(Smart)Watch Out! Encouraging Privacy-Protective Behavior through Interactive Games

Meredydd Williams*[a], Jason R. C. Nurse[b], Sadie Creese[a]

[a]*Department of Computer Science, University of Oxford, UK*
[b]*School of Computing, University of Kent, UK*

Abstract

The public frequently appear to overlook privacy, even when they claim to value it. This disparity between concern and behavior is known as the Privacy Paradox. Smartwatches are novel products that offer helpful functionality. However, although they often store sensitive data (e.g. text messages), owners rarely use protective features (e.g. app permissions). Campaigns have sought to increase privacy awareness, but initiatives tend to be ineffective. We therefore explore the efficacy of a serious game in encouraging protective smartwatch behavior. The application is designed with Learning Science principles and evaluated through a study with 504 smartwatch owners. After soliciting concerns and behavior, our treatment group [n = 252] play the online simulation. Our control group do not participate [n = 252], as we seek to limit extraneous variables. In a follow-up session, all users report posttest responses and qualitative justifications. We appear to encourage protective behavior, with our treatment group using privacy features more often. We also significantly reduce the prevalence of the Paradox, realigning behavior with concern. These quantitative findings are complemented by an inductive analysis of user rationale. Smartwatch behavior is influenced by several factors, including privacy awareness and data sensitivity. Finally, we use Protection Motivation Theory (PMT) to develop intervention recommendations. These include risk exposure tools and protective demonstrations. To our knowledge, this is the first tool to encourage protective smartwatch behavior.

Keywords: Privacy, behavior change, smartwatch, serious game, wearable, IoT

Email addresses: meredydd.williams@cs.ox.ac.uk (Meredydd Williams*), j.r.c.nurse@kent.ac.uk (Jason R. C. Nurse), sadie.creese@cs.ox.ac.uk (Sadie Creese)

1. Introduction

The public claim concern for their privacy, as expressed through a range of polls and surveys (Pike et al., 2017; Morar Consulting, 2016; Rainie et al., 2013). However, we often exhibit behavior considered in tension with the principle (Felt, Ha, Egelman, Haney, Chin and Wagner, 2012). This disparity between concern and behavior has been labeled the Privacy Paradox (Barnes, 2006). The issue can occur for several reasons, including lack of awareness (Deuker, 2009), social norms (Phelan et al., 2016) and hyperbolic discounting (Kokolakis, 2017). Such disparities can place users at risk, as they may not recognize (or be able to prevent) the threats that they face (Deuker, 2009).

The Internet-of-Things (IoT) is expanding rapidly, populating our world with connected devices. But many of these exciting products, including voice assistants (Ford and Palmer, 2018), smart TVs (Ghiglieri et al., 2017) and wearables (Lee et al., 2016), can pose privacy risks. Smartwatches offer clear convenience benefits through storing messages, contacts and location data (Do et al., 2017). However, users rarely apply protection (Udoh and Alkharashi, 2016), placing their personal information under threat (Dai et al., 2016). This has led to the Paradox becoming prevalent in smartwatch environments (Williams et al., 2017). As these devices continue to proliferate, can we align behavior with privacy concerns?

Deuker (2009) believed the Paradox could be mitigated through awareness, as did Pötzsch (2009) in her design requirements. Furthermore, when asked to justify the Paradox, users frequently cite a lack of understanding (Williams et al., 2017). However, highlighting an issue is not sufficient to change behavior (Bada et al., 2015). In addition to awareness and education, individuals must be encouraged to adjust their actions (Dolan et al., 2010). Serious games offer greater promise, since they embed incentives within an engaging environment (Connolly et al., 2012). Furthermore, studies suggest that such apps can aid retention more than conventional instruction methods (Wouters et al., 2013). Since privacy games have proved successful (Barnard-Wills and Ashenden, 2015; Raynes-Goldie and Allen, 2014), we explore their influence in a smartwatch environment.

