trust against their expectations of the content from the privacy policy. This extension automatically scans the organization's privacy policy, comparing this to the actual data usage for their website and producing a concise summary for the user of the results. By combining this intention to raise awareness with a method of evaluating privacy policies in a usable way, the Privacy Paradox extension provides users with a better platform for evaluating their trust in each website, with respect to their own privacy expectations.

In presenting the Privacy Paradox we make three contributions. First, we provide a design and implementation exemplar for an extension for evaluating how the privacy paradox is played out. Second, we present a case study application that demonstrates how the Privacy Paradox extension can be used to evaluate how the privacy paradox is played out. Third, we demonstrate how the Privacy Paradox extension can be incorporated into privacy education/awareness interventions.

In Section 2, we review of related research, considering user privacy and how user access websites, and considering key concepts relevant to this study. In Section 3, we describe the the methods used within the study; these include a study plan, methods for eliciting requirements, and an analysis of the design and implementation elements of the extension and their accompanying evaluation methods. This is followed by a requirements analysis in Section 4, before introducing the designs for the extension and accompanying artefacts in Section 5. In Section 6, we formatively evaluate the Privacy Paradox extension, and summatively evaluate it by considering whether websites are creating realistic privacy policies which meet user expectations and do not contradict the intentions of those websites. Finally, in Section 7 we conclude with a critical analysis of the implications of results, threats to validity, and consider future work.

2. Related work

2.1. *Attitudes to privacy*

Individuals have different observable attitudes towards their privacy depending on demographic and personal experience. Several surveys have identified privacy as some of the most pressing concerns of those using information technology Acquisti and Grossklags (2004), and Kokolakis suggests that a phenomenon known as the privacy paradox shows how individuals do not make privacy conscious decisions Kokolakis (2017).

The privacy paradox – a phenomenon covered extensively throughout privacy literature – highlights the discrepancy between an individual's intentions to protect their privacy and how they actually behave online; the difference between what personal information individuals intend to disclose, and the information which they do disclose are often different. The privacy paradox affects almost everyone, regardless of their awareness or expertise.

Privacy decision making is often determined based on a privacy trade-off. This behaviour is observed in the attribute-attribution framework, which stipulates that attitudes develop and may be reinforced by violations of privacy Norberg and Horne (2007). Norberg & Horne explain that these attitudes can contradict observed behaviours because individuals have an increased likelihood to focus on the immediate benefits from the disclosure of personal information. There are other explanations for this behaviour though, with no accepted definitive reason, including social theory-based theories. For example, younger people are generally perceived to be less concerned with privacy, and choose to maintain their digital image over privacy concerns Blank et al. (2014).

Cranor claims that individuals are typically less willing to provide information when that information is personally identifiable to them, but this could be balanced out: 28% of participants were more likely to provide personal information to a website if it had a privacy policy, while 58% of participants indicated that they would be more likely to provide information if the website also had a trust seal Cranor et al. (1999). However, this work also indicated that none of the participants understood what a trust seal was or how it worked; this indicates that attitudes towards privacy can be influenced by properties of trust, regardless of whether they have been verified as trustworthy.

2.2. *Privacy awareness*

Privacy awareness is a measure of how aware individuals are of the privacy landscape and the privacy decisions which they

make. In theory, by raising privacy awareness, Potzsch contends that people can make informed decisions which should lead to less privacy-invasive behaviour (Pötzsch, 2009). However, in practice, individuals tend to disclose large amounts of personal information, regardless of their attitude and awareness towards privacy. This suggests that users – regardless of privacy awareness – are equally susceptible to the privacy paradox. Nonetheless, a study of the 3 most popular social media websites (Facebook, Twitter, and Instagram) indicated that, by increasing user awareness, the willingness to share information decreases (Paramarta et al., 2019). However, this was typically balanced by trust. When expressed through privacy awareness techniques, this increased the willingness to share personal information. The result of this balance supports Potzsch's claim that users, regardless of privacy awareness, are equally susceptible to the privacy paradox.

Improving an individual's privacy awareness is one of the most suggested techniques in resolving the privacy paradox, but is not an accepted solution. Previous work has shown that, by raising privacy awareness, individuals might disclose less of their personal data than before (Deuker, 2010).

2.3. *Trust*

The constant development of relationships over the internet - amid changing regulations and technology - means that the reliance on trust in technology mediated interactions continues to grow. In parallel, the risk consumers face is also constantly growing; placing trust in a potentially unknown company, with no physical interaction and no physical presence is risky, whether the consumer is technically minded or absent-minded.

During initial or one-off interactions, the signalling of trust-warranting properties is particularly important, and in repeating or returning interactions, an acceptable level of trust is usually already present through temporal embeddedness (Flavián and Guinalíu, 2006). Many trust-warranting properties are often not considered by individuals - including those covered within Section 2.6 - and this can lead to misplaced trust in a website.

Technology can be used in 3 ways in trust interactions, and submits signals prior to a trusting action, be the channel for a trusting action, and be used for fulfilment (Riegelsberger et al., 2005). Properties of contextual trust create the means to warrant trust in another actor. These properties can be continually assessed and evaluated - both cognitively and pre-cognitively - and are essential to Technology Mediated Interactions (TMI). In any transaction, risks, benefits, and trust warranting properties are prevalent.

While Riegelsberger, Sasse & McCarthy welcome the potential of trust mediating technologies to enable interactions that would otherwise not have been possible, they do not consider the individual or legal implications of unwarranted trust interactions and their consequences.

2.4. *Privacy policies*

Data privacy has evolved to cover a much broader scope including the increasing usage of database, cookies, and trackers. Privacy policies help build consumer trust by reducing the fear that a users' personal information will be disclosed. Their content influences user interaction with websites where there is a requirement to provide personal information (Wu et al., 2012).

Additionally, privacy concerns were found to have a significant influence on trust (Peterson et al., 2007). However, Wu et al. do not consider the likelihood that a user will initially engage with privacy policies, and make the assumption that all users who visit a website will read the them, which is a misconception (Wu et al., 2012).

In 2018, Obar and Oeldorf-Hirsch found that 74% of consumers accept privacy policies without reading them (Obar and Oeldorf-Hirsch, 2018). They claim that the "biggest lie on the internet" is agreeing to terms and conditions and suggests that the practice of ignoring privacy policies is so widespread that it points to regulatory failure. Of the participants who did read the privacy policies, 96% were found to spend 5 min or less reading them; a (typical) privacy policy of around 8000 words would take around 15–17 min to read. This suggests that even those who do read privacy policies, do not take the time to digest them; in fact, Obar and Oeldorf-Hirsch found that 86% of study participants who did read the privacy policy spent less than a minute doing so.

While much work in this area focus on links between privacy policy, trust, and the privacy paradox, little exists on the effect of the EU General Data Protection Regulation (GDPR) (European Commission, 2018) on these links. According to Linden, the introduction of GDPR in May 2018 has contributed to an increase of 4.9% in websites now having privacy policies since the law came into effect (Linden et al., 2020).

2.5. *Parsing*

As Obar and Oeldorf-Hirsch [ibid] demonstrated, very few individuals take the time to read a privacy policy. By parsing a privacy policy, it may be possible to create a summary which individuals are more likely to engage with. Parsing is a method of analysing strings of text into logical components to form a conceptual representation. In the context of the proposed artefact, a website privacy policy could be parsed to extract relevant data which can be summarised and presented to the user.

The Platform for Privacy Preferences (P3P) was an attempt to parse privacy policies in a computer-readable format. P3P was published by the World Wide Web Consortium in April 2002, and adopted by browsers shortly after (Cranor et al., 2002). P3P encodes privacy policies in XML format which can then be interpreted by the browser (Cranor, 2003). Its aim was to make it easier to digest, interpret and understand privacy policies - key issues identified by Obar and Oeldorf-Hirsch. However, P3P was obsoleted in August 2018 due to several critical problems and ambiguity such as the absence of formal semantics (Olurin et al., 2012). When considered alongside Obar and Oeldorf-Hirsch's findings, it is sensible to assume that the use of P3P would be limited anyway. Nonetheless, as a potential solution to aspects of the 'privacy paradox', the technology remains an important consideration for modern-day online data privacy.

In 2006, Weitzner et al. adapted P3P as a rule-based policy management system for open deployment on the

web (Weitzner et al., 2005). This work suggests that the web fails at satisfying critical policy requirements such as privacy protection and that by extending the use of rule-based policies to privacy, e.g. an access policy is created to allow users of a particular role to access specific data based on the privacy settings of that specific data.

Several technical challenges prevent either P3P or policy-aware web from becoming successful. For either implementation to work, wider changes to local and remote technologies are necessary. Both technologies require support from both the webpage and client (browser) to work; Internet Explorer and Edge were the only major browsers which supported P3P, although support has since ended for these too, and Weitzner et al.'s policy-aware web is not yet supported by any major browser.

2.6. Web tracking

Web tracking uses technical tracking methods to determine specific information about an individual. This information can restore previous browsing sessions, display personalised adverts, and determine the location of an individual. Most people believe that personal information collection is used primarily for targeted advertising, but recent studies have found that web tracking is being used for many other purposes, including: price discrimination, personalisation of search results, and government surveillance (Bujlow et al., 2017). Moreover, third-party tracking is becoming an increasing privacy threat; around 46% of the 10,000 most popular websites are monitored by at least one third-party tracker (Li et al., 2015).

Storage-based tracking mechanisms depend on storing data on personal computers. The most common form of storage-based tracking mechanism is a cookie: a small piece of data stored in the user's browser. Local storage is another form of storage-based tracking mechanism which allows for larger data to be stored in a similar way.

Fingerprinting is a method of identifying a device by creating a unique key. It uses a broad range of technologies to create a unique identifier for a device, based on a range of factors, such as operating system, browser version, and screen size. Fingerprinting does not need to create any storage on the user's device, but changes to the device such as increasing or decreasing the browser size can cause inconsistency when identifying devices.

Once a tracking mechanism has successfully identified a user, a website may then request additional data on that user from their own, or a third-party database. This is done in the form of GET/POST requests, these are requests made to websites or web services which usually send a response. Once a user is identified, a GET/POST request could be made to retrieve that user's profile picture, but is more frequently used for requesting data about that user from third parties, including their browsing history.

The privacy paradox extension presented in this paper, considers all the above tracking methods and attempts to intercept or access them to display this information to the user.

2.7. Security & privacy extensions

As of March 2015, Google Chrome had the largest browser user base with 63.7% of users (Tsalis et al., 2016). With the most browser users, Google Chrome is the logical choice for development of an extension. Due to its popularity, the Chrome Web Store contained around 43,000 free extensions in 2017 (Sjösten et al., 2017). While it is recognised that this saturation could decrease visibility; (50% of all Chrome extensions have fewer than 16 installs (Extension Monitor, 2019)), according to a study by Tsalis in 2016, privacy extensions make up the smallest share of security and privacy add-ons in the Chrome Web Store, at only 7.7% (Tsalis et al., 2016). Moreover, privacy extensions fall well below the average across all browsers of 14.15%, indicating a potential gap which, we contend, the privacy paradox extension could exploit.

Existing privacy informative extensions such as Ghostery and Disconnect aim to influence the privacy awareness of users, much like the Privacy Paradox extension we present. However, these extensions focus on blocking trackers, rather than analysing collected data and privacy policies. This is where the Privacy Paradox extension is novel, by providing individuals with clear information on website data collection, and allowing individuals to determine for themselves if their privacy expectations are being met.

Websites attempt to address user awareness issues with privacy policies (Schaub et al., 2016), but few extensions attempt to improve their awareness, viability, or readability - and none can achieve all three. Extensions can empower users by highlighting tracking activities, yet no tool exists to inform users of the data they are giving away; individuals are unable to determine whether the privacy implications of using a website meet their privacy expectations.

In their work on extension design (Schaub et al., 2016), Schaub et al. determined that alerts should be used sparingly to prevent them from becoming annoying. The information displayed in the main panel should also be relevant, understandable, and actionable to users; where setup videos, tutorials and/or a website were available, users had greater trust in the extension.

2.8. Security problems with chrome extensions

Google Chrome extensions combine a mix of popular technologies including: HTML (Hyper Text Markup Language), CSS (Cascading Style Sheets), and JavaScript. HTML and CSS - which are both static languages - are used for creating the views within extensions, while JavaScript is used to provide application logic (Mehta, 2016). Information about the Chrome Extension is located within a manifest file. A manifest file is a JSON-formatted file and can include the name, description, and version of the extension as well as more specific details such as permissions and icon locations. Further, extensions are sandboxed within the browser, meaning that they are isolated from other extensions for improved security.

When designing extensions for Google Chrome, the security of the extensions itself is rarely considered; Carlini et al. discovered 70 vulnerabilities across 40 extensions, in a security review of 100 Chrome extensions (Carlini et al., 2012). During the security review, scripts fetched over HTTP were re-

sponsible for 56% of the vulnerabilities found. All these vulnerabilities could be prevented using HTTPS requests (in place of HTTP requests). Thus, in creating the the Privacy Paradox extension, script injection will be avoided.

Extensions have access to special privileges within the browser, making them an appealing target for attackers (Google, 2020). Precautions should be taken to follow good security architecture practices when designing an extension to enable code transparency and demonstrate the trustworthiness of the extension itself.

2.9. *Design and evaluation approaches*

Several design approaches have been considered for creating an extension, including prototyping. However, two design approaches have been considered in more depth; these are Nielsen's heuristics and IBM Design Thinking.

Neilsen's heuristics contain 10 principles for user interface design which were originally developed for heuristic evaluation and 'refined based on a factor analysis of 249 usability problems' (Nielsen, 1995). The 'user control and freedom', 'flexibility and efficiency of use', and 'aesthetic and minimalist design' principles may be especially useful for the the Privacy Paradox extension as they will ensure that the design are effective, efficient, and satisfactory. However, given the time limitations imposed by this study, it may be difficult to implement all the design principles.

IBM Design Thinking is a human-centred approach for creating 'human-centred outcomes at speed and scale' (IBM Studios, 2016). It involves an approach of 'applying design thinking at the speed and scale of modern enterprise demands' and consists of three principles: a focus on user outcomes, multidisciplinary teams, and restless reinvention.

While IBM Design Thinking is focused towards enterprise usage, there are elements of it which may be useful in design creation of the Chrome extension, specifically the human-centred elements described as a focus on user outcomes.

Several evaluation approaches are available for evaluating models, approaches, designs, and outcomes (Preskill and Russ-Eft, 2012). Of these, we consider the Behavioural Objectives Approach and Goal-Free Evaluation.

The Behavioural Objectives Approach, focuses on the degree to which the objectives of a program have been achieved (Preskill and Russ-Eft, 2012). In the context of this study, this approach can be used to evaluate whether the Chrome Extension achieved its objectives.

The Goal-Free Evaluation focuses on actual, rather than the intended, outcomes (Preskill and Russ-Eft, 2012). Our work can be used to evaluate what the actual effects of the Chrome Extension are (including any unintended side-effects). However, it should be considered that guidance for conducting a goal-free evaluation is limited (Youker, 2019).

Evaluation results can be affected by external factors. The Hawthorne effect is an experimenter effect whereby participants in a human-centred study may exhibit atypically high levels of performance as the result of an understanding that they are being studied (Macefield, 2007). This change in performance occurs because participants believe that changes in studies will improve their ability to perform. The aim of the the Privacy Paradox extension is to improve the privacy awareness of individuals and participants may appreciate that there will be a desire for them to display improved privacy awareness when using the artefact. As such, participants may subconsciously perform more effectively. However, there is a large amount of controversy surrounding the Hawthorne effect, and there are many different interpretations of the original phenomenon (McCambridge et al., 2014). Only experimental solutions to the Hawthorne effect exist (Mayo and Dooley, 1968), and can impact the behaviour of participants (Sedgwick and Greenwood, 2015). However, it could be considered unsafe to criticise an experimental study based on the Hawthorne effect alone.

3. Methodology

3.1. *Study aims and objectives*

The aims and objectives of our study were determined using the 'SMART' criteria used in programme planning to describe programme expectations (Toffler, 2013). SMART objectives must follow specific criteria to be specific, measurable, achievable, realistic, and time-bound; this approach to defining objectives creates goals which are more likely to be attainable.

The first aim was to evaluate whether two websites, Amazon and Facebook, meet their prescribed privacy policies, before then considering whether these meet the expectations of a sample group of users interacting with them (see Table 1). We achieved this by evaluating the trust between users and these websites with respect to the data collection that takes place, compared with the content of the privacy policies, thereby helping us determine whether these websites act with social responsibility (with regards to user expectations) to abide by their privacy policies (see Section 6).

To support the creation of Privacy Paradox, the results of the background study were used to shape the expectations of the extension, by understanding existing technologies, capabilities, and ensuring that the extension could be considered novel when compared with other privacy-based tools.

3.2. *Requirements elicitation*

Requirements were elicited based on an understanding of similar extensions, and through other requirements gathering processes, including discussions with prospective users. The combination of personal and research-based requirements gathering – including the analysis of similar extensions in

Table 1 – Summary of study objectives.

Objective	Description
O1	Evaluate whether Amazon's and Facebook's privacy policies meet user expectations.
O2	Evaluate whether the data collection taking place on Amazon's and Facebook's websites meet user expectations.
O3	Create a Google Chrome extension to improve the privacy-awareness and evaluate website violations.

Section 2.7 – allowed for an in-depth set of requirements to be specified. The resulting requirements focus on isolating the opportunity for creating something novel, while ensuring that the purpose of the extension remains well defined, in line with the expectations laid out during requirements gathering discussions. The requirements are detailed in Section 4.

3.3. *Design method*

We used a prototyping methodology to designed and develop the extension, enabling continuous development and improvement throughout each stage of this study. Prototyping is a good tool for facilitating ongoing evaluation and rapid development thus, users are regarded more likely to adopt any system created (Avison and Fitzgerald, 2006). Thus, prototypes of the extension were developed early to speed up development, and improve the likelihood of overcoming obstacles earlier in the Software Development Lifecycle (SDLC). Prototypes were also used to provide context during the design and implementation stage.

3.4. *Evaluation design*

We conducted formative and summative evaluations to evaluate the extension. Formative evaluation was conducted on the extension designs from a usability perspective using early prototypes of the extension. During the evaluation, 41 participants took part in formative assessment and 19 participants were involved with summative assessment. Formative assessment was spread across 4 focus groups and 11 1-to-1 interview sessions. An ethics approval was requested from and provided by the University prior to the commencement of the study.

3.4.1. *Formative evaluation*

Various methods for determining a suitable design for the extension were considered including focus groups, interviews, and surveys. From these, conducting a survey was determined to be the most efficient and effective method of evaluating the design of the extension. A survey was used in order to facilitate access to a larger pool of participants, when compared to the other evaluation methods, due to the ability to complete the evaluation quickly and remotely. This approach gave participants the opportunity to engage with the extension throughout the SDLC. The formative evaluation was conducted with anonymous participants and focused on three usability principles: effectiveness, efficiency, and satisfaction. This evaluation was then used to enhance the extension, prior to any summative evaluation taking place.

3.4.2. *Summative evaluation*

Summative evaluation was used to determine which designs users preferred from the extension as part of the prototyping phase. The summative evaluation of the extension was conducted in focus groups after the feedback from the formative evaluation had been quantified and incorporated. Focus groups were selected and participants were encouraged to provide constructive feedback for the extension based on their capabilities to provide detailed individual and collective comments.

Two focus groups consisted of technical participants, while the other two consisted of non-technical participants. By separating the focus groups in this way, it was possible to draw conclusions on differences between technical and non-technical usage of the extension, and how this relates to privacy behaviour.

During the evaluation, Goal-Free evaluation and the Behavioural Objectives approach (as described in Section 2.9) were used to evaluate the success of the extension in highlighting privacy-issues. Goal-Free evaluation was used to realise the true outcomes of the extension and can be achieved by providing participants with a platform for open discussion about the outcomes of using the extension. The Behavioural Objectives approach was be used to determine whether participants behaved with an increased privacy-awareness after using the extension; this was achieved by allowing the participants to use specific websites with and without the extension and determining differences in behaviour, where desired behaviour would show an increased desire to attempt to understand the privacy risks facing them on each website. This desire could be expressed through extensive use of the extension, its popup, or by showing an increased interest in understanding the privacy-related consequences prior to making privacy-based decisions.

Following the focus groups, participants were given the option to participate in an interview which analysed their behavioural changes [if any] after using the extension over a longer period. The interview consisted of open questions to determine whether privacy expectations were being met, and whether participants noticed changes in their privacy awareness.

More detailed information on the approaches taken to evaluate the extension are explained within Section 6.

4. Requirements

Requirements were gathered through the process explained in Section 3.2. A mixture of background (document-based) research and user feedback was used to determine the requirements which are outlined below.

4.1. *Extension specifications*

The extension was devised to provide privacy transparency to users in two key areas: privacy policies, and tracking methods. From this, a set of requirements were elicited, and summarised in Table 2.

The first three requirements (R1-R3) formed the primary focus of the extension design and feature prominently in the finished extension. Requirements R4 - R6 focus on providing additional value, novelty, and detail to the extension.

4.2. *Access, analyse, and present privacy policy*

The extension shall be able to access, analyse and present the privacy policy of a website in a summative form, and thereby allow the user to make a faster and more transparent privacy decision based on a better understanding of the privacy implications they may face.

Table 2 – Requirements list.

Requirements Id	Specification	Reference
R1	Access, Analyse, and Present Privacy Policy	[4.2]
R2	Intercept and Display Trackers	[4.3]
R3	Detect Privacy Violations	[4.4]
R4	Full Report of Findings	[4.5]
R5	Report Privacy Violations	[4.6]
R6	Privacy Summary Popup	[4.7]

Satisfying this requirement helps users circumvent the challenges faced in Section 2.4, specifically related to the tolerance of time spent analysing policies, and the high rate of individuals who ignore them completely.

The privacy policy summary will be categorised into four clear sections: Data Collection, Data Usage, Choices, and Tracking.

4.2.1. *Data collection*

Personal data being collected by the website shall be obtained and summarised as a readable list. To conserve space, the extension will only show whether personal data is collected.

4.2.2. *Data usage*

If data is being collected, how it is used shall be obtained and displayed. This includes whether personal data will be shared to third-parties, or for personalization purposes. Additionally, data security will be considered within this category; if the privacy policy does not mention any steps which are in place to protect personal data, the extension shall assume that there are none.

4.2.3. *Choices*

The choices users can make regarding their personal data shall be summarised within this category. This will include whether users can reject trackers and data collection, and whether there is any consequence for doing so. Furthermore, the extension shall determine whether users can request a copy of their personal information, as they are legally entitled to do under Article 15 of the GDPR (European Commission, 2018).

4.2.4. *Tracking*

A summary of the ways which users may be tracked shall also be included. The tracking types will include analytics, usability tracking, tracking for personalised advertising, and tracking cookies.

4.3. *Intercept and display trackers*

Trackers – specifically those described in Section 2.6 – shall be displayed via the extension so the user can view the data collection techniques in action against them in real-time.

This requirement provides transparency of use by alerting users to the personal data being collected, and allows for comparisons to be made between the user's expectations based on the privacy policy, and the reality perceived from real data collection. This empowers users to make better privacy decisions, and act upon potential privacy violations which they may not have otherwise been alerted to. By providing users with a thorough understanding of the implications on their privacy through the personal data being collected, privacy-based decisions awareness should be improved.

Five categories of tracker should be displayed to users: Session Trackers, Advertising Trackers, Ad-Block Detection, Location, and Fingerprinting.

4.3.1. *Session trackers*

Session trackers identify users and enable website personalisation. They typically provide a continuous user experience, such as keeping users logged in between pages. However, they can also be used for identifying individuals from third parties and enable targeted advertising practices. As a result, users should be warned when session trackers are being used.

4.3.2. *Advertising trackers*

Advertising trackers collect information about users from other webpages, enabling targeted advertising. If a website is collecting personal data for the purposes of showing targeted adverts, users should be made aware from the extension.

4.3.3. *Ad-block detection*

Ad-Blocking tools have become common within most browsers. These are used for hiding or removing adverts from webpages. The extension should be able to detect when a website is running explicit tests to check whether an ad-blocker is being used, as this may suggest that the website will attempt to circumvent the ad-blocking and thereby restrict the users choices when using the website.

4.3.4. *Location*

Websites tracking the specific location of their users should be detected by the extension. While the usage of location data may be justifiable, that may not always be the case. As such, the extension should make users aware of when their location was specifically requested and/or exposed by the website.

4.3.5. *Fingerprinting*

Fingerprinting identifies devices based on characteristics. This practice can be used in a similar way to session trackers. The extension should be able to detect device fingerprinting and inform the user if specific device or personal details were requested as part of this process.

4.4. *Detect privacy violations*

The extension shall use the data resulting from the 'Access, Analyse, and Present Privacy Policy' and 'Intercept and Display Trackers' requirements to detect discrepancies between the stated usage of personal data in privacy policies and actual usage through trackers.

This shall be done on a comparative basis where statements made in the privacy policy can be compared against the actual result as determined by the extension (e.g. the privacy policy claims cookies are not in use but the extension detects the usage of cookies).

Privacy policy specific privacy violations will also be counted. A privacy policy specific violation could include failure to mention data security measures or providing no option to request a copy of your personal data. The absence of a privacy policy from a website will be considered as a privacy violation.

The number of violations will be displayed to the user from within the extension. Full details of each privacy violation shall be recorded and presented to the user in the full report of findings.

4.5. *Full report of findings*

A full report of findings from the extension shall allow users to achieve further transparency of use when analysing the privacy impacts of websites. This should put them in a better position to make privacy-based decisions. The full report should show the name of the website which it was created for and include a timestamp of when it was created. The full report shall include the following sections: Summary of Findings, Tracker Count, Tracker Analysis, Privacy Policy Analysis, and Additional Information.

4.5.1. *Summary of findings*

The first section in the full report shall summarises the tracking data and privacy policy analysis displayed within the extension. This will show the same information as is shown in the extension to aid clarity and understandability.

4.5.2. *Tracker count*

A section showing the number of trackers shall include a count for the following trackers: cookies, local storage, and GET/POST requests.

4.5.3. *Tracker analysis*

The report shall include a justification for why the extension determined each of the trackers in the 'Intercept and Display Trackers' requirement to be in use or not. For each detected tracker, at least one justification will be provided. For example, if location tracking was found to be in use, the justification could be that a request for 'GPS' was made by the website.

4.5.4. *Privacy policy analysis*

The report shall break down the findings of the privacy policy into smaller categories, and justify whether they have been detected in a similar way as Tracker Analysis. The extension will look for – and display to the user – a set of privacy-related data from the privacy policy as outlined in Table 3

4.5.5. *Additional information*

Any additional information collected by the extension shall be displayed to the user including any email addresses found within the privacy policy, and any violations described in the 'Detect Privacy Violations' requirement.

4.6. *Report privacy violations*

The extension shall list potential privacy violations and provide users with the capability to reporting the violation(s) to the website. Users will be shown the name of the website they are reporting from the extension and a timestamp of when the violation(s) were detected.

When reporting a violation, the following information will be visible to users: Summary of Findings, Additional Information, and Email Creator.

The Summary of Findings and Additional Information will contain the same information as those described in the 'Full Report of Findings' requirement.

4.6.1. *Email creator*

Users shall be able to create an email from within the extension. This will involve entering an email address, a body of text, and allowing the user to send the crafted email from their personal email account. The email creator shall make the experience of writing a privacy violation email easier by providing all the necessary information on the same page.

4.7. *Privacy summary popup*

A popup shall appear when creating a new account on a website. The popup shall display a summary of the websites privacy policy, allowing the user to make an informed decision on whether the privacy implications of using the website outweigh the potential benefits, prior to creating an account.

The popup will include the same information as determined in the 'Access, Analyse, and Present Privacy Policy' requirement including Data Collection, Data Usage, Choices, and Tracking.

The popup shall add an icon to register buttons on websites, whereby hovering over this icon will show the popup. As such, the extension will be able to detect when a user is creating a new account to show them the popup only when relevant.

4.8. *Architecture*

4.8.1. *Extension security*

For the purposes of this extension, security is not to be considered a primary concern. The extension does not create or introduce any new behaviour which could compromise the security of the user's device or personal data. Instead, it focuses on intercepting data and behaviour which already exists. The extension will also not collect any information on its users either locally or remotely; a privacy policy will be created to reflect this (see Section 5.5) (Fig. 1).

During the design and implementation of the extension, best practices were adhered to; this includes following Google Chrome's Extension Security guidelines (Google, 2020).

Table 3 – Tracker items.

Tracking	Usability Tracking	Analytics	Cookie Usage
Advertising	Advertising	Recommendations and Personalisation	
Data Type	Data Types	Personal Data Retention	
Collection	Direct Data Collection	External Data Collection	Personal Data Release
Control	Ability to Request Personal Data	Ability to Reject Data Collection	
Security	Personal Data Security	Consequences of Rejecting Data Collection	

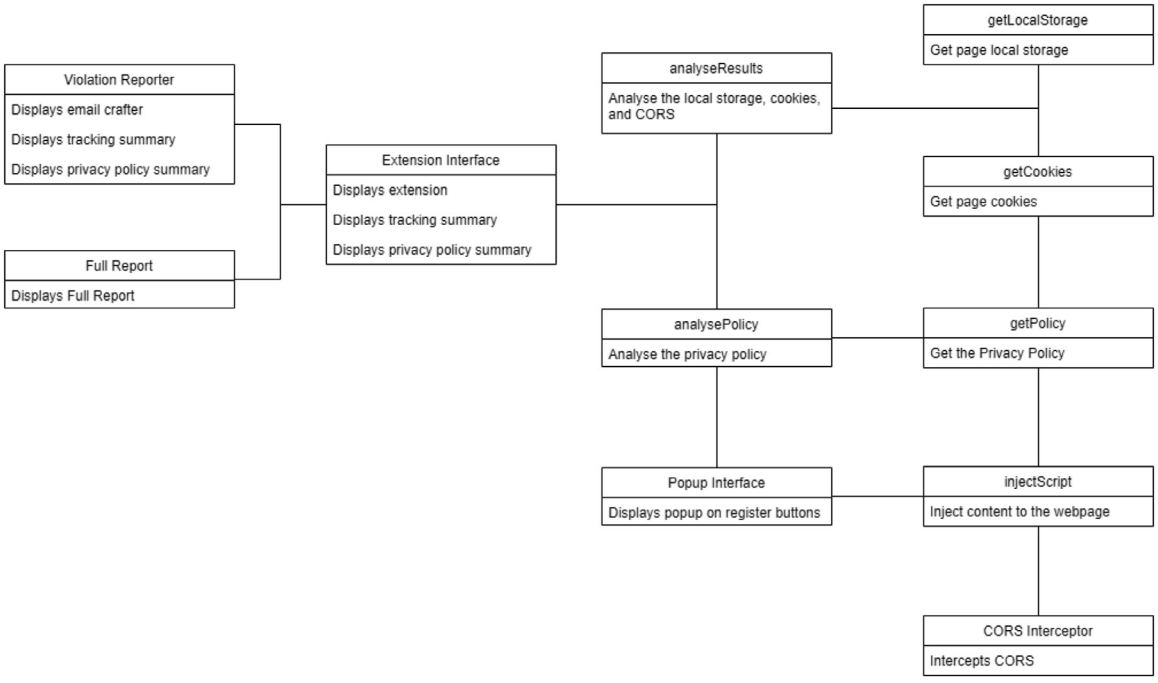

Fig. 1 – Extension architecture.

4.8.2. *Messaging API*

Google Chrome extensions have access to the same API as the browser, which provides an efficient method of passing messages between the extension itself and its back-end logic through the *chrome.runtime* messaging API. This enables messages to be sent and retrieved outside the boundary of the extension and therefore allows for external pages such as the websites privacy policy to be fetched.

4.8.3. *Versioning*

Semantic versioning practices were followed to ensure consistency. This involves splitting the version into 3 numbers separated by a full-stop, e.g. 1.0.2. The first number is used to define a major version, the second number is used to define a minor version, and the third number is used to define the patch version.

4.8.4. *Permissions*

Privacy policies are not always stored on the same domain as the rest of a website, and therefore the extension needs access to the '<all_urls>' permission, which allows GET/POST requests to Cross-Origin Domains. Therefore, the extension required manual review by Google when uploaded to the Chrome Web Store, and subsequent reviews each time it is updated.

5. Design & implementation

5.1. *Extension designs*

In designing the Google Chrome Extension, the name 'Paradox' was chosen. This was intended to represent the privacy paradox, and subtly raise awareness to the theory. Three wireframe designs were created to be presented to users. The three designs include different ways of displaying tracking and privacy policy specific information, offering alternate usability experiences.

During the creation of each design, Nielsen's heuristics for usability (see Section 2.9) were considered and implemented, to make each design relevant, necessary, and productive for its prospective audience.

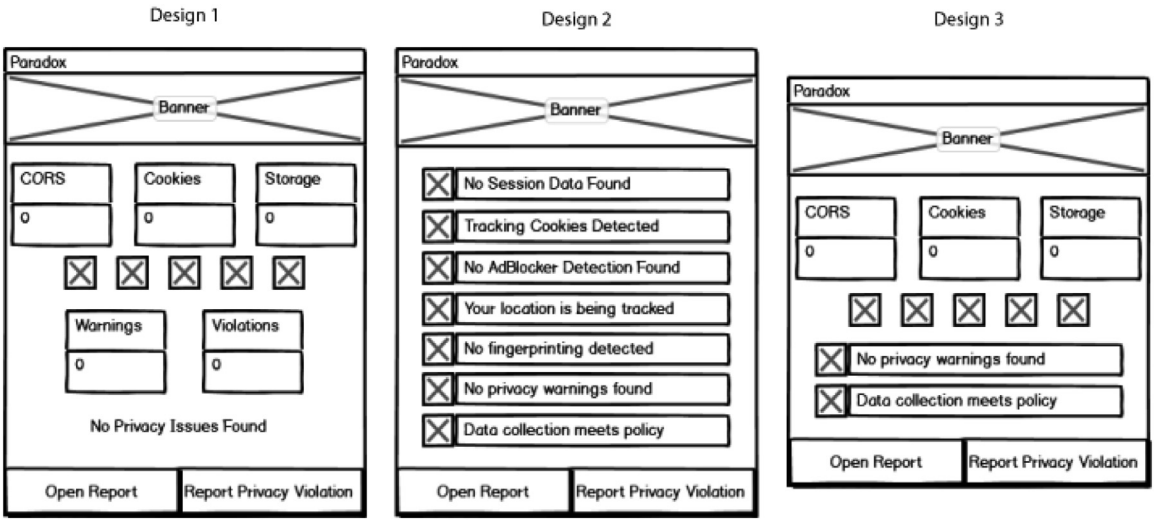

Fig. 2 – Paradox extension designs (Wireframe).

Fig. 3 – Paradox extension designs (Graphical).

In addition to Nielsen's heuristics, IBM Design thinking approaches was used to create human-centred design. As discussed in Section 2.9, this design approach was compared with others, and it was determined that this approach compliments Nielsen's heuristics while enforcing a focus on user outcomes. Therefore, a hybrid combination of both approaches was followed to ensure that focused, human-centred design remained central to the creation of the extension. Further, the first author's personal experience as a certified IBM Enterprise Design Thinking Co-Creator was employed to ensure IBM Design thinking approaches were used to ensure designs meet user expectations.

By separating layouts from graphics, colours, and images, a single preferred design could be realistically identified. To facilitate this, we used both wireframe and graphical designs. Wireframes were used to introduce concepts and layouts (Fig. 2), while graphical representations were used to demonstrate potential colour schemes and images (Fig. 3). Participants were asked their opinions for the layout of each design - considering effectiveness, efficiency, and satisfaction - after being shown the wireframes, but before being introduced to the graphical representations. This ensured that participants opinions for each design were not influenced by graphics.

All the designs included a banner where the Paradox extension logo was be displayed to the user. Additionally, all the designs featured twin buttons: 'Open Report' and 'Report Privacy Violation'. The appearance of these buttons were unchanged in each design as they represented important features of the extension, which enabled functionality. Therefore, it was important to maintain visible continuity.

The first of the three designs – 'Design 1' – used a bold visual approach which focuses on expressing the number of trackers and warnings as physical numbers to the user. This would enable users to make privacy-based decisions based upon the size of the number in each box.

The second design – 'Design 2' – focused on a heavier text-based approach. Each tracking method included an icon,

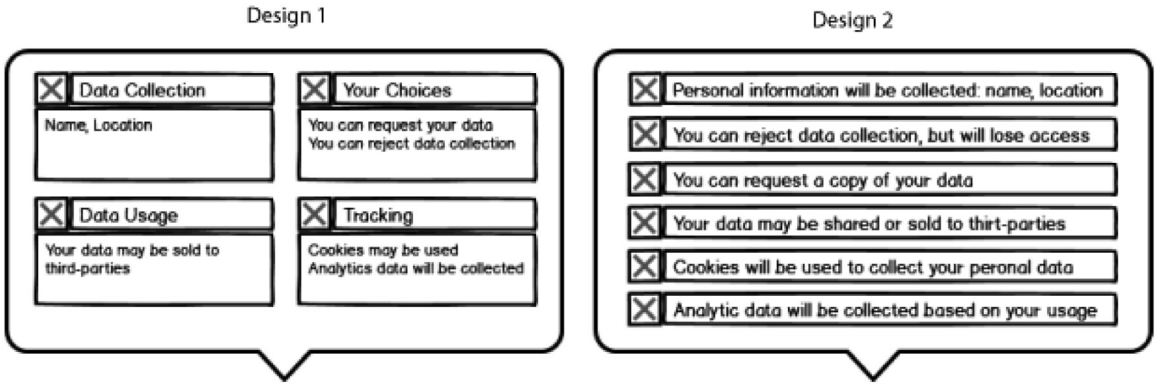

Fig. 4 – Paradox popup designs (Wireframe).

which used the 'traffic light' metaphor, and was accompanied with textual rationale.

The final design – 'Design 3' – combined the previous two designs. It expressed the trackers as physical numbers together icons for individual tracking methods, and displayed the privacy policy related data using icons with textual rationale.

In addition to the wireframe designs, graphical representations of each design were created to support the requirements specified in Section 4.

5.2. *Popup designs*

Designs for the summative policy analysis popup were also created in the design process. Two designs were created to represent different ways in which the policy analysis could be presented to users.

The first design – 'Design 1' – separated the analysis into 4 categories: 'Data Collection', 'Your Choices', 'Data Usage', and 'Tracking' (Fig. 4). Each category had a unique icon with either a green tick or an amber warning. The colour red was not used in this design to encourage users to investigate potential privacy policy issues, rather than write them off as critical problems without evaluating. Each category also contained short sections of text which justify the colour and image of the icon. The intention of this design was to improve clarity by using easy to understand headings, enabling non-technical users to make decisions on privacy-based concerns, easier and quicker.