We seek to encourage the use of smartwatch privacy features, such as app permissions and screen locks. To achieve this, we constructed an online game for wearable behavior change. This approach supported a large-sample evaluation (with smartwatch users), and such simulations are common in top-tier privacy research (Rajivan and Camp, 2016; Wang et al., 2015; Kelley et al., 2013; Jackson and Wang, 2018). The app was carefully designed with principles from Human-Computer Interaction (HCI) and Learning Science (Quinn, 2005; Annetta, 2010).

Players were then tasked to solve smartwatch privacy challenges, such as restricting the permissions of invasive apps. Through learning and practicing protective techniques, they could refine new skills. To ascertain the influence of the game, we conducted a pretest-posttest study with real smartwatch owners [$n = 504$]. Participants disclosed their concerns and behavior, before being divided into control and treatment groups. They then gave updated responses one week after the gameplay session. We studied behavior frequency, assessing how consistently a protective feature (e.g. screen lock) was used.

Protection became significantly more frequent in the treatment group, suggesting privacy games might encourage action. This is our primary contribution, since behavior change is a challenging endeavour. The Paradox also became significantly less prevalent, while our control group did not differ in either case. We then sought to dissect privacy rationale: the reasoning behind individuals' privacy decision-making. These justifications were analysed through a rich process of inductive analysis. Several factors appeared to influence behavior, including privacy awareness and perceived data sensitivity. Finally, we considered our findings through the lens of Protection Motivation Theory (PMT) (Rogers, 1983). In this model, threats and abilities are weighed to determine behavior. The theory was used to recommend privacy interventions, including risk exposure tools and protective demonstrations. Our contributions are as follows:

1. Primarily, through an evaluation with 504 smartwatch users, evidence that privacy-protective behavior can be encouraged.
2. A rare qualitative analysis of the privacy rationale of smartwatch owners.
3. The design and development of, to our knowledge, the first game to address smartwatch privacy.

This paper is structured as follows. Section 2 outlines the background, including previous work, smartwatch features and the assumed threat model. We then introduce the game in Section 3, highlighting the principles that informed its design. Section 4 details our research hypotheses, supporting the study procedure in Section 5. Both our quantitative and qualitative findings are outlined in Section 6. Finally, Section 7 concludes with contributions, limitations and future work.

2. Background

2.1. Privacy and the Privacy Paradox

Definition. Since privacy is contextual (Nissenbaum, 2009) and subjective (Palen and Dourish, 2003), people might value it in one environment but not another. To set the context, we scope our work to informational privacy, defined as

"*the interest an individual has in controlling, or at least significantly influencing, the handling of data about themselves*" (Clarke, 1999).

Privacy Paradox. We explicitly define the issue as '*the disparity between claimed informational privacy concerns and informational privacy behavior*'. This aligns with Barnes' original concept (2006): that individuals oppose privacy invasion and yet share their data widely. Opinion polls (Rainie et al., 2013; Morar Consulting, 2016) suggest that the public care about their privacy. In a 2017 survey (Pike et al., 2017), 84% claimed to be worried about their data. 70% of those surveyed also stated that their concerns had recently increased. However, despite these assertions, we frequently place our privacy at risk. We avoid policies (McDonald and Cranor, 2008), ignore permissions (Felt, Ha, Egelman, Haney, Chin and Wagner, 2012) and rely on lax default settings (Bonneau and Preibusch, 2010). Keith et al. (2013) studied Privacy Calculus, which compares the risks and benefits of disclosure. In their 1,025-person experiment, they discovered that although individuals may have intentions, they often fail to act. Acquisti and Gross (2006) found that even those who expressed reluctance would share data on social networks. This is similar to the attitude-behavior gap (Fazio and Roskos-Ewoldsen, 2005) found in psychology research. The disparity is particularly concerning for privacy, as users might place themselves at risk.