The second design – 'Design 2' – addressed 6 separate, but specific privacy concerns: personal information collection, the ability to reject data collection, the ability to make an information request, whether information will be passed to third-parties, whether cookies are used, and whether analytical data will be collected. Each concern comprised of an icon (green tick or amber warning) and justification. This design was intended to make it clearer to distinguish between important privacy concerns (Fig. 4).

5.3. *Other designs*

Designs for the 'Full Report' and 'Report Privacy Violation' pages were created prior to their implementation. Both designs contain identical banners for consistency and split the pages into sections of relevant information to make them easier to read.

5.4. *Full report*

The 'Full Report' page was designed to present a detailed summary (Fig. 5). The data found and displayed in the Paradox extension was presented under key-headings, with specific issues identified under sub-headings. 'Tracker Analysis' and 'Privacy Policy Analysis' sections shall provide justification for the explanations provided in the summary. The tracker analysis was split into: session data; advertising trackers; ad-block detection; location tracking; and device fingerprinting. Each section shall contain at least one justification explaining why the Paradox extension determined that specific tracker to be prevalent or not. The privacy policy analysis was split into: advertising; cookies; data release; data retention; data security; direct data collection; external data collection; reject data collection; reject data collection consequence; recommendations; third party sharing; usability tracking; analytics; and data types. These factors were chosen as they represent common themes throughout privacy policies and allowed comparison with relevant laws such as GDPR. For each factor, at least one justification was displayed to the user.

5.5. *Privacy violation reporter*

The 'Privacy Violation Reporter' page design provided a summary of information collected by the Paradox extension that provides the user with a method for creating an email on the same screen (Fig. 6). The design also showed users potential email addresses they can contact regarding privacy-related issues for the website they are reviewing. In addition, the page enabled the user to compose an email, which can be sent directly from the page or copied for sending from the user's personal email account.

5.6. *Google chrome web store*

Four graphics were created for use in the Chrome Web Store to make the Paradox extension look more appealing and demonstrate functionality. The first of these aimed to summarise

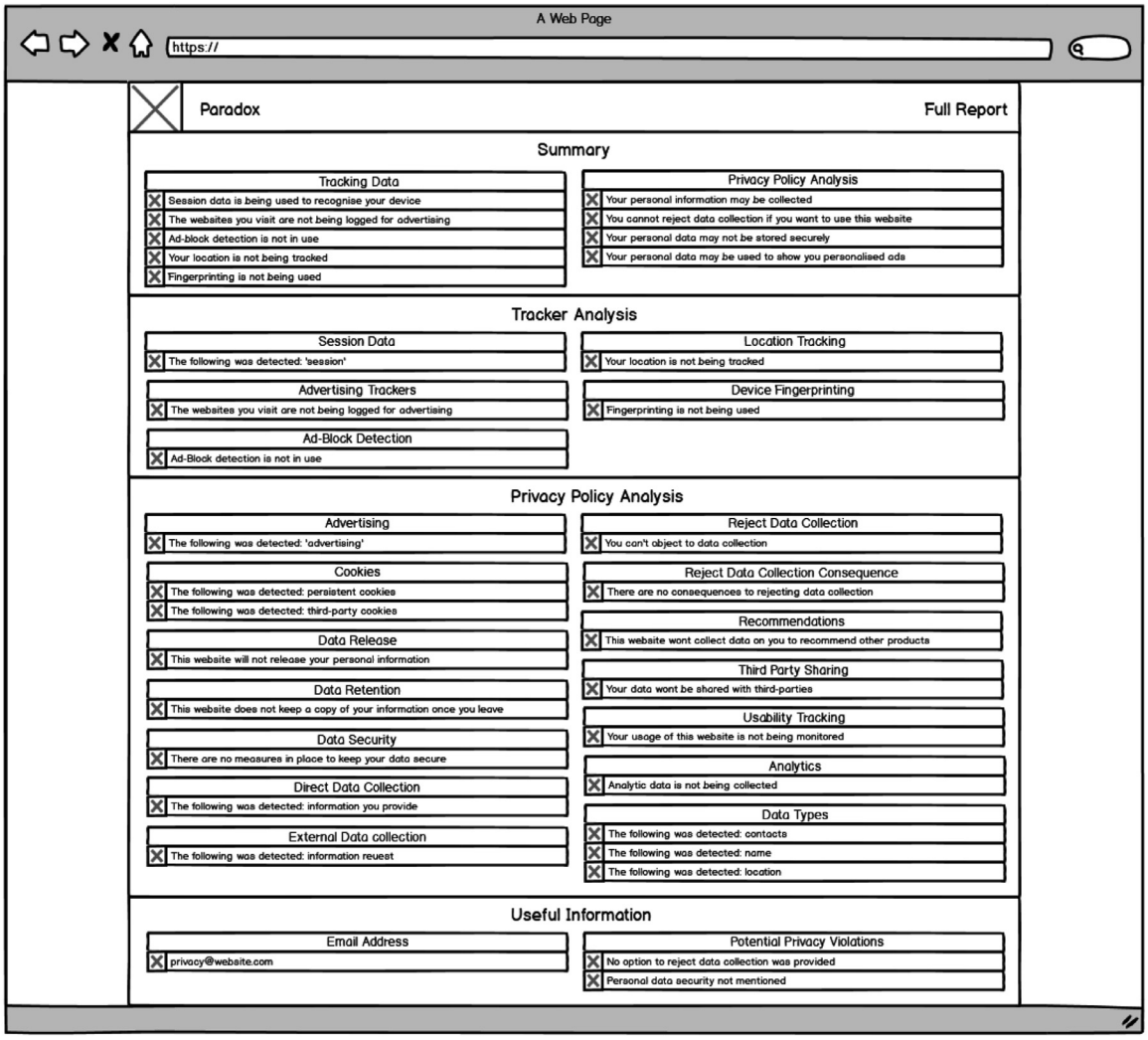

Fig. 5 – Paradox full report page design (Wireframe).

the core functionality of the Paradox extension, including its ability to intercept trackers, analyse privacy policies, and report privacy violations (Fig. 7). The second image, depicted in Fig. 7, focused on demonstrating the user interface to users and shows the extension and popup interfaces from Version 0.1.

The images aimed to provide a basic, clear, and informative way for users to understand the simple concepts and functionality of the extension. Additional detail – including a section on how to use the extension – could be found on the the Paradox extension website; a link directly to the website is included on the Chrome Web Store. Additionally, a direct link to the privacy policy was also added, along with a link to email support@privacyparadox.co.uk for additional information.

5.7. *Website*

As stated in Section 2.7, users with access to 'setup videos, tutorials and/or a website' had a greater trust in the extension (Fig. 8). Therefore, a website for the extension was created in attempt to achieve this greater trust. The website includes an overview of the extension, a page on how to use the extension, and a privacy policy. A link to this website is provided from the extensions Google Chrome Web Store page. The website was created to provide an overview of the Paradox extension, as well as some information on how to use it; the site followed the same colour scheme and design patterns as the extension designs in Fig. 3.

The homepage showed a clean and simple title, with an option to visit the Google Chrome Web Store to get the extension; Fig. 9 provides a quick overview of the key components of the extension.

The 'Using Paradox' page is targeted at providing a simple, visual expression of how to use the extension. Images are used to help convey information with arrows pointing towards relevant information to help users understand and use the extension (Fig. 10).

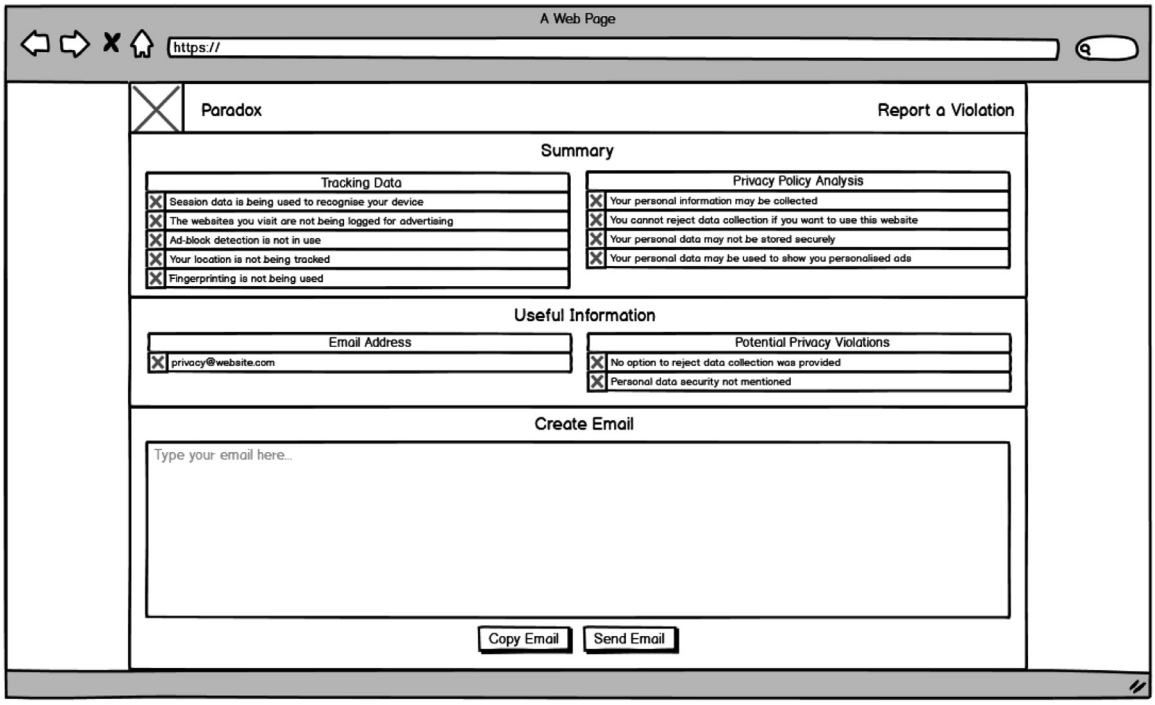

Fig. 6 – Paradox report privacy violation page design (Wireframe).

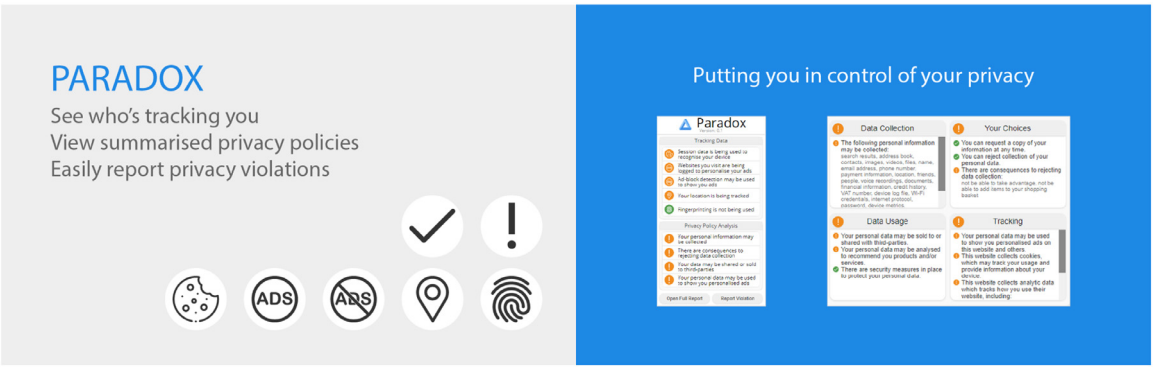

Fig. 7 – Paradox images for chrome web store.

5.8. *Extension code implementation*

The extension can be split into 3 distinct parts: the logic, the interface, and injected content. Source code is available via https://github.com/shortbread31/Paradox.

The key functionality of the extension was split between tracking detection, and policy evaluation. Tracking detection analysis Cross-Origin Requests, Cookies, and Local Storage, for: session data, ad-focused tracking, ad-block detection, location data, and fingerprinting. The occurrences of each tracking method were counted and displayed to the user. Privacy policy evaluation finds and analyses the privacy policy of the website to create a summary of the number of warnings (where data is being collected, used, or shared) and violations (where the policy is either in breach of GDPR, or contradicts actual data collection).

The extension also added a popup to the 'Create Account' and 'Register' buttons, which displayed a summary of the analysis of the privacy policy. This provides users with a summative analysis of the policy prior to completing account creation. This is not intended to replace the ideal scenario of reading a privacy policy in full, rather it is a means to aid the process. However, it is anticipated that users may choose to use the popup to replace the necessity of reading the entire privacy policy.

An option to view a full report of findings opened an external webpage with a detailed report of all analysis carried out by the extension. The full report was intended to help technical users make privacy-based decisions when the base information provided by the extension is not enough. Furthermore, the additional data captured and presented in the full report could be used for reporting a violation.

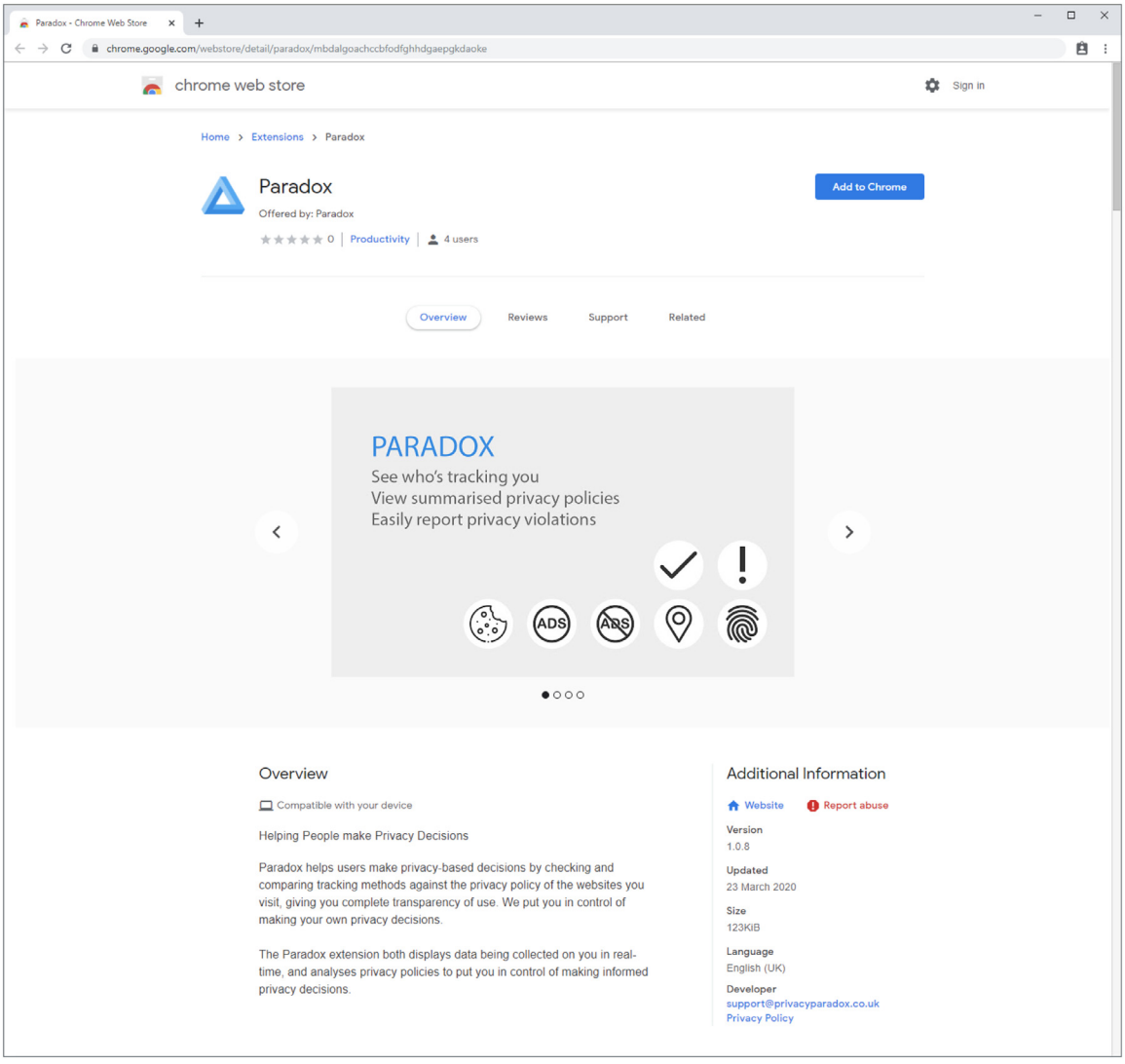

Fig. 8 – Chrome web store listing for paradox.

There was an option to report a violation from the extension, which may be recommended to users upon discovering privacy violations. To report a violation, a new window was be opened, which provides users with the relevant email address for the website, and the findings from the extension. This enables users to build an email template, which can then be sent on from their personal email address.

5.9. *Extension algorithms*

The logic of the extension handled the collection of trackers, and analysis of the privacy policy, as well as injecting a script into the webpage to intercept cross-origin communications; the majority of this is carried out in a content script.

Cookies and local storage were obtained by making requests to document storage objects, and were added to an object which could be sent to the extension interface. The interface was sent this object once the user attempts to open it and sent updates if the object changes.

```
//Get cookies from host
function getCookies() {
 if (document.cookie.split(';').length > 0) {
 return document.cookie.split(';');
 } else {
 return [];
 }
}
//Get local storage from host
function getLocalStorage() {
 if (localStorage.length > 0) {
 return localStorage;
 } else {
 return [];
 }
}
```

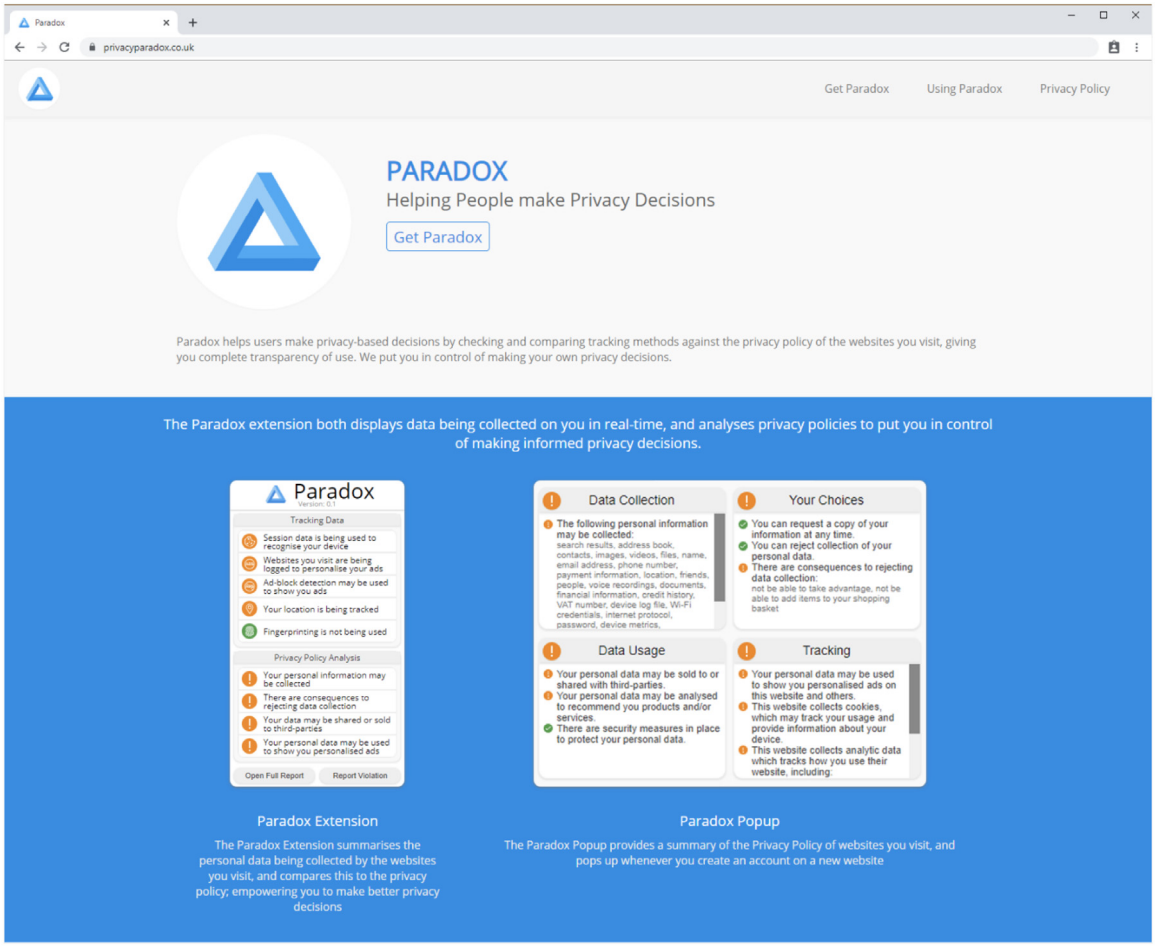

Fig. 9 – Paradox website homepage prototype.

The location of the privacy policy was determined by searching the webpage for the term 'privacy' within the sites anchor tags and retrieved the full address if found. A get request was performed on this address and the result (the privacy policy webpage as a string) was parsed.

To parse the policy, a search list of common terms was used to identify common themes (e.g. the string 'we share information about you' would be detected as third-party sharing) as indicated in Fig. 11. Once parsing was complete, an object was created. Each common theme had a true or false value, which defined whether that theme was found within the policy; where a theme was found, an additional parameter 'data' was returned containing all of the strings identified by the search list for this theme; these are shown in Fig. 12.

5.10. *Interface*

Once opened, the interface received the most recent results from the logic as an object. The interface then analysed the object and searched for keywords such as 'fingerprint'. The results of the analysis were then rendered; where a match is found, a warning and justification were displayed, and where no match is found, a green tick and explanation was shown.

```
const adsList = ['advert', 'advertisement'];
const adblockList = ['adblock', 'adblk'];
const locationList = ['location'];
const sessionList = ['session'];
const fingerprintList = ['analytic', 'fingerprint',
'browserwidth','browserheight', 'screenwidth',
'screenheight', 'wd='];
```

The privacy policy object was analysed and, where at least 1 warning is found (e.g. the policy makes no mention of personal data security measures), a warning was displayed to the user. The interface also analysed the privacy policy object comparatively with identified keywords to determine if there were any discrepancies between the policy and actual data collection, then – if found – displayed a warning to the user.

5.11. *Injected content*

The content injected a script into the webpage to intercept GET/POST requests, which could then be parsed by the interface. This must be done outside of the extensions as direct access to the website code is required to intercept and cap-

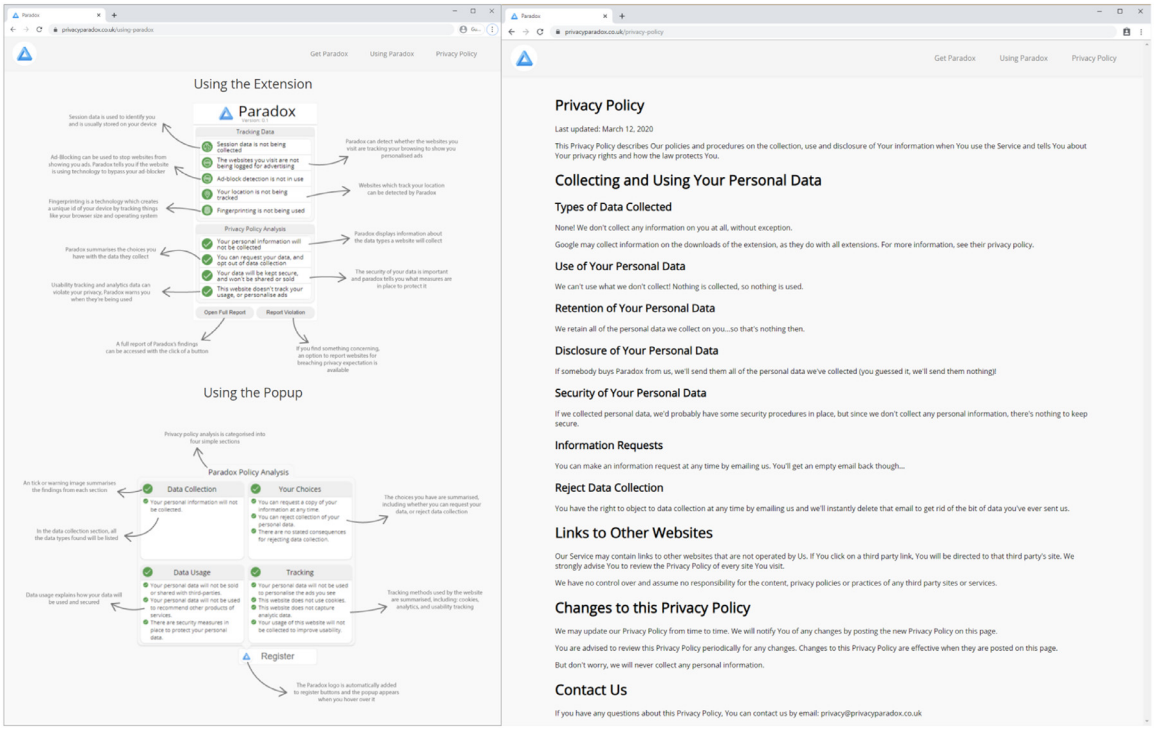

Fig. 10 – Paradox website prototype.

ture data in transmission. The script neither stopped nor prevented requests from taking place, it only captured them once they have been transmitted.

5.12. *Popup code implementation*

As the popup is rendered onto the webpage, it was implemented outside of the extension, and was handled by the same script intercepting GET/POST requests. The parsed privacy policy object was sent to the script from the logic, which was analysed, and relevant data was added to the popup. To display the popup, the script searched for a button element referencing the term 'register', and other similar terms. If found, the paradox logo was drawn on top of the button and could be hovered over to display the policy analysis.

5.13. *Requirements fulfilment*

The 'Extension Specifications' requirement in Section 4.1 expressed six requirements for the extension: 'Access, Analyse, and Present Privacy Policy' (R1); 'Intercept and Display Trackers' (R2); 'Detect Privacy Violations' (R3), 'Full Report of Findings' (R4); 'Report Privacy Violations' (R5); and 'Privacy Summary Popup' (R6).

In Fig. 11, a method for accessing and analysing the privacy policy is demonstrated. This shows that R1 has been satisfied.

Script injection was used to intercept trackers on active webpages. This data was then presented within the extension, thereby fulfilling R2.

Once the extension popup had received the privacy policy analysis from the content script, the results of the analysis were used to determine if any violations were prevalent. The code below demonstrates how a privacy violation can be detected:

```
if (result.location.cookies || result.location.cors ||
 result.location.storage) {
 const justification = 'location tracking despite
 claiming no personalinformation will be collected';
 if (!violationJustification.includes(justification)) {
 updateViolations(justification);
 }
}
```

Privacy violations were shown within the extension, and users could read full descriptions in the full report and privacy violation reporter, fulfilling R3. Two additional pages were created to display further, specific information and provide additional functionality in line with R4 and R5. These are the full report and privacy violation reporter pages. Finally, popup designs were created and implemented as specified by R6.

6. Evaluation

Formative evaluation was carried out on the extension designs from a usability perspective. This evaluation fed directly into the development of the extension, prior to any summative evaluation taking place. Summative evaluation of the extension took place in focus groups once feedback from formative evaluation has been quantified and incorporated.

A total of 41 participants took part in the design survey; 75% of the participants stated they were technically compe-

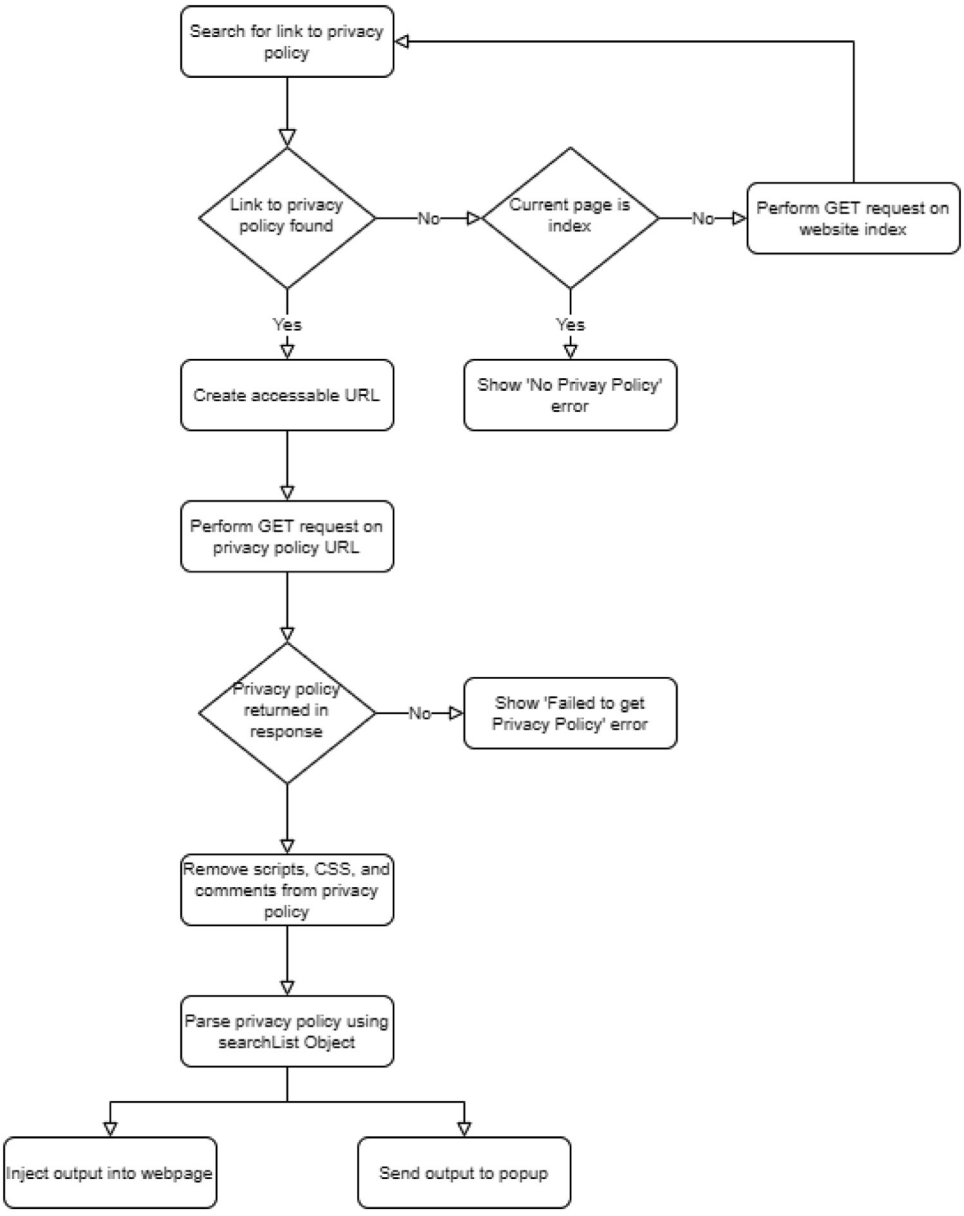

Fig. 11 – Find and parse privacy policy flow diagram.

tent. For the extension, Design 2 (Fig. 2) was determined to be the preferred design and for the popup, Design 1. (Fig. 4) was preferred overall.

6.1. *Formative design evaluation*

To ensure that the survey would both appeal and be relevant to a wide user base, the survey was designed to provide contextual understanding for the extension. Participants were provided with the purpose of the extension, where it would exist within the web browser interface, and how they could interact with it. The formative evaluation participants have been labelled from FP1 to FP41.

Designs for the full analysis report and privacy violation reporter were not created for the survey. This was to ensure a high quality of response by keeping the length of the survey as short as possible to maintain attention. The designs for these elements of the extension followed the design requirements elicited from the survey to maintain a fluent user experience.

The questions in the survey were mixed between direct and open questions; where a specific response was desired, a closed question was asked (e.g. 'Of these designs, which do you think is the most effective, efficient, and satisfactory?') and where open-minded thinking was desired, an open question was asked (e.g. 'If you could make this design better, what would you change?').

The survey was split into 4 sections: background information, extension designs (wireframe), extension designs (graphical), and popup designs.

The survey results can be downloaded from https://figshare.com/articles/dataset/Privacy_Extension_Design_Survey_csv/12652883.

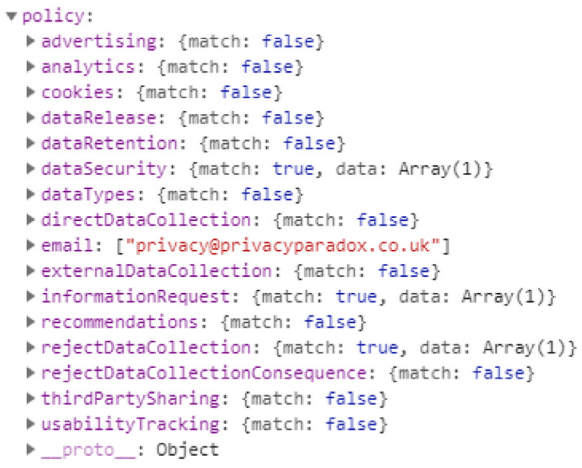

Fig. 12 – Privacy policy object.

6.1.1. *Background information*

The background information section explained the context of the Paradox extension to the participant and asked them whether they consider themselves to be technical or non-technical. This identified whether there are differences in design requirements for technical and non-technical users.

6.1.2. *Extension designs (Wireframe)*

In this section, participants were shown the three wireframe designs, and are asked if each is effective, efficient, and satisfactory. Participants were then asked which single design applies these three principles the best.

Open ended questions were asked to determine why their preferred design was better than the others, and what they would change to improve on it. These questions were designed to elicit thought-provoking responses which may present information that has not been considered during the design process.

6.1.3. *Extension designs (Graphical)*

This section focused on the colour scheme of the designs, and determined whether images were clear. For each design, participants were asked whether the colour scheme makes the content readable, and whether the images and their meanings weree clear.

Participants were then asked if their opinion of the most effective, efficient, and satisfactory design has changed, and asked the same open questions from Section 6.1.2.

6.1.4. *Popup designs*

In this section, participants were presented with 2 popup designs and their context for use. The questions asked were the same as those in Section 6.1.2, while referring to the popup designs instead.

6.1.5. *Findings*

Of the 41 participants, 10 (24.4%) claimed to be non-technical and 31 (75.6%) claimed to be technical (Fig. 13). Considering how ubiquitous web browsers has become, this result is not surprising. Nonetheless, the sample size from each group was still large enough to draw comparisons on design preferences between non-technical and technical users.

Participants were asked about the effectiveness, efficiency, and satisfaction of each design. Design 2 (Fig. 2) was rated as both the most effective and efficient but received the lowest rating for satisfaction, whereas Design 3 (Fig. 2) had the most satisfactory design (Fig. 14).

When asked for an overall favourite, the participants reflected the result of the previous question, with 43.9% choosing Design 2 (Fig. 2).

Participants reasoning for their chosen design was focused on understandability, e.g. "*Clearer to understand and a lot easier on the eyes*" [FP4], "*Simpler to understand*" [FP28], and "*The easiest to understand at first glance*" [FP19]. It was commented "*what does the box 'CORS' mean: Cross-origin resource sharing? Non-technical like me would not understand what this actually means. Design 2, although looks busy, actually identifies areas of concern* [FP27]".

Participants made the following suggestions: "*Perhaps a button to switch between 'basic' and 'detailed' information*" [FP30], "*Add information bubbles when the cursor is placed over each item*" [FP13], and "*Highlight what the images represent…Are they metaphors?*" [FP11].

When asked whether the colour schemes demonstrated in the graphical representations made it easier to read the content, Design 2 (Fig. 2) and Design 3 (Fig. 2) both received very positive results, with 37 out of the 41 participants agreeing in both cases. However, Design 1 (Fig. 2) received a lower score, with only 30 out of the 41 participants agreeing that the colour scheme improved the readability of the design (see Fig. 15).

Participants found that the images were far clearer and easier to understand in Design 2 (Fig. 2) when compared to the other designs; 36 of the 41 participants said that the images in this design were clear (see Fig. 16).

Participants conclusively decided that Design 2 (Fig. 2) was the most effective, efficient, and satisfactory design after seeing both the wireframe and graphical designs. In total, 56.1% of participants chose this as their preferred design (see Fig. 17).

Some participants explained that they were red/green colour-blind and that the second design was the only one which they could understand due to the explanations next to each icon. Other comments included: "*Maybe explain why it's in two sections by just like subheadings*" [FP19], "*Provide a graphical/textual distinction between "*good*" and "*bad*" - colour on its own is insufficient. I'd suggest you have a green tick or red cross superimposed over a grey icon to indicate success or failure - if you can't tell the difference without colour, it's a poor UI*" [FP39], "*Could the no privacy issues found change colour at some point so it's clearer*" [FP5], and "*Use amber for warnings not red. Red looks like you are already in big trouble on this page*" [FP2] (Fig. 18).

Participants believed that Design 1 (Fig. 4) was more efficient and satisfactory than Design 2 (Fig. 4). For effectiveness, the results for each design were very similar, with 31 and 32, respectively.

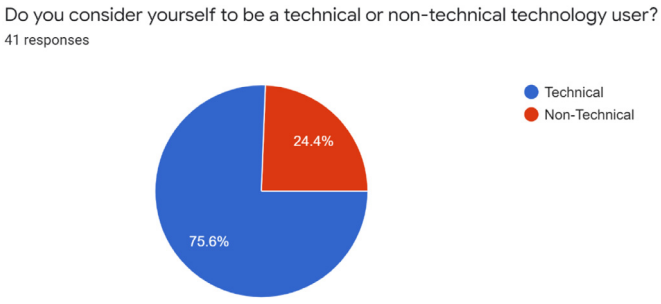

Fig. 13 – Technical and non-technical participants.

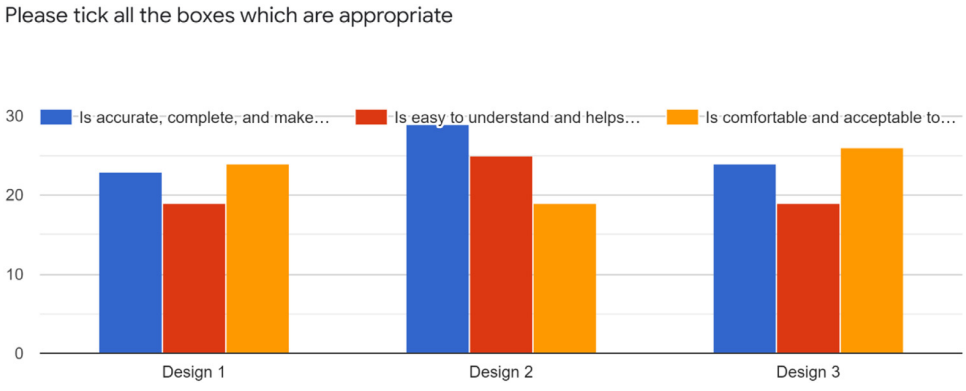

Fig. 14 – Effectiveness, efficiency, and satisfaction of wireframe extension designs.