Justifications. Several justifications have been given for the Paradox, including cognitive heuristics (Gambino et al., 2016), social representations (Oetzel and Gonja, 2011) and concern-reward imbalances (Hallam and Zanella, 2017). In their 2018 work, Bandara et al. (2018) used Construal Level Theory to dissect the phenomenon. They concluded that privacy is usually perceived as abstract, and therefore intentions rarely lead to behavior. While Internet-of-Things (IoT) analysis is sparse, Williams et al. (2017) undertook a comparative study. They solicited the concerns and behavior of the public, comparing the Privacy Paradox across product categories. They found that smartwatches were most prone to the issue, with protective features rarely used. Since their participants blamed insufficient knowledge, privacy instruction appears a logical solution. While these works offer useful insights, we differ by actively seeking to reduce the Paradox.

2.2. Privacy Behavior Change

Overview. To reduce the Paradox, we must actively adjust privacy behavior. Previous work suggests the issue can be mitigated by raising awareness. Knowledge is also important, as a person might not recognize they are endangering their data (Williams et al., 2017). Deuker (2009) attributed the disparity to incomplete information, bounded rationality and psychological variables. He explained how

awareness-building could address the first two components, but did not conduct empirical research. Pötzsch (2009) claimed the Paradox could be reduced through aligning concern to action or action to concern. Therefore, if we wish to encourage protective behavior, we should remind individuals of their trepidation.

Learning interventions have prompted protection, but not in the context of smartwatches. Albayram et al. (2017) conducted a study of 228 smartphone users, seeking to incentivize screen locking. After watching an instructive video, their treatment group reported adjusting their behavior. Albayram et al. (2017) later analyzed how risk, self-efficacy and contingency influence the use of Two-Factor Authentication (2FA). They found participants were more willing to try 2FA after watching a video. We follow the methodological structure of the first study, but provide novelty through our serious game.

Three-step behavior change. While awareness can highlight risk, behavior is often not adjusted by providing information (Bada et al., 2015). Sasse et al. (2007) theorized that change must be delivered through three steps. Attention should first be drawn to an issue, before learning is provided. Finally, training should be given to help users refine their behavior. Even if participants then understand protection, they must also be motivated to act (Bada et al., 2015). Rather than mandating compliance, user incentives should align with tasks (Renaud et al., 2014). For example, individuals might feel empowered by taking control of their data.

The above theory has been applied in many privacy studies. Kirlappos and Sasse (2011) explored how security education might deter phishing. They assessed the efficacy of warning notices, comparing these against signs of reliability. The authors concluded that effective learning should be driven through the three-step approach. Furman et al. (2011) also considered security education through this model. Through their in-depth interviews, they found their participants were aware of privacy risks. While some even possessed basic knowledge, few had the skillset to guard their data. Through providing awareness, information and training, we seek to encourage protective behavior.

Protection Motivation Theory. Within this work, we deconstruct behavior through Protection Motivation Theory (PMT) (Rogers, 1983). While other models were considered (including the Theory of Reasoned Action (TRA) (Fishbein, 1979) and the Theory of Planned Behavior (TPB) (Ajzen, 1991)), they posed limitations. TRA did not appear suitable, since it does not include behavioral constraints or self-efficacy (Briggs et al., 2017). TPB was also deemed insufficient, as risk is not factored into considerations (Norman and Conner, 1996). We preferred PMT, which has been recommended for cyber security behavior change (Briggs et al., 2017). It has also been deemed effective through large-scale meta-

analysis (Floyd et al., 2000). In PMT, threats (perceived severity and vulnerability) and benefits are weighed against coping approaches (response efficacy and self-efficacy) and response costs. This all contributes to intention, and potentially behavior. If a risk is perceived and privacy settings are understood, a person might be expected to act. We seek to highlight these risks while encouraging protection.

This theory has proved successful in previous privacy research. Chenoweth et al. (Chenoweth et al., 2009) used the model to predict the usage of protective technologies. They studied how the factors would affect the intention to use antivirus tools. Through an analysis of 232 participants, response costs and perceived vulnerability appeared to be influential. Another study used PMT to assess how concerns impact protective behavior (Youn, 2009). This was undertaken through a questionnaire distributed to 144 students. Perceived vulnerability was most important, though perceived rewards also had significance. In addition to the above studies, PMT has been used to study behavioral intentions (Anderson and Agarwal, 2010) and encourage online protection (Shillair et al., 2015*a*). Although these works did not explore smartwatches, they align well with our research.