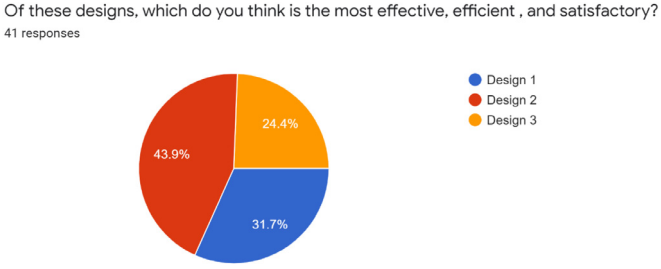

Fig. 15 – The most effective, efficient, and satisfactory wireframe extension design.

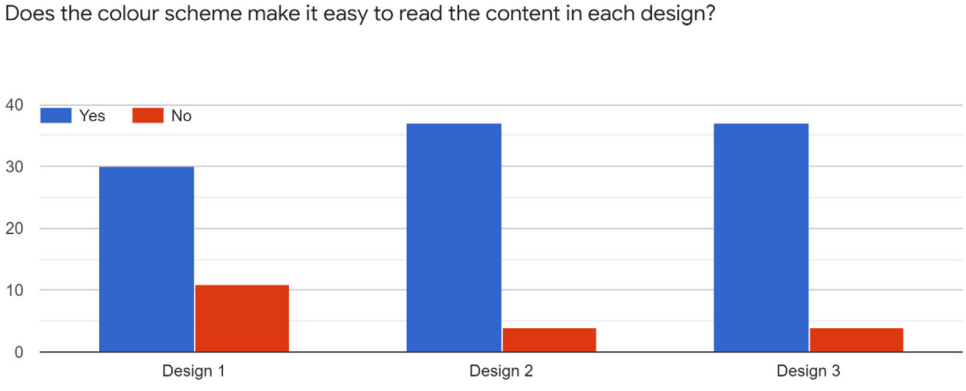

Fig. 16 – Effect of colour scheme on readability.

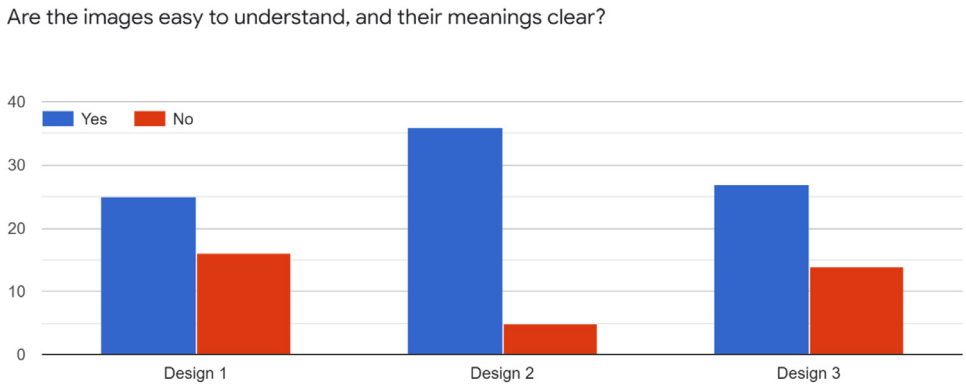

Fig. 17 – Image meaning understandability.

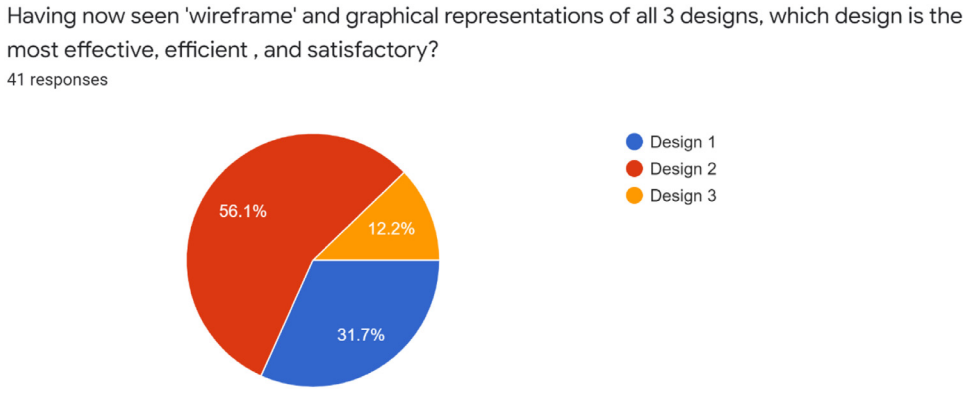

Fig. 18 – The most effective, efficient, and satisfactory extension design.

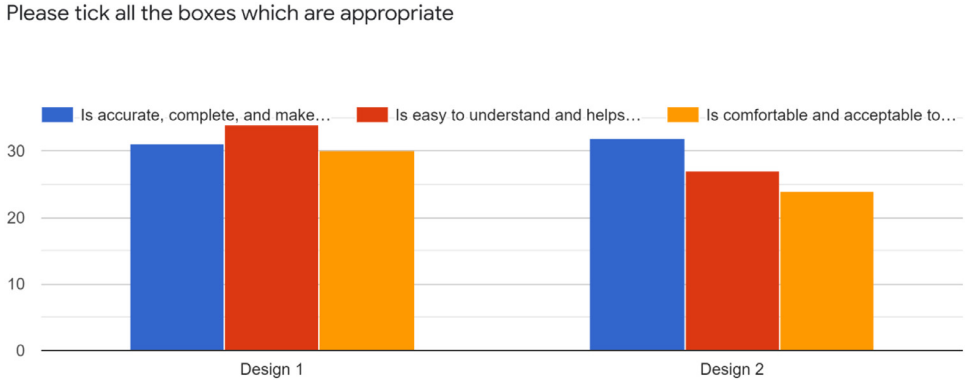

Fig. 19 – Effectiveness, efficiency, and satisfaction of popup designs.

The preferred popup design was Design 1 (Fig. 4). However, the result was very close, with 56.1% of participants preferring it (see Fig. 19).

Participants made the following comments on the popup designs: "*Maybe have a box that gives the software's over all opinion on its safety as a summary of the four boxes*" [FP19], and "*This is effectively giving users permission not to read the T&C document provided because you've summarised it for them. Does that make you responsible if there is a privacy breach if there is something in the T&C that you have told the user is ok? I know most people don't read the T&C anyway but be careful about taking that responsibility for them.*" [FP15] (Fig. 20).

6.2. *Summative evaluation - pilot*

A pilot of the focus group was carried out with a single participant. The aim of the pilot was to ensure the remote meeting presentation software was suitable and determine whether the structure and quality of the evaluation material was suitable for eliciting relevant and useful responses from the participants. The participant involved was of a technical background and was not involved in any further evaluation, nor have their responses been included in the results from the focus groups in Section 6.3.

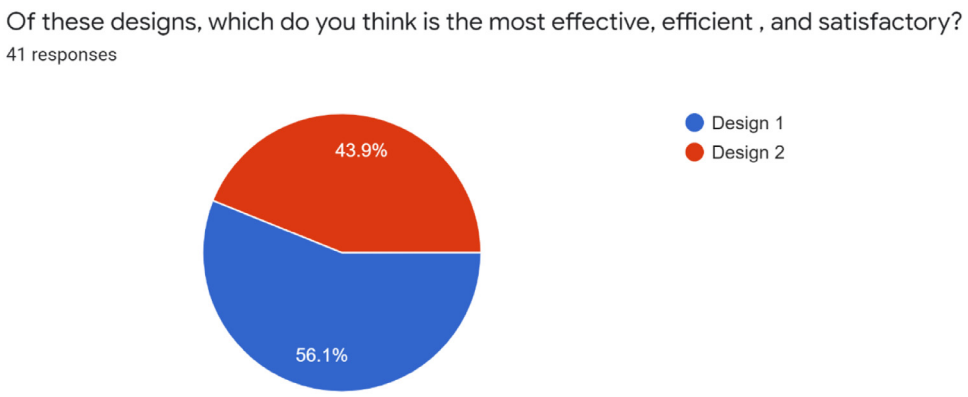

Fig. 20 – The most effective, efficient, and satisfactory popup design.

6.2.1. *Pilot findings*

The pilot determined several issues with the evaluation material, which were: Lack of Context, Misunderstanding Expectations, and Lack of Images.

6.2.2. *Lack of context*

The initial evaluation material provided very little context about the study, privacy, privacy policies, or tracking prior to introducing questions on those topics. The pilot participant found it difficult to understand or relate to some of the questions given that they had no prior knowledge of these topics. As a result of this, additional slides were added to introduce privacy, privacy policies, tracking, and provide context of the study itself.

6.2.3. *Misunderstanding expectations*

A very brief introduction to privacy expectations was given to the pilot participant. However – when asked what their privacy expectations were when using the web – they were unsure of how to answer the question. This was partly due to the lack of context [6.2.2], but also because only a limited amount of detail on privacy expectations was provided. This was improved by adding additional slides to cover privacy expectations in depth, and by talking about them as a group, rather than asking about them directly.

6.2.4. *Lack of images*

While the pilot participant claimed that the user interface of the Paradox extension was simple and easy to pick up, they found it difficult to relate the information they were being given about the extension with it at the time. The participant explained that having images of the extension and talking about each section directly would clear up any potential misunderstandings. As a result of this, images of the extension were added to the presentation slides and were delivered with a direct explanation during the focus groups.

6.2.5. *Pilot observations*

During the pilot, the participant showed concern over some of the data types being collected and was particularly concerned that all the websites they visited claimed to share personal data with third parties. In contrast, the participant was not concerned about analytic data collection or any given consequences to rejecting data collection, such as preventing a user from adding an item to their basket on an e-commerce website if they reject data collection, although they did suggest that the latter was morally wrong to do.

Interestingly, the participant suggested that the Paradox extension had changed their opinion of not only the sites they visited, but other sites as well. Their reasoning was that now they were aware of the ways the tested websites were using their personal data, they expected that similar websites would do the same thing. As a result, the participant said their expectations have changed after using the Paradox extension, but not just for the websites visited while using the extension.

Based on these findings an additional question, covering changes in expectations for websites which were not visited during the focus group, was added. This was to determine whether a similar pattern would be repeatable among participants. In addition to this, the format of the evaluation was changed whereby participants were offered the opportunity to participate in a follow-up interview a week after their focus group; participants were made aware that this was optional.

6.3. *Summative evaluation - focus groups*

Due to the emergence of COVID-19 in Spring 2020, focus groups had to be rearranged to take place remotely. Rather than meeting in-person, remote meeting presentation software was used to connect participants. The responses from participants were reviewed through an informal thematic analysis based on an affinity diagram style feedback-grid (IBM Studios, 2016).

6.3.1. *Focus group participants*

Four focus groups were conducted, each with 5 participants (Groups A to D). Small groups were used to ensure that everybody was able to speak and provide quality data while adhering to a reasonable timeframe. Two of the groups were made up of non-technical participants, while the other two groups were made up of technical participants.

Participants were recruited primarily based on their technical ability, and were recruited via their educational work or current employer; the technicality of participants was determined based on a combination of a self-assessment by each

Table 4 – Focus group participant numbers.

Technical		Non-Technical	
Group A SP1, SP2, SP3, SP4, SP5, SP6	Group B SP7, SP8, SP9, SP10	Group C SP11, SP12, SP13, SP14	Group D SP15, SP16, SP17, SP18, SP19

participant and the consideration of their expertise, including their current and previous employment roles, and prior education. The demographic information of participants was considered with a balance of participants from each bracket (such as age and gender) in each group.

All four focus groups were successfully carried out, taking between 50 min and 1 h 7 min. Of the 20 initial participants, one [non-technical participant] was unable to take part due to technical difficulties when joining via the online meeting software. A further participant was unable to participate during their allocated time slot, but participated in a later group instead.

A PowerPoint presentation was created as a visual aid for participants during the focus group. The slides were displayed via the meeting presentation software to participants during the focus groups.

6.3.2. *Participant information*

The summative evaluation participants were labelled from SP1 to SP19, where SP1 to SP10 were technical participants and SP11 to SP19 were non-technical participants (Table 4).

6.3.3. *User expectations*

Participants were asked about their privacy expectations when using the web. These expectations were broken down into the four categories specified in the 'Access, Analyse, and Present Privacy Policy' requirement.

Almost all the participants made it clear that they had not considered their privacy expectations before, and several stated that they "*don't have any*". Only once they were introduced to different aspects of their privacy exposure, such as the personal information being collected on them, did they begin to consider what their expectations might be.

6.3.4. *Personal information*

Across the focus groups, both non-technical and technical participants were generally quite restrictive about the personal data which they wanted to be collected. Around half of the participants accepted that some basic information such as their name, age, and gender could be collected freely. The other half stipulated that without specific consent and purpose, no personal information should be collected at all, e.g. "*I would expect the basics*" [SP8], "*I have a very negative opinion of personal data collection completely, I always assume the worst although I don't know what the worst actually is*" [SP9], and "*I wouldn't expect any of that [Personal] information to be taken unless it was needed for a specific purpose*" [SP6].

6.3.5. *Device information*

Almost all of the participants across the focus groups suggested that they were happy for basic information such as device type, browser type and IP address to be collected, as long as it was for the purpose of improving their user experience [by rendering content which is suitable for their device]; "*I would [expect this] if it makes a difference to how the page is being rendered*" [SP6]. A small number of participants said that they did not expect device information to be gathered at all, e.g. "*It's something I hadn't even thought about which I now have to worry about even more*" [SP9] and "*I can't see what device information has to do with them*" [SP10].

6.3.6. *Financial information*

All the participants agreed that financial information should not be collected unless their permission had been explicitly given, and then destroyed immediately afterwards. They also stated that this information should only extend to the minimum amount required to complete the relevant transaction; "*[websites should only collect] the minimum amount of information they need to make a transaction*" [SP19]. Some participants also suggested that financial information should be instantly destroyed after a transaction by default, e.g. "*I personally don't like websites keeping that data*" [SP5] and"*it should be destroyed afterwards*" [SP12]. Contextually, all participants agreed that a website which has no intention of selling you something should not be collecting this type of information, while websites which were being used to sell products/services should be allowed this information temporarily, and only when provided directly by them.

6.3.7. *Information on others*

In general, the participants all agreed that social media websites obtaining information about their friends, family and others was acceptable, but stated that other types of websites should have no justification for collecting this type of information and therefore shouldn't be collecting it at all, e.g. "*I wouldn't have really expected that unless it was a website like Facebook*" [SP9] and "*It's down to the website*" [SP19]. One focus group had a slightly different opinion, collectively stating that this type of information should "*absolutely not*" [SP11] be collected and that it was "*none of their business*" [SP12].

6.3.8. *Selling or sharing with third parties*

Both non-technical focus groups made it clear that they were entirely unhappy with the concept of third party sharing and did not want it to happen unless they had explicitly agreed to it, e.g. "*I don't want them to [sell or share my information]*" [SP17].

In contrast, both technical groups made it clear that they were aware that third party sharing was a common practice and something they had become used to, thereby suggesting that they had come to expect it, e.g. "*I assume they're always selling and sharing my information with third parties*" [SP9] and "*Third party sharing is just something that happens*" [SP7]. Despite this some of the technical participants made it clear that

they were not happy with third party sharing despite accepting and expecting it, e.g. "*it's really annoying*" [SP5]. Members of the technical groups tended to show more concern about who the third parties were and what data they had rather than concern for the practice of third party sharing itself, e.g. "*You'd probably have to do quite a bit of digging to find out the specific companies [whom your data is being shared with]*" [SP8].

6.3.9. Recommending products/services

Typically, participants accepted the usage of their data to recommend them products/services. Some participants suggested that it can be a good thing as long the recommendations felt suitable and were relevant, e.g. "*it's a business tactic to get you to buy more, and it is useful sometimes*" [SP10] and "*when it's being used appropriately, it's useful*" [SP4].

SP17 – who was very unhappy about this type of data usage – said "*if I buy a present for my wife on the internet and she uses the same laptop, she'll see what I'm buying her*". This demonstrates a way in which this type of data usage is not only invasive for some people, but also undesirable.

6.3.10. Data security

Technical participants were concerned with their personal data security in general and suggested that they assumed every website they visit is insecure. Despite this, they said that their expectations were that relevant laws are followed to keep personal data secure, while admitting that they were not too sure if this happened in practice, e.g. "*There's a lot of laws and regulations in place and you kind of hope that they're being followed, but you don't really know. There's not much [clarity] on websites I've seen, I just hope it's okay*" [SP7] and "*I just assume everything is insecure*" [SP10]. SP1 and SP3 suggested that while personal data security measures should be transparent to users, too much information could compromise security and that finding a balance was key, e.g. "*a brief overview should be provided, but nothing which can compromise themselves*" [SP3].

Non-technical participants explained that they wanted it to be clearer to them that their data was going to be kept secure, e.g. "*I think websites should be telling you what they're doing*" [SP17]. However, they said that they often trusted websites without any justification and would not read about their personal data security even if it was made transparent to them, e.g. "*I think it should be on there somewhere, but I think you should have to go looking for it. If I go on to a website, I'm not going to read everything they do to protect my security before I actually do what I want to do*" [SP18]. SP16 suggested, "*In a nicer world [I would assume that websites take steps to make security measures transparent], but they probably don't*".

6.3.11. Requesting a copy of personal information

All the participants agreed that they should be able to request a copy of their personal information though some said that they had never thought about or considered doing so and therefore had no real expectation towards the capability of doing it, e.g. "*You should always be able to view your personal information*" SP7 and "*I don't even think about it to be honest*" [SP11]. Some participants also suggested that the option should be made clearer or that they should simply be able to access a copy of all their personal information from a website without having to make a request for it, e.g. "*I want to be able to see it, but I don't want to have to request it*" [SP13].

6.3.12. Rejecting data collection

All the participants agreed that they should have the choice to reject data collection, and that the capability to do so should be made clearer or easier to access, e.g. "*I think you should be able to [request a copy of your personal information] but you don't seem to be able to*" [SP9] and "*I don't know how you would go about doing that*" [SP7]. Many of the participants from non-technical and technical backgrounds suggested that rejecting data collection might have undesirable consequences on the websites, or that they simply would not be able to use them anymore. SP3 went further and suggested that websites which allow users to reject data collection should have the choice to turn users away if they decided to do so; if the business cannot make money without collecting your personal information, the participant considered this a valid reason to turn users away. SP6 agreed with this: "*it depends on the business model of the website, which should be made clear*".

6.3.13. Personalised advertising

Most participants were content with personalised advertising trackers, with some suggesting that it can often be useful; SP7 agreed, saying, "*I think personalised advertising is a good thing*". SP10 commented that they did not like it when they were shown "*irrelevant adverts*" and SP8 agreed, saying "*sometimes when I search for one thing, I get lots of ads I don't want*" while SP5 suggested they did not like that this type of tracking, which could be used to increase prices on items they had recently searched for on other websites.

6.3.14. Cookies

All the participants accepted cookies, and understood that they could be useful. The participants made it clear that they were only comfortable with cookies if they were only collecting relevant data which would not be shared or sold e.g. "*It's quite useful to go back to a website and have what you were doing resume from the time before*" [SP6] and "*I don't like cookies that track you across various websites*" [SP17]; SP17 added that they did not mind cookies which were used to restore website sessions.

One of the non-technical focus groups expressed concerns that cookies affect the performance of their device, e.g. "*they clog up my computer and I have to clear them all the time*" [SP12]. This behaviour suggests that the security and privacy concerns associated with cookies are not necessarily considered, and – in some cases – performance concerns take precedence.

6.3.15. Analytics

Generally, all the participants agreed that they were comfortable with analytics being used, e.g. "*it can be a good thing, but it also depends on what they pull out while they're doing it*" [SP9]. However, SP11 described their usage as being "*nosey*". SP9 was unaware that analytic data was collected at all but considered that it would not have a negative impact on their personal user experience; SP7 added "*I didn't know mouse tracking was a thing, that's interesting to know*".

6.3.16. Usability tracking

All the participants were content with usability tracking taking place, with many suggesting that it was a good thing and beneficial for them in the long-term, e.g. "*that is a good thing, absolutely*" [SP9] and "*it'll help to make the website more accessible*" [SP15]. As such, the participants collectively decided that they both expected and wanted websites to use usability tracking.

6.3.17. Expectations on Amazon

Participants were asked whether their expectations were different for Amazon; the results were mixed.

One of the non-technical groups was more concerned about Amazon due to the number of services and devices they offer. In contrast. the other focus group collectively suggested that they were more trusting of Amazon compared with other websites, and that they would therefore be more lenient with them if any of their expectations were not met. Both non-technical focus groups made it clear that they still had the same expectations, despite admitting that they expected the reality to differ from their expectations.

Of the two technical groups, one collectively suggested that they expected Amazon to take a larger amount of financial information but that otherwise their expectations remained the same. The other technical group said that their expectations were the same, and SP4 went further, saying that "*I expect them to be better as I see them as role models. They should be demonstrating appropriate use of data [as] they have a responsibility*".

6.3.18. Expectations on Facebook

Participants were asked whether their expectations were different for Facebook. Across all the groups, everyone believed that Facebook would collect more information and do more with it.

The non-technical focus groups both said that they expected Facebook to be less secure than other websites. One of these groups also agreed that they had lower expectations of Facebook overall with both SP15 and SP16 admitting that they did not use Facebook at all as a result of this.

In the technical groups, SP9 said they did not use Facebook due to concerns over how their data would be used. One user also reiterated his earlier statement, suggesting that like Amazon, Facebook is a role model and that their behaviour should demonstrate this.

6.3.19. Paradox results on Amazon

Participants were asked to use the Paradox extension on Amazon and consider the results from the full report feature, the results are listed in Table 5.

6.3.20. Paradox results on Facebook

Participants were asked to use the Paradox extension on Facebook and look at the results from the full report feature (Table 6).

6.3.21. Extent of expectations met

Based on the results from the Paradox extension Results on Amazon and Facebook, a table summarising the areas in which expectations were and were not met was produced; cells where at least 50% of participants expectations were not met have been highlighted in red (Table 7). The 'combined' cells calculate the combined number of non-technical and technical users whose expectations were met.

The table indicates some interesting patterns, including: Disconnected Expectations of Data Types, Amazon met fewer Expectations than Facebook, and Differences in Third Party Sharing Expectations.

6.3.22. Disconnected expectations of data types

Across every group, 'Data Types' was the only category to not meet a single participants expectation on both Amazon and Facebook. None of the participants believed that the data being collected on them met their expectations and every participant was shocked or surprised at the amount of data being collected, and how intrusive some of it was. Many participants were also concerned about how unnecessary much of the data appeared to be.

6.3.23. Amazon met fewer Expectations than Facebook

Overall, Amazon met user expectations 67% of the time while Facebook met user expectations 84% of the time. Amazon scored particularly badly with the non-technical groups; whose expectations were only met 62% of the time.

When compared to Facebook, users were particularly unhappy with Amazon's policy on releasing personal data; no participants said that this met their expectations. In contrast, Facebook said that it did not release any personal information.

Amazon also fell short of user's expectations for 'Data Retention' with only 16% of users stating that this met their expectations compared with 100% on Facebook. Furthermore, participants felt their expectations were not being met by Amazon's 'External Data Collection', with it only meeting expectations 32% of the time across non-technical and technical users.

Finally, non-technical users were unhappy that Amazon would prevent them from using core services if they chose to reject personal data collection; this met only 22% of non-technical user's expectations.

A combination of the above factors may contribute to explaining why Facebook received a much higher score than Amazon. In addition to these factors, it is worth considering that there may be slight differences in expectations as described in 'Expectations on Amazon' and 'Expectations on Facebook'.

6.3.24. Differences in third party sharing expectations

The results show that non-technical users had higher expectations for third party sharing. For Amazon, 0% of non-technical participants said their expectations had been met, while 60% of technical participants expectations were met. A similar pattern was visible on Facebook where only 22% of non-technical users said their expectations towards 'Third Party Sharing' had been met, but 100% of technical users agreed that their expectations were met.

This demonstrates a clear difference in expectations, and shows that the non-technical participants were far less accepting of third party sharing than the technical participants.

6.3.25. Violation reporting

Participants were asked to browse the web and look at a few websites with the Paradox extension to determine whether

Table 5 – Evaluation results on Amazon.

Evaluation	Participant Response
On Amazon, the Paradox extension detected different numbers of trackers for each participant. As a collective, there were typically 30–800 trackers detected; SP18 however, found 232 trackers detected by the Paradox extension, with 221 of these in Local Storage.	The non-technical groups were generally very concerned by the number of trackers in use with SP17 asking, "*Why do they need so many?*" and "*What are they tracking?*". Collectively the participants agreed that there were more trackers in use than they thought there would be, although SP11 and SP13 admitted that they "*didn't know what to expect*". One of the technical groups said that despite there being more trackers than expected, they were not particularly surprised. The other technical group agreed that they had expected large number of trackers and that they were not surprised with the results.
The Paradox extension detected that Amazon was using third party advertising, advertising, and interest-based ads.	All four groups collectively agreed that they both expected this, and that they were comfortable with it.
The Paradox extension detected that Amazon was using cookies and unique identifiers on its website.	All four groups agreed that they expected this, with SP19 suggesting that cookie usage was "*pretty normal*". However, SP17 added that they were not necessarily comfortable with it, using the word "*unfortunately*".
The Paradox extension detected that Amazon releases account information and exchanges information with others. Data release is the practice of releasing information to a company or individual without notice or consent.	None of the participants (technical or non-technical) were happy about this and it did not meet their expectations, e.g. "*A bit too ambiguous*" [SP12] and "*I'm not happy with that*" [SP11]. They were especially concerned with the ambiguity and that it was not clear what data could be released and SP7 suggested that "*they're very vague with the wording, it's very illusive*". SP6 said they were "*very worried*" while many others said they were "*not happy*".
The Paradox extension found that Amazon retains personal data for 'as long as is required'.	All the participants said that they did not mind their data being retained. However, they all expressed concern over the ambiguity in the term 'as long as it is required', e.g. "*as long as it is required' is very vague*" [SP9] and "*it's just a bit vague*" [SP2]. SP17 questioned whether the data was being held for as long as required by Amazon, or if it was held for as long as required by the user, asking,"*who's definition of 'as long as required'?*"; SP15 added, "*required by you or required by them?*". Participants suggested that they were uncomfortable with the idea that Amazon could retain their personal information for as long as Amazon deemed to require it; there were suggestions that the specific criteria for retaining this information should be made transparent.
The Paradox extension detected that Amazon has a range of security measures in place to protect personal information.	All the participants said that they were happy with this result and that this had met their expectations, e.g. "*it's all positive by the indicators - the green ticks*" [SP17]. None of the participants indicated that Amazon's policy on data retention had changed their opinion on data security. Further to this, participants did not seem to consider the fact that their data could be released by Amazon, thereby making data security effectively redundant.
The Paradox extension detected that Amazon was collecting information which was provided by the user.	All the participants said that this met their expectations and that they were happy with this, assuming the information being requested was the minimum required for its purpose. SP3 was slightly more specific and stated that they were happy with this "*depending on the information*" which was being collected.
External data collection was detected on Amazon by the Paradox extension. This consisted of information which was provided to Amazon from other sources.	The non-technical groups both expressed that they were displeased with this, e.g. "*this makes me uncomfortable*" [SP15], "*did not expect that*" [SP12], and "*why would they do that?*" [SP11]. SP19 added, "*It means you don't really know who's got what data on you*". However, SP3 said that "*it depends on the type of data*" and suggested that they would be happy with some specific acceptable use of external data collection. The technical groups both agreed that their expectations were based on the context of data being provided from other sources. They agreed that in some cases external data collection can be necessary, but suggested they were worried about how external data collection would be used on Amazon, e.g. "*when it comes to the financial side they have to do that don't they? I'm not sure I'm happy with other things but there's things you expect there and things you might not expect depending on what kind of things it is they're getting from other sources*" [SP9].
Amazon states clearly in its policy that users can make an information request, as is expected under GDPR legislation.	Across all four focus groups, the participants agreed that this met their expectations, e.g. "*you should be able to*" [SP16].
The Paradox extension detected that Amazon used the personal data it collects to recommend features and personalise user experience when using their website.	Three of the four focus groups were happy with this as they felt it met their expectations, e.g. "*I'd rather see ads that are relevant if I have to see ads*" [SP18]. However, in one of the non-technical focus groups there was a mixed response. Generally, this group believed that recommendations were okay sometimes but usually found them intrusive.

(*continued on next page*)

Table 5 (*continued*)

Evaluation	Participant Response
Amazon provides its users with the capability to reject data collection when using its services. Users can choose not to provide their personal information, withdraw their consent, or opt out, and object to the processing of their personal information.	Both non-technical groups were pleased with this and stated that it met their expectations. In the technical groups there was a slightly more varied response; while all the participants were pleased, SP9 said that they had not expected Amazon to do this. SP3 suggested that while this was a good thing, the option to reject data collection was not obvious enough and that users should be made more aware of how to access this capability and added, "*it's only good if it's easy to find*".
At Amazon, the Paradox extension detected that there were consequences to rejecting data collection. These consequences included "*not being able to take advantage*" of Amazons services, and "*not being able to add items to your shopping basket*". In effect, these consequences are designed to force people to accept data collection if they want to use Amazon.	In the non-technical focus groups, only 22% of the participants expected that there would be a consequence to rejecting data collection. Despite this, none of the participants were happy that they would not be able to use Amazon without accepting data collection; the groups both labelled it as "*morally wrong*" and SP16 said "*If you reject it you're stuffed then aren't you?*". In the technical groups, 40% of the participants were unhappy with this and SP9 suggested that "*we can't use it then*". The other participants agreed that Amazon was within their right to do this and SP1 reasoned that "*[Amazon] needs your data to check out, so this could be a functional requirement*".
The Paradox extension detected that Amazon communicates with third-partied in a multitude of ways. Amazon states in its privacy policy that it might 'sell or buy' or 'share' information and states that 'third parties are involved in your transactions'.	In the non-technical focus groups, all of the participants were notably shocked by this having previously stated in the 'Selling or Sharing with Third Parties; expectation that they did not expect third party sharing to occur on any website without explicit consent, e.g. "*[third party sharing] certainly doesn't meet expectations, I didn't think Amazon would go quite that far*" [SP17] and "*I'm not happy with this, it's too far*" [SP11]. In contrast, both technical groups unanimously admitted that they expected third party sharing to occur, e.g. "*other parties sell through Amazon, so they need to know your personal information*" [SP5]. SP1 did express concern that they were not comfortable not knowing who the third parties were.
The Paradox extension detected that Amazon uses usability tracking on its website.	All the participants agreed that the presence of usability tracking met their expectations and many of them suggested that they were happy it was being used, e.g. "*you can understand that*" [SP16] and "*I'm happy it's there*" [SP10].
The Paradox extension detected that analytics were being used on Amazon. This included the use of mouse tracking, click detection, and scroll detection.	Typically, the participants either expected the presence of analytics or did not mind that it was there, e.g. "*you would expect that to be there, they want to know what you're browsing and when*" [SP16]. SP11 mentioned that there were more tracking types than they had expected but that they were still comfortable with their usage.
The Paradox extension detected 27 different types of data which Amazon said it could collect. This information varied from basic personal information such as an email address, to much more intrusive data such as a user's credit history, VAT number, and WiFi password.	No single participant was happy with the volume or types of data being collected at Amazon. In the non-technical groups, participants agreed that the data being collected was "*too much*" and that it "*did not meet expectations*". Participants were clearly shocked with one branding the data types as "*intrusive*" and claimed that there was "*far too much data collection*", e.g. "*half of those I'd like them to take off*" [SP16]. In the technical groups, participants were equally concerned. Participants said that they were "*worried*", and "*scared*"; SP10 said that there was "*lots of stuff they don't need*" and that they were "*not happy, especially if this is being sold on*" while SP9 said, "*There's a lot more than I expected*". Overall, the participants agreed that their expectations had not been met "*at all*" and some called for Amazon to provide justification for what this data was being used for.

they could find any privacy violations. All but one of the participants found at least one violation when they were free to browse the web with the Paradox extension.

All the non-technical participants agreed that they would consider reporting a violation using the Paradox extension's violation reporter if they found one. However, SP18 said that it would depend on their opinion towards the severity of the violation and SP16 questioned whether a report would be taken seriously by a company like Amazon or Facebook.

All the technical participants – bar SP9 – suggested that they would consider reporting a violation if they found one. SP9 – who said that they would not consider reporting a violation using the Paradox extension – explained that they were concerned that reporting a website for a privacy violation would expose them to that website, and were concerned that those websites may then treat them differently.

6.3.26. *Privacy awareness*

Participants were asked whether they believed that the Paradox extension had made them more aware of their personal privacy. Every participant agreed that they were more aware after using the Paradox extension.

In the non-technical groups, participants explained that they were more aware, and that their expectations towards websites had been changed for the worse, e.g. "*I think more about the way I behave online now*" [SP17]. When asked whether

Table 6 – Evaluation results on Facebook.

Evaluation	Participant Response
On Facebook participants recorded between 5–50 trackers.	All the participants agreed that the total number of trackers they had on Facebook was typically large. However, most participants expressed that they were not particularly concerned about this and that it met their expectations. In one of the technical focus groups there was a slightly more mixed feel, with some participants suggesting that they were a "*little worried*" at the high number of trackers Facebook was storing on their devices; SP13 indicated that they had "*loads more*" after around 10 s had passed and found this especially concerning.
Facebook use personalised advertising and show sponsored content when you browse their website	All the participants agreed that advertising on Facebook was expected and that they were comfortable with it.
The Paradox extension detected the usage of cookies on Facebook for the purpose of uniquely identifying users	All the participants said that Facebooks usage of cookies met their expectations.
The Paradox extension found that Facebook - unlike Amazon - does not suggest that it may release personal information.	All the participants said they were pleased about this, but also admitted that this exceeded their expectations, e.g. "*that's quite good actually*" [SP2] and "*from experience, Facebook don't even give data to law enforcement either, which is interesting*" [SP5]. SP11 suggested that they were "*surprised*" while SP17 asked "*has this changed recently?*" and added "*that's precisely how it should be*". This pattern was present across technical and non-technical groups.
The Paradox extension detected that Facebook retains personal information 'until your account is deleted'.	All the participants were content with this and suggested that their expectations had been met, e.g. "*I suppose you would expect that*" [SP16]. However, SP13 queried "*When is no longer necessary?*" and suggested that this could allow Facebook to store personal information after an account is deleted. SP11 also questioned whether data really is deleted after Facebook claims that it has been.
The Paradox extension detected that Facebook was using TLS security measures as a method of protecting personal data. While significantly less security features were discussed in Facebooks privacy policy in comparison with Amazon, their presence was not omitted entirely.	Almost all the participants expressed concern that there seemed to be a lack of transparency about personal data security in Facebook's privacy policy. In one of the non-technical groups, expectations were generally met; this group suggested that "*lots of the data [on Facebook] is public, so they might not care [about security] as much*". In the other non-technical group, only SP14 said that their expectations had been met, adding that "*we all know how Facebook are a little bit shady with what they do with data, so I'm not surprised [that there is less here], but I'm disappointed*". The other participants agreed that they were not happy and were displeased that there was not more information present. In the technical groups, participants were concerned with the lack of transparency and were worried about certain technologies which were not listed, e.g. "*worried about [the] security of my data as there is so much [of it]*" [SP7]. SP4 added that "*when we were looking at Amazon wasn't there a long list of ways data was being secured? It would be interesting to see what things are not being used to secure our data on Facebook*".
The Paradox extension detected that direct data collection methods were being used on Facebook, including – interestingly – face recognition technology.	All the groups expressed that they were very concerned about the usage of facial recognition technology in obtaining personal identifiable information about them. Across the non-technical groups, most of the participants suggested that they did not mind or expected this from Facebook, but they did brand it as "*creepy*" and "*scary*" and SP11 asked "*why would they want to do that?*". SP17 mentioned that they were "*very, very concerned with that*" and spoke about a similar attempt to use facial recognition technology in Russia via the public app FindFace in 2016 which could be used to discover peoples social media profiles by taking a picture of them. In response, SP18 suggested "*everyone's phone does it anyway, I kind of expect it*". In the technical groups, one group was very concerned about the usage of facial recognition and suggested that their expectations had "*not at all*" been met, e.g. "*that's quite worrying, it doesn't meet my expectations at all*" [SP8]. In the other technical group, there were still some concerns about the usage of facial recognition but overall, these participants were not too worried and considered that their expectations had been met.
Facebook was detected using external data collection by the Paradox extension. This included information that was provided by other people, and information provided by partners.	All the participants agreed that this met their expectations, e.g. "*partners like Instagram I think they share with which is okay, I expect it*" [SP3]. However, SP6 replied that "*this assumes partners are somehow related to the website, but they might actually not be*".
The Paradox extension detected that Facebook offers its users the capability to make an information request from its website.	All the participants agreed that this had met their expectations; SP16 added that they "*absolutely* " expected this.
The Paradox extension detected that Facebook was using personal information to "*make suggestions*" and "*give you tips*".	All the participants agreed that this met their expectations. SP19 participant suggested that they were "*happy*" with this usage of their personal information and SP7 said "*I don't mind it, it's a given*".
The Paradox extension determined that Facebook provide their users with an option to reject data collection.	Across the groups, participants seemed surprised that they were able to reject data collection on Facebook. Multiple participants suggested that they were "*happy*" while others said that this was "*good*".

Table 6 (*continued*)

Evaluation	Participant Response
Unlike Amazon, the Paradox extension did not detect any consequences for rejecting data collection at Facebook.	None of the participants expected that Facebook would not suggest any consequence to rejecting personal data collection; SP3 said that they were "*somewhat surprised to be honest*". SP16 said that they were "*pleased*" while others expressed that this had "*exceeded expectations*".
The Paradox extension detected that Facebook uses third party sharing to share information about its users.	In the non-technical groups, only SP11 and SP18 suggested that their expectations had been met; "*they've got to have some way of paying their employees*" [SP18]. The other participants were not happy with their information being shared and SP17 suggested that it made them feel "*uncomfortable*". In contrast, both technical groups considered this behaviour normal and said that it met their expectations, e.g. "*it's expected*" [SP2].
The Paradox extension detected that usability tracking was being used on Facebook for the purpose of improving their products.	All the participants agreed that this met their expectations and many participants reiterated that they believed usability tracking was a positive thing which could improve their user experience, e.g. "*that's fair enough*" [SP10].
The Paradox extension detected many ways in which Facebook uses analytics to track how its users browse their website. This included some less specific information such as: 'how you use features' and 'actions you take'.	All the participants said that the analytics being collected by Facebook met their expectations, e.g. "*that's okay really, it's expected*" [SP10]. However, SP11 did state that some of them were "*a bit strange*".
The Paradox extension detected 35 different types of data which Facebook can collect on its users. This included some generic data types such as an email address, but also included some irregular data types such as: call log, health information, battery level, Bluetooth signals, and places you like to go.	None of the participants said that the data types being collected met their expectations. In the non-technical groups, participants were very unhappy with the data that was being collected. The participants in these groups made a variety of comments about the data types, including: "*that's crazy*", "*it's none of their business*", "*way over the top*", "*very personal*", and "*unethical*". SP13 added "*that's too much, I didn't expect that*" and SP16 said "*that's way over the top, it's far too much information that they're gathering*". The technical groups suggested that they already had low expectations of Facebook but SP7 said, "*this is still much worse [than I expected]*". SP1 called for a justification on why all this data was required, saying, "*some of this is expected but some of it I would like to see justified*". SP3 said, "*there are some very odd things here. I don't want a lot of this*" while SP4 added "*I'm curious to know how it gets all of this information*".