Privacy research. Behavior change has been achieved in non-smartwatch environments. Egelman et al. (2009) found users were willing to pay for protection when privacy indicators were provided. The icons highlighted when site policies conflicted with a person's preferences. While this technique was successful, the researchers augmented user interfaces. This approach would be challenging on proprietary smartwatches. 'Nudging' has also proved effective, guiding individuals to privacy-conscious decisions (Wisniewski et al., 2016). In a six-week trial, Wang et al. (2014) studied the influence of nudges on Facebook. When users' audiences were made salient, unintended disclosure was reduced. Most notably, nudging has succeeded in mitigating the Paradox. Jackson and Wang (2018) designed personal notifications for mobile apps. Through a 241-person online simulation, behavior was realigned with intentions. However, rather than evaluating later actions, this was achieved within a single session. This poses a limitation, since the efficacy of nudging tends to decrease over time (Voyer, 2015). Furthermore, once this soft paternalism is removed, users often neglect the lessons (Bruyneel and Dewitte, 2016). This introduces the potential for serious games, as highlighted below.

2.3. Serious Games

Definition. Serious games are defined as "*any form of interactive computer-based game software...that has been developed with the intention to be more than*

entertainment" (Ritterfeld et al., 2009). We develop what is termed an 'intervention game': one that's primary goal is to encourage behavior change. This aligns with the terminology used in a wide range of influential applications (Kato et al., 2008; Huizenga et al., 2009; Lovio et al., 2012; Deriso et al., 2012).

Efficacy. Applications are not a panacea and can suffer from several issues. Contextualization is challenging, with this limiting the relevance of instruction. Individuals might also become distracted and overlook the information. Nevertheless, apps can aid retention more than conventional instruction methods (Wouters et al., 2013). Wouters et al. (2013) conducted a meta-analysis, comparing serious games with traditional instruction. Their evaluation contained 5,547 participants across 39 studies. Games were found to be significantly more effective for retention, with this concerning the recall of information after several weeks. Therefore, with users lacking privacy awareness, this approach appears apt.

Security games. While few games have been developed for smartwatches, related work is promising. Anti-Phishing Phil (Sheng et al., 2007) taught users how to avoid phishing campaigns. This online game challenged participants to evade risky predators. It was found that players became better able to identify fraudulent websites. The authors (Kumaraguru et al., 2009) also created PhishGuru: a training application with fake emails. Individuals were targeted through spearphishing and challenged to avoid the links. Even 28 days after participation, players were found to retain knowledge. The learning techniques in this application, including contextualization and immediate feedback, were adopted in our game. Hale et al. (2015) built on similar concepts, developing a rich browser simulation. Users navigated websites and were asked to rate the reliability of content. The participants successfully identified attacks, whether targeting social networks or news portals. We also seek to encourage learning through simulated interactions. NoPhish, developed by Canova et al. 2014, approached the topic in a smartphone environment. Players were awarded points for protection and penalised for unsafe decisions. Our application similarly implemented these techniques.

CyberCIEGE (Irvine et al., 2005) did not concern phishing, but taught trainees how to protect their networks. Through a 3D simulation, users were rewarded for securing assets and completing objectives. Although it was not evaluated by its creators, Raman et al. (2014) found trained users outperformed their control group. Flushman et al. (2015) also considered security, using alternate reality games (ARGs) to teach university students. After participating, individuals were seen to act in a more-protective manner. While the aforementioned games were effective, they all targeted a traditional desktop environment.