Table 7 – Number of participants whose expectations were met.

	Amazon		Facebook		
	Non-Tech	Tech	Non-Tech	Tech	Total
Advertising	9 (100%)	10 (100%)	9 (100%)	10 (100%)	38 (100%)
Cookies	9 (100%)	10 (100%)	9 (100%)	10 (100%)	38 (100%)
Data Release	0 (0%)	0 (0%)	9 (100%)	10 (100%)	19 (50%)
Data Retention	3 (33%)	0 (0%)	9 (100%)	10 (100%)	22 (58%)
Data Security	9 (100%)	10 (100%)	6 (67%)	2 (20%)	27 (71%)
Direct Data Collection	9 (100%)	10 (100%)	5 (56%)	5 (50%)	29 (76%)
External Data Collection	1 (11%)	5 (50%)	9 (100%)	9 (90%)	24 (63%)
Information Request	9 (100%)	10 (100%)	9 (100%)	10 (100%)	38 (100%)
Recommendations	6 (67%)	10 (100%)	9 (100%)	10 (100%)	35 (92%)
Reject Data Collection	9 (100%)	10 (100%)	9 (100%)	10 (100%)	38 (100%)
Consequences to Rejecting Data Collection	2 (22%)	6 (60%)	9 (100%)	10 (100%)	27 (71%)
Third Party Sharing	0 (0%)	6 (60%)	2 (22%)	10 (100%)	18 (47%)
Usability Tracking	9 (100%)	10 (100%)	9 (100%)	10 (100%)	38 (100%)
Analytics	9 (100%)	10 (100%)	9 (100%)	10 (100%)	38 (100%)
Data Types	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Combined Total	84 (62%)	107 (71%)	112 (83%)	126 (84%)	

their online behaviour would be different after using the Paradox extension, SP17 said, "*very much so - it's astonishing the amount of data they collect on you, which is concerning and worrying*".

The technical groups admitted that their behaviour would be changed after using the Paradox extension. Participants agreed that the Paradox extension was a great way of visually "*seeing things in a more readable way*" and that "*some of the information is really worrying*". SP4 said, "*a lot of the time you see something pop up and people click 'yeah - I don't need to worry about this' but some of the information - the types particularly - were really quite worrying*" and that "*once data has been shared with another third-party it could be used for literally anything*". SP9 also noted that while their behaviour would change, "*the problem is [that] you can't use the Paradox extension until you've clicked on the website*".

6.3.27. *Impact on other websites*

Participants were asked whether they believed that the Paradox extension has made changes to their wider behaviour, including their expectations towards other websites which they had not visited with the extension. All the participants agreed that their opinions and expectations towards other websites had changed after using the Paradox extension.

Non-technical participants agreed that they would be more likely to check other websites with the Paradox extension before using them and that their wider expectations had changed. SP14 said, "*websites have far more information than I actually realise*" and "*I intend to use this on other websites now*".

Technical participants were worried about the extent of third party sharing on other websites and how their data can be used on other websites. Participants made a variety of comments, including: "*I now wonder about other websites, especially smaller websites*" and "*any other website could have privacy violations like this*". SP4 added that "*there's no clarity on how data will be used, so you have to assume that it will be used for anything*".

6.3.28. *Feedback*

Participants provided a variety of general feedback on the study, extension, and focus groups. Some of this feedback is outlined below in: General Feedback and Improvements to the Paradox extension.

Participants provided a range of feedback for the Paradox extension, all of which was positive: "*very clever tool [SP16]*", and "*very impressive [SP12]*".

SP9 also said that they now "*realised things I didn't know before*" as a result of the focus group.

A number of suggestions were made towards the Paradox extension which have been described at depth in Section 8.2.

From the focus groups specifically, suggestions were made towards increasing the visibility of violations within the full report, including tooltips within the full report to explain what each section means, adding direct links from the full report to the privacy policy, and porting the Paradox extension for use in Firefox.

6.4. *Interviews*

Semi-structured interviews were carried out to determine whether the changes in expectations, opinions and privacy awareness observed during the focus groups, could demonstrate more permanent behavioural changes. Interviews were carried out exactly a week after the focus groups; participants were asked not to change their browsing behaviour and to check the Paradox extension whenever it felt natural; participants were asked whether they were able to use the Paradox extension naturally throughout the week in Section 6.4.2.

The interviews were carried out using the same meeting software as the focus groups as the participants were now already familiar with the technology.

Participants were asked 5 separate questions related to their behaviours, opinions, and expectations while using the Paradox extension, in summary:

- Paradox Usage - Participants were asked whether they had managed to use the Paradox extension regularly throughout the week.
- User Expectations - Participants were asked if they found any more websites which did not meet their expectations as a result of using the Paradox extension for a week.
- Privacy Violations - If participants found any privacy violations, they were asked to explain whether they considered reporting it using the privacy violation reporter.
- Privacy Awareness - To determine if the participants still believed they were more privacy aware after using the Paradox extension, they were asked about whether their actions on websites they visited had changed since using the Paradox extension, and whether they believed their privacy awareness had changed.
- Impact on Other Websites - Participants were asked to consider whether they still believed that their opinions or expectations towards websites they had not visited with the Paradox extension, had changed since they had started using the Paradox extension. This was asked to determine whether the Paradox extension has a wider impact on the browsing habits of its users, or whether any potential change of privacy awareness was limited to websites which participants had seen a Paradox extension report for.

6.4.1. *Participant information*

Following the focus groups, eleven participants agreed to take part in an optional follow-up interview; participants will retain the same participant number throughout this evaluation, which is demonstrated in Table 8.

The interviews took around 5 min to complete, with some taking up to 10 min. Participants were given opportunities to speak freely about the Paradox extension and their privacy as well as being asked structured questions. The responses were analysed and have been presented as findings below.

Table 8 – Interview participant numbers.

Technical	Non-Technical
SP1, SP5, SP6, SP8, SP9, SP10	SP15, SP16, SP17, SP18, SP19

6.4.2. *Paradox usage*

All but one of the technical participants had managed to use the Paradox extension throughout the week, e.g. "*every website I've been on, I've checked it to see what come up*" [SP5]. SP6 was unable to use the Paradox extension since the focus group, saying, "*I haven't been able to, but not from lack of wanting to - time has just not allowed it*".

Of the non-technical participants, all managed to use the extension throughout the week since the focus group, e.g. "*Yes, I have. I visited quite a few websites with it*" [SP15].

6.4.3. *User expectations*

Aside from SP6 – who had not used the Paradox extension since their focus group – the technical users all found that their expectations had not been met on at least one website they had visited with the Paradox extension since their focus group, demonstrating a wider issue between user privacy expectations and privacy actuality. There were multiple cases where websites did not provide a method for requesting personal information, while websites without privacy policies were also found. SP10 also added that "*[the types of data being collected] did not meet my expectations at all*" for the websites which they had visited with the Paradox extension.

All the non-technical participants found websites where their expectations had not been met, apart from SP16; however, SP16 did add that they were "*unhappy with some data types*". SP18 found 6 violations on 1 website, including tracking methods which were being used on them, but were not declared within the privacy policy. Three of the participants also found websites which did not have a privacy policy.

SP17 said that they had used a VPN while browsing the internet at one point and found that their location was being exposed despite this; "*even though I had a VPN, it immediately knew my location*". The participant said that they had asked themselves, "*How are you bypassing my VPN?*" and that the Paradox extension revealed that "*session data on my device was being collected*". The participant expressed concern that they now believed the website was storing information about their device from previous visits to remember device location, which they had not known about before.

Collectively, most participants had found at least one website where their expectations were not being met since their focus group. This demonstrates that the gap between user privacy expectations and privacy reality is not limited to a small number of websites.

6.4.4. *Privacy violations*

None of the technical participants said they had reported a privacy violation, despite all admitting they had found at least one. The participants all suggested that they had considered the idea and would be likely to report a website in the future, e.g. "*I have considered it*" [SP5] and "*I didn't want to because they were small sites, I would if it was a major site*" [SP10]. SP8 suggested that they were "*not sure where the report will go or what will happen*".

There was a mixed response to violation reporting from the non-technical participants. When asked if SP17 would consider reporting a big company, they said, "*No, I wouldn't. I think It would be ignored. If it was a smaller company, yes I would*". SP16 was also reluctant to use the violation reporter, suggesting that "*I wouldn't be confident enough*".

SP15 found that their company intranet had no privacy policy by using the Paradox extension. As a result, they contacted the website administrator directly to ensure that one was put in place; "*we need to have a privacy policy*". Further to this, they asked the website administrator to use the Paradox extension when designing the policy, to ensure that it conformed as it would be expected to.

6.4.5. *Privacy awareness*

The technical participants all agreed that they remained more privacy aware since they started using the Paradox extensio, e.g. "*more aware of how companies try to find loopholes*" [SP1] and "*I've learnt things that I never knew existed before*" [SP8]. In addition, SP6 added, "*I would be very interested to use it more over time to see how other sites compare*".

The non-technical participants also agreed unanimously that they were more privacy aware, "*without a doubt*", e.g. "*I didn't realise how many things they actually knew about me. I'm more cautious about where I look*" [SP16] and "*It shows you data being collected which I hadn't considered*" [SP15].

6.4.6. *Impact on other websites*

Every participant, apart from SP8, suggested that their opinions or expectations towards websites which they had not visited with the Paradox extension had been changed by websites which they had visited with the Paradox extension. SP8 said that they were not as concerned as they were directly after the focus group, "*but it is quite worrying*". Other participants suggested that they were more concerned, e.g. "*if one or two big companies are doing it, then other companies will be doing it as well. It might not stop me using them though as a proportion of my internet functionality would be lost*" [SP1] and "*I did that a lot, I went onto one site, then went on to another to see if that website also did it. Most of the time, yes, they also do it*" [SP10].

All the non-technical participants agreed that their opinions towards other websites had been changed as a result of using the Paradox extension, e.g. "*I'm concerned that similar websites are collecting the same, if not more information*" [SP15] and "*yes it's raised awareness of this and the importance of how much is being collected. I'm more likely to compare websites with the Paradox extension*" [SP19].

6.4.7. *Encouraging discussion on privacy awareness*

SP6 found that they were shocked by the amount of data which was being collected and decided to share the results they had found with the Paradox extension; "*I discussed it with my wife and told her a number of the things that I now know are being collected and she was shocked at the amount of data and somewhat bemused at the types of data. My wife's reaction was like 'Oh my, you must be kidding!'*". This demonstrated that there is an interest in privacy awareness and shows that users can often be unaware of the privacy implications of using websites. However, it also demonstrates that through the Paradox extension, a participant was encouraged to speak out about privacy, and thereby potentially improve the privacy awareness of a non-participant.

6.4.8. *Acceptance of data collection*

One participant – SP9 – had demonstrated that they were already very sceptical and concerned about their privacy prior to using the Paradox extension. They suggested that – through the Paradox extension – they have come to accept that their data will be collected and shared, often without them knowing about it. As such, they admitted that their behaviour had changed and they had therefore decided to embrace or accept privacy issues facing them, rather than hide from them; "*I don't do cash back offers either as I don't want every company in the world having my email address and misusing/leaking it. I know it means I miss out on discounts, but I see it as the cost of extra security. But from what you showed me on Facebook gathering stuff about non-members, then not so sure anymore*".

6.4.9. *Changing privacy settings*

SP17 indicated that they had taken more of an interest in the privacy choices they can make after using the Paradox extension. As such, they made the decision to change the privacy settings for their Amazon devices; "*I changed the privacy settings on my Amazon devices as a result of using the Paradox extension*".

6.4.10. *Making purchases elsewhere*

SP19 said that they had made the decision not to make a purchase based on the feedback given to them by the Paradox extension for that website; "*I was going to buy something online but didn't because the website wouldn't let me have a copy of the data it was collecting according to the Paradox extension*". The participant indicated that this was particularly concerning "*especially as financial information was involved*". This demonstrates a change in behaviour, where a participant is showing a high level of privacy awareness and considering their privacy to be more valuable than the potential benefits of making a purchase from the website.

6.4.11. *Feedback*

Participants were positive about the Paradox extension overall. SP16 admitted that the Paradox extension "*has opened my eyes*" and SP17 asked, "*Can I keep using it? I want to carry on using it*". All the participants praised the Paradox extension, and many requested that its development continues beyond this study.

A number of suggestions were made towards the Paradox extension which have been described at depth in Section 8.2.

From the interviews specifically, suggestions were made towards porting the Paradox extension to Firefox, providing a clearer message to users when the Paradox extension has detected a privacy violation, and automatically reporting violations by sending them centrally under the Paradox extension's name, rather than through users.

At the time of writing, the Paradox extension has received two public reviews on the Chrome Web Store. Both reviews were positive and left a five-star rating.

Additional feedback was received (although not requested) after the Paradox extension was made public on the Chrome Web store. All the feedback received was positive with one specific comment saying, "*I don't think I ever want to use Amazon ever again [after] looking at what their T&Cs say about data collection from your plugin!*".

7. Discussion

7.1. *Data and anonymity*

Through the use of the Paradox extension, this study has demonstrated a large number of ways in which data is collected and used - often unbeknown to the user. Participants didn't consider either their privacy expectations, or potential privacy consequences. Much of how data can be collected and used is governed by the GDPR, which suggests that users place trust in this regulation without necessarily understanding it's implications. While anonymity is becoming a more important factor (with the increasing availability of technologie such as VPNs), anonynimity doesn't change how data is collected, stored and used. Whether a website collects personal data through an anonymity tool or not, that data is often the same (e.g. telephone number, address) and therefore the ways it can be stored and used remain the same.

7.2. *Usage observations*

When the Paradox extension was being designed to detect keywords from within privacy policies, multiple policies were manually reviewed with the aim of improving accuracy. In Facebook's privacy policy, the privacy impact of third-party tracking is particularly apparent: 'Partners provide information about your activities off Facebook - including information about your device, websites you visit, purchases you make, the ads you see and how you use their services - whether or not you have a Facebook account or are logged in to Facebook' (Facebook, 2018) (Fig. 21). The quote implies that Facebook may collect personal information, even if that person has never visited or heard of Facebook before. While third-party tracking is a relatively well-known practice, the legal implications of storing personal data on someone without their knowledge or consent are unclear.

The Paradox extension detected that Amazon uses GET/-POST requests, but the number of requests was based on whether the user was using ad-blocking technology. When an ad-blocker was not in use, the number of GET/POST requests detected was between 2–3, but when an ad-blocker was being used, that number rose to between 30–47 requests (Fig. 21). All the additional requests were related to advertising and comprised of a unique id, followed by a JavaScript function that contained links to many advertising partners. This approach is designed to bypass the usage of an ad-blocker, making it obsolete. By doing this, Amazon is taking control away from the user and forcing them - unknowingly - to view personalised

Fig. 21 – Third party data collection at facebook.

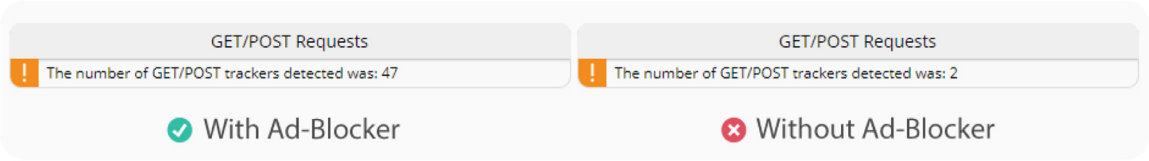

Fig. 22 – GET/POST requests on Amazon with and without an ad-blocker.

adverts and tracking regardless of personal preferences. Other technologies also influenced the number of requests made by Amazon, including cache, which – when cleared – resulted in a larger number of requests being made (Fig. 22).

7.3. *Participants had no prior privacy expectations*

None of the participants had considered their privacy expectations in any depth prior to the evaluation conducted with them. Only SP9 expressed concern for their privacy before using the Paradox extension, any many participants suggested that they had never thought about what their privacy expectations were, with some initially suggesting that they "*didn't have any*" [Section 6.3.3].

Some participants implied that they just expect websites to abide by privacy law and act responsibly, however they also admitted that they did not know if this really happened in practice. Website privacy seems to be something that very few people are considering - as demonstrated in Section 6.3.3. Trust and reputation perhaps have a much more significant effect on biases of using specific websites, both of which are easier to determine than privacy realities.

7.4. *Privacy expectations are different to privacy reality*

Evidence from Section 6.3.21 demonstrates that typically, users expectations are not met on websites. This pattern was also noticed after participants used the Paradox extension for a week, as described in Section 6.4.3. The most critical expectation not to be met was for the types of personal information websites were collecting; no participants were happy with the amount of personal information Amazon and Facebook were collecting.

Many users do not consider their privacy expectations before using a website – as explained in Section 7.3 – but when these expectations are considered, they are frequently not met. At Facebook, participants found that their expectations were met 84% of the time, while at Amazon this dropped to 67%.

90% of interviewed participants - who had used the Paradox extension since the focus group - found at least one other website which did not meet their expectations while using the extension for a week.

Privacy expectations of websites simply do not match up to user expectations. While there were cases where websites did meet user's privacy expectations, the number of times users reported that their privacy expectations were not met, demonstrates a clear divide between privacy expectations and privacy reality. This has implications; many users may unknowingly be using websites where their privacy is not being handled in a way that meets their expectations.

7.5. *Participants did not understand the extent of third party sharing*

Third-party selling and sharing split the opinions of the technical and non-technical participants in this study, as indicated in Section 6.3.24. Typically, non-technical users stated that they were unhappy with third parties having access to their data while technical users were far more comfortable with the idea.

It is possible that technical users are more aware of third party selling/sharing and have therefore become more familiar and accepting of it, thereby expecting it without necessarily understanding it. By contrast, non-technical users who are less familiar with the practice, made clear that they were far more concerned with it.

Despite the differences in opinion, all the (technical and non-technical) participants agreed that third party selling/sharing should not take place without prior consent. This seemed to be a universal opinion, with many participants also referring to various laws where they believed this was required. However – as demonstrated in Section 7.2 – Facebook admits to receiving personal information from third parties 'whether or not you have a Facebook account or are logged in to Facebook'. This has potential implications as it can make it difficult for users to track where their information is being sent, what information is being sent, when it is being sent, and whether they have any control over it - and in the case of Facebook, users won't necessarily even know that it is happening as their personal information may be shared with Facebook, whether they have an account or not.

7.6. *Paradox improved privacy awareness*

Participants overwhelmingly admitted to becoming more privacy aware after using the Paradox extension. All participants from the focus groups and interviews had become more aware of their privacy when using websites. This demonstrates that the participants were all willing to develop a better privacy mindset and suggests that a tool like the Paradox extension has the potential to cause a widespread change to privacy awareness.

In addition, the Paradox extension inadvertently encouraged participants to openly discuss their privacy expectations and findings outside of any evaluation. One participant unexpectedly demonstrated this by discussing their privacy expectations openly, as described in Section 6.4.7. While the Paradox extension can help individuals to understand the privacy

implications of specific websites, by inducing privacy discussion among its users, a much wider impact could be made. As such, the Paradox extension could be used alongside privacy awareness education/training as a tool for demonstrating privacy realities, which do not align with business or user expectations.

7.7. *Paradox was effective as an educational tool*

During the summative evaluation, participants demonstrated improved privacy awareness. Throughout the summative evaluation, much of the information provided to participants about their privacy was previously unknown to them. Additionally, usage of the Paradox extension enabled participants to directly apply, and understand, the new concepts which were introduced to them during the focus group or interview.

Thus, as a privacy education/awareness tool, the Paradox extension - when supported by a focus group and/or it's accompanying slides - demonstrated that it could successfully be used within privacy awareness/education training. The evaluation results support this; during the hour-long focus groups, every participant agreed that their privacy awareness had been improved.

The Paradox extension has been made publicly available via GitHub to enable anyone to use it to evaluate their own or other websites and for use as an educational tool. Paradox can also be downloaded via https://privacyparadox.co.uk or on the Chrome Web Store.

7.8. *Threats to validity*

7.8.1. *COVID-19*

Throughout this study, the COVID-19 pandemic caused severe disruption to people's lives; social distancing became essential and government-led isolations forced people to remain indoors. As a result of this, it is likely that individuals spent more time on the internet, and therefore it is possible that they spent more time on Amazon and Facebook.

This has the implication that people may have had a higher trust for Amazon and Facebook during this study, because they were deemed to be more essential. Online shopping dramatically increased as people deferred from visiting local stores, which were often largely out of stock. As such, it is likely that more people used Amazon for essential goods, and this may have had an impact on peoples trust for the website. Similarly – with more time spent indoors – people would be more likely to connect with friends and family through social media websites such as Facebook, as they were unable to visit them without breaching Government advice. Facebook would likely have gained an increased trust as it became a platform used for connecting people who would otherwise be unable to communicate.

This may have had an overall implication on the privacy expectations of the websites. The participants may have been more accepting of potential privacy issues because of how essential these websites had become during this pandemic. However, the results still show that users were very serious about their privacy and many participants were unhappy with aspects of the findings within the privacy policy summaries.

Furthermore – due to difficulties brought by the pandemic – a smaller sample size of participants was used. However, the demographic variety and expertise of participants was of a high standard, and the sample size considered appropriate; Marshall et al. suggest that 'Single case studies should generally contain 15 to 30 interviews' (Marshall et al., 2013).

7.8.2. *Differences in website types*

It is worth considering that Amazon and Facebook offer very different services. They were picked for this study due to their prominence as online offerings, but while Amazon is primarily an e-commerce (and cloud) company, Facebook is a social media website. Participants may have more favourability towards one of these websites, as it is likely that they would already be familiar with both sites, and therefore already have formed an opinion towards them. As such, participants may have expected that website to meet less of their expectations to begin with, and thus prove a threat to validity.

The difference in the services each of these websites offer is also a risk, as considered in Section 7.8.1 there may be some favourability towards one service, due to prior reliance creating a greater trust.

7.8.3. *Hawthorne effect*

The Hawthorne effect – as described in Section 2.9 – suggests that participants may perform better under evaluation conditions, as they know that they are being analysed. In the context of this study, it is possible that users behaved differently and anticipated that they were expected to demonstrate more privacy awareness as the evaluation progressed.

One way in which this was managed, was by determining the privacy expectations of users prior to introducing the Paradox extension. As such, it could be determined exactly what participants were expecting before they were shown the results from the Paradox extension and therefore, they could not claim that they had a different expectation than what they had previously stated.

7.8.4. *Website privacy evolution*

Websites will continue to monitor the laws encapsulating privacy, but the desire for them to capture personal information will remain the same due to its value. Many privacy tools already exist - and in many cases - websites have already identified ways around them. For example, Ad-Blockers can be avoided through detection mechanisms which then prevent page content from being rendered; Amazon take this approach further, and use alternate means to fetch ads when an ad-blocker is in use (as explained in Section 7.2).

This is the same for the Paradox extension; websites could potentially carefully reword their policies to conform to the Paradox extension's analysis and hence bypass keyword analysis which would otherwise detect potential privacy violations. However, this can be overcome by continually updating Paradox.

8. Conclusion

8.1. Summary

This study has demonstrated that - in line with the privacy paradox - participants did not fully understand the implications on their privacy, nor did they do anything specific to address it. Through using the Paradox extension, the expectations of participants were changed, and their privacy awareness was improved. This suggests that participants were more understanding of the privacy implications they faced when using websites, and some even demonstrated behavioural changes indicating improvements in privacy attitude over a short period of time.

In addition, this study has identified that many users do not have any privacy expectations or have never thought about what they might be. Clear differences between the privacy expectations of users and privacy realities were discovered, and invariably, participants suggested that many of the websites they use do not meet their privacy expectations.

8.2. Future improvements

Further research – using the Paradox extension – could be conducted to determine the privacy tolerance of individuals when they use websites. Through determining individual privacy tolerance, it may be possible to understand the personal privacy implications of the privacy paradox.

Individuals could be asked to use the Paradox extension while using websites to determine which privacy-related characteristics of websites cause behavioural changes. This could provide an understanding of which privacy expectations cause the greatest change in behaviour when they do not meet user expectations.

For this study, the measures which individuals take to protect their privacy - and the impact these measures have - were not considered. This was to focus on understanding whether websites meet privacy expectations, rather than understand the protections individuals use and their effectiveness.

Future work could be conducted to determine how privacy tools such as VPNs and the Tor protocol affect the privacy expectations of individuals, and to to use Paradox to determine how effective these tools are.

8.2.1. Increasing prominence of violations

During a focus group, it was suggested that the violations being detected by the Paradox extension were the most critical piece of information available to the user. Despite this, currently, the violations are displayed under 'Useful Information' which lies at the bottom of the report. Often, users found that - due to the size of the report - the 'Useful Information' section was often not visible to the user unless they scrolled to the end of the report.

As part of future work, the 'Useful Information' section will be moved to a more prominent position towards the top of the report. This will allow users to quickly see how many violations have been found by the Paradox extension and determine what they are with minimal effort.

8.2.2. Tooltips in full report

Feedback was received which requested that additional information was provided about each heading in the report to provide users (whether technical or non-technical) with a better context for understanding the implications of the findings in the full report.

To implement this, future work will seek to amend each section to include a tooltip. An image (such as a question mark) will appear next to each title and - when hovered-over - will show a brief description providing a context for the specific section. Ideally, this will help engage non-technical users more by helping to understand how to read the report.

8.2.3. Direct links to privacy policy

SP1 suggested that it would be useful to jump to areas in a privacy policy related to the findings displayed by Paradox in the full report. To rectify this, a button could be added which would take the user directly to the section of the policy which the Paradox extension is using for its analysis.

This can be implemented by storing the location of strings on the privacy and adding a button to each item in the policy analysis, which will allow users to directly access the relevant part of the policy.

8.2.4. Clearer violation warning

SP5 suggested that it was not clear whether there were any violations until they had opened the Paradox extension. They added that for many people, violations could end up being missed completely and privacy implications would not be made clear. The participant added that some method of prominently displaying to the user via a banner, popup, or message directly on the webpage would be incredibly useful and prevent people from missing privacy violations when browsing websites.

To implement this, when the Paradox extension detects a privacy violation, a red Paradox logo could be added to the HTML of the website and made to appear on top of the webpage. This will enable users to click on the logo to view the violation(s) found. Users can then also be given the option not to be shown violations warnings on the same website, once they have dismissed the warning.

8.2.5. Automatic violation reporting

Multiple participants suggested that they were uncomfortable with reporting violations, either because the process was too complicated, or because they were worried about attaching their name to such a report. It was therefore suggested that violation reporting should be done automatically, so that users do not have to take steps to report violations themselves. As such, violations would be centrally reported, and reporting would be done under the name of the Paradox extension, rather than the name of the user.

To implement this, a central database will need to be set up where privacy violations can be sent by users to the Paradox extension. The Paradox extension would then send out automatic emails to websites to inform them that a violation was detected on their website, and provide constructive information on how to improve it. In addition, the Paradox extension would need to ensure that multiple emails are not sent to the same website when receiving similar privacy violation reports

over a short time period. Ultimately, the decision to report a website would remain with the user, who would be given the choice to select which violation(s) to report, and add an optional comment. However - unlike the current system - their personal information would not be attached to the email in any way, as this would be handled centrally by the Paradox extension.

8.3. *Using paradox to assess or improve privacy awareness*

Formative education already takes place throughout the education system to inform students on the impacts of cyberbullying and safe ways to use the internet. The Paradox extension had an impact on the privacy awareness of every participant involved in this study, and could therefore be used as an education tool for demonstrating the privacy consequences of real websites, which students are likely to use, such as Facebook. This could change how students behave online by demonstrating how their personal information may be used, shared, or sold.

Businesses could benefit from using the Paradox extension by defining what they consider to be unacceptable privacy consequences. They could then distribute Paradox to their employees to ensure that the websites being visited adhere to the privacy expectations laid out by the business. This could be accompanied by a privacy awareness training/education programme for the employees using Paradox, which may benefit from the use of materials created in this study].

After each focus group conducted during this study, every participant claimed that their privacy awareness had been raised. As such, the material used during the focus groups could be used - along with the Paradox extension - as privacy awareness training/education material.

By getting participants to think about their privacy expectations and compare these with privacy realities, this study has demonstrated that individuals frequently discover differences between expectation and reality, and thereby have a higher privacy awareness as a result. This same concept could be applied on a wider scale to improve the privacy awareness of individuals more generally.

The Paradox extension could be used as a tool when designing a new website or privacy policy to ensure that it is meeting user expectations. This could be achieved by evaluating new websites and privacy policies against the Paradox extension internally, orby running an external evaluation which could make use of existing material.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Callum Pilton: Conceptualization, Methodology, Software, Validation, Writing - original draft. **Shamal Faily:** Supervision, Writing - review & editing. **Jane Henriksen-Bulmer:** Writing - review & editing.

REFERENCES

Acquisti A, Grossklags J. Privacy attitudes and privacy behavior. In: Advances in Information Security, vol 12. Boston, MA: Springer; 2004. p. 165–78. chapter 13

Avison DE, Fitzgerald G. Information Systems Development: Methodologies, Techniques & Tools.. fourth ed. McGraw-Hill Education; 2006.

Blank, G., Bolsover, G., Dubois, E., 2014. A new privacy paradox: young people and privacy on social network sites. Prepared for the Annual Meeting of the American Sociological Association.

Bujlow T, Carela-Espanol V, Lee BR, Barlet-Ros P. A survey on web tracking: mechanisms, implications, and defenses. Proc. IEEE 2017;105(8):1476–510. doi:10.1109/JPROC.2016.2637878.

Carlini N, Felt AP, Wagner D. An evaluation of the google chrome extension security architecture. In: Proceedings of the 21st USENIX Security Symposium; 2012. p. 97–111.

Cranor LF. P3P: making privacy policies more useful. IEEE Secur. Privacy 2003;1(6):50–5. doi:10.1109/MSECP.2003.1253568.

Cranor, L. F., Langheinrich, M., Marchiori, M., Prseler-Marshall, M., Reagle, J., 2002. The Platform for Privacy Preferences 1.0 (P3P1.0) Specification. [online]URL: https://www.w3.org/TR/P3P/.

Cranor, L. F., Reagle, J., Ackerman, M. S., 1999. Beyond concern: understanding net users' attitudes about online privacy. 9904010.

Deuker A. Addressing the privacy paradox by expanded privacy awareness – the example of context-aware services. IFIP Adv. Inf. Commun.Technol. 2010;320:275–83. doi:10.1007/978-3-642-14282-6_23.

European Commission, 2018. Regulation (EU) 2016/679 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation – GDPR). 10.5771/9783845266190-974

Extension Monitor, 2019. Breaking Down the Chrome Web StoreURL: https://extensionmonitor.com/blog/breaking-down-the-chrome-web-store-part-1.

Facebook, 2018. Data PolicyURL: https://www.facebook.com/policy.php.

Flavián C, Guinalíu M. Consumer trust, perceived security and privacy policy: three basic elements of loyalty to a web site. Ind. Manage. Data Syst. 2006;106(5):601–20. doi:10.1108/02635570610666403.

Google, 2020. Stay Secure - Google Chrome. [online]URL: https://developer.chrome.com/extensions/security.

IBM Studios. In: Technical Report. IBM Design Thinking Field Guide. IBM; 2016. URL: https://www.ibm.com/cloud/garage/content/field-guide/design-thinking-field-guide

Kokolakis S. Privacy attitudes and privacy behaviour: a review of current research on the privacy paradox phenomenon. Comput. Secur. 2017;64:122–34. doi:10.1016/j.cose.2015.07.002.

Li T-C, Hang H, Faloutsos M, Efstathopoulos P. TrackAdvisor: taking back browsing privacy from third-party trackers. In: Mirkovic J, Liu Y, editors. In: Passive and Active Measurement. Cham: Springer International Publishing; 2015. p. 277–89.

Linden T, Khandelwal R, Harkous H, Fawaz K. The privacy policy landscape after the GDPR. Proc. Privacy Enhancing Technol. 2020;2020(1):47–64. doi:10.2478/popets-2020-0004.

Macefield R. Usability Studies and the Hawthorne Effect. J. Usability Stud. 2007;2(3):145–54.

Marshall B, Cardon P, Poddar A, Fontenot R. Does sample size matter in qualitative research?: A review of qualitative

interviews in is research. J. Comput. Inf. Syst. 2013;54. doi:10.1080/08874417.2013.11645667.
Mayo ST, Dooley RJ. Book reviews: Desmond l. Cook. The impact of the hawthorne effect in experimental designs in educational research. U.S. department of health, education, and welfare, office of education, project no. 1757. Columbus, Ohio: Ohio state university, 1967. pp. ii + 160. Educ. Psychol. Meas. 1968;28(4):1255–9. doi:10.1177/001316446802800434.
McCambridge J, Witton J, Elbourne DR. Systematic review of the hawthorne effect: new concepts are needed to study research participation effects. J. Clin. Epidemiol. 2014;67(3):267–77. doi:10.1016/j.jclinepi.2013.08.015.
Mehta P. Creating Google Chrome Extensions. Berkeley, CA: Apress; 2016. doi:10.1007/978-1-4842-1775-7.
Nielsen J. 10 Heuristics for User Interface Design. Nielsen Norman Group; 1995. p. 1–2.
Norberg PA, Horne DR. Privacy attitudes and privacy-related behavior. Psychol. Mark. 2007;24(10):829–47.
Obar JA, Oeldorf-Hirsch A. The biggest lie on the internet: ignoring the privacy policies and terms of service policies of social networking services. Inf. Commun. Soc. 2018;23(1):128–47. doi:10.1080/1369118X.2018.1486870.
Olurin M, Adams C, Logrippo L. Platform for privacy preferences (P3P): current status and future directions. In: 2012 10th Annual International Conference on Privacy, Security and Trust, PST 2012; 2012. p. 217–20. doi:10.1109/PST.2012.6297943.
Paramarta V, Jihad M, Dharma A, Hapsari IC, Sandhyaduhita PI, Hidayanto AN. Impact of user awareness, trust, and privacy concerns on sharing personal information on social media: Facebook, twitter, and instagram. In: 2018 International Conference on Advanced Computer Science and Information Systems, ICACSIS 2018; 2019. p. 271–6. doi:10.1109/ICACSIS.2018.8618220.
Peterson D, Meinert D, Criswell J, Crossland M. Consumer trust: privacy policies and third-party seals. J. Small Bus. Enterp.Dev. 2007;14(4):654–69. doi:10.1108/14626000710832758.
Pötzsch S. Privacy awareness: a means to solve the privacy paradox? IFIP Adv. Inf. Commun.Technol. 2009;298(216483):226–36. doi:10.1007/978-3-642-03315-5_17.
Preskill H, Russ-Eft D. Evaluation models, approaches, and designs. In: Building Evaluation Capacity; 2012. p. 102–81. doi:10.4135/9781412983549.n5.
Riegelsberger J, Sasse MA, McCarthy JD. The mechanics of trust: a framework for research and design. Int. J. Hum. Comput. Stud. 2005;62(3):381–422. doi:10.1016/j.ijhcs.2005.01.001.
Schaub F, Marella A, Kalvani P, Ur B, Pan C, Forney E, Cranor LF. In: NDSS Workshop on Usable Security. Watching them watching me: browser extensions impact on user privacy awareness and concern; 2016. doi:10.14722/usec.2016.23017.
Sedgwick P, Greenwood N. Understanding the hawthorne effect. BMJ (Online) 2015;351(September):1–2. doi:10.1136/bmj.h4672.
Sjösten A, Van Acker S, Sabelfeld A. Discovering browser extensions via web accessible resources. In: CODASPY 2017 - Proceedings of the 7th ACM Conference on Data and Application Security and Privacy; 2017. p. 329–36. doi:10.1145/3029806.3029820.
Toffler, A., 2013. Writing SMART, Short-term Outcome Objectivescited 3 March 2020 URL http://ncwatershednetwork.org/wp-content/uploads/2016/05/Writing-SMART-short-term-outcomes-and-objectives.pdf.
Tsalis N, Mylonas A, Gritzalis D. An intensive analysis of security and privacy browser add-ons. In: Lambrinoudakis C, Gabillon A, editors. In: Risks and Security of Internet and Systems. Cham: Springer International Publishing; 2016. p. 258–73.
Weitzner DJ, Hendler J, Berners-Lee T, Connolly D. Creating a policy-aware web: discretionary, rule-based access for the world wide web. In: Web and Information Security; 2005. p. 1–31. doi:10.4018/978-1-59140-588-7.ch001.
Wu KW, Huang SY, Yen DC, Popova I. The effect of online privacy policy on consumer privacy concern and trust. Comput. Hum. Behav. 2012;28(3):889–97. doi:10.1016/j.chb.2011.12.008.
Youker BW. What, how, and why? A comparative analysis of 12 goal-free evaluations. J. MultiDiscip. Eval. 2019;15(33):16–29.

Callum Pilton is a former student of Bournemouth University having completed his B.Sc. in 2020 while studying Forensic Computing & Security. He is currently working as a software developer at IBM with several years of experience within the industry. He has been involved in multiple projects involving web technologies, anonymity, and messaging.

Dr Jane Henriksen-Bulmer is a Lecturer in Computing, with several years of experience in industry working as a project manager, business analyst and researcher. She completed her Ph.D. in Privacy in 2019 and since then has been teaching, among other things Business System Analysis and Design, Information Assurance and Business Continuity at BU. During her time at BU Janehas worked on several outreach projects, collaborating with several external organisations and charities on research projects. This includes 3 commercialisation projectsthat sought to take university research and turn this into commercially viable products. Before joining the Ph.D. programme at BU, Jane completed a MSc in Information Technology from Bournemouth University in 2015, gaining a distinction. She also holds a Masters in Business Administration (MBA), gained from Curtin University in Perth, Western Australia in 2013 and a first-class honours degree in Law from the Open University.

Dr Shamal Faily is currently working at Bournemouth University as a Principal Lecturer in Systems Security Engineering within the Department of Computing and Informatics. Before joining BU as a lecturer in 2013, Shamal was previously a post-doctoral researcher at the Department of Computer Science at the University of Oxford, and a teaching fellow at the Information Security Group at University College London. He completed his DPhil in Computer Science at the University of Oxford in 2011. Prior to his doctoral research, he was a software engineer within Logica's Space business.