Privacy games. Games have also been theorized for privacy instruction. Im-

maculacy (Suknot et al., 2014) is a proposed interactive story, in which the player is immersed in a dystopian world. Scores increase based on privacy-protective decisions, incentivizing users to reflect on their actions. Our application follows similar principles, rewarding individuals when they configure smartwatch settings. Barnard-Wills and Ashenden (2015) developed a card game, seeking to highlight the value of personal data. It was entitled 'Privacy' and evaluated on three levels: as a game, as an instructive tool, and as a research aid. While most participants reported enjoying the experience, the game did not simulate real interactions. Furthermore, activities were not contextualized within a simulated environment. These facts do not detract from the content of the game. However, privacy is contextual (Nissenbaum, 2009) and contextualized apps have been found to be effective (Steiner et al., 2009). The Open University adapted the game into an online application[1]. This app highlighted how data can desired by many entities. While it possessed rich interactivity, the game has not been empirically evaluated.

Raynes-Goldie and Allen (2014) took a different approach to developing privacy applications. They constructed 'The Watchers', a digitally-augmented board game. It sought to encapsulate the informational transactions that children experience. Our application shares similarities, since it encourages practice within a safe environment. However, we differ by conducting an evaluation of efficacy. Kumar et al. (2018) sought to co-design privacy games in their 2018 study. They engaged directly with children to extract recommendations for future apps. These suggestions included relatable narratives and a variety of privacy consequences. While Kumar et al. did not evaluate a novel application, we adopt their design recommendations. 'Ministry of Sharing'[2], developed by the Open University, was targeted at a broader audience. This app challenged individuals to make disclosure decisions in common scenarios. Based on user actions, they were assigned privacy classifications. While the game sought to encourage protection, it also lacks empirical evaluation.

Principles. We designed our game based on techniques from Learning Science and HCI. While this work is empirical rather than theoretical, we sought to root the game on sound principles. These principles, and the literature supporting their efficacy, are outlined below.

Reinforcement. Positive reinforcement is a common principle used in serious games (Linehan et al., 2011). It is a subcomponent of operant conditioning,

[1] `http://www2.open.ac.uk/openlearn/privacy/game/`
[2] `https://www2.open.ac.uk/openlearn/ministry-of-sharing/`

where individuals develop a causal relationship between behavior and outcome (Skinner, 1953). In positive reinforcement, a desirable stimulus is presented as the consequence of a particular action. Since an individual should favour this outcome again, they are encouraged to repeat their behavior. Linehan et al. (2011) outlined empirically-validated guidelines for serious games. They discussed how the award of in-game points might encourage the repetition of particular actions.

Negative reinforcement is also an effective technique. This approach encourages behavior through the removal of a negative stimulus (Linehan et al., 2011). For example, an individual losing a game might be forced to restart. This time-consuming issue might encourage the user to avoid defeat. The technique was implemented in the Fish'n'Steps game by Lin et al. (2006). In their application, physical activity controlled the development of an animated fish. Negative reinforcement was used, since a lack of exercise would damage their character. Through a 14-week study, the authors found their game encouraged activity.

Incentives. To encourage behavior change, incentives are often required. By aligning new actions with a person's desires, they should be motivated to act. As highlighted by Bada et al. (2015), this approach can be more effective than mandating compliance. They described how campaigns often instruct individuals to act securely. However, this may be ineffective if users are not motivated.

While Richter et al. (2015) considered gamification, their techniques are applicable to gameplay environments. They discussed a variety of means in which new behavior could be encouraged. Indeed, incentives could be driven through empowerment and social comparison. This might be provided by interactivity and leaderboards, respectively.

Serious games can encourage actions through intrinsic motivation. This is defined as undertaking an action for its inherent satisfaction (Ryan and Deci, 2000). For example, a person might play a game purely because it is enjoyable. Extrinsic motivation emerges when behavior is driven by an external reward (Ryan and Deci, 2000). In a gameplay environment, individuals might receive a financial prize for topping a leaderboard. Breuer and Bente (2010) considered how such principles could be applied to serious games. They outlined that personalization and customization were sound approaches to motivate users.

Customization. Customization is an effective principle in serious games (Annetta, 2010). Applications might appear more engaging when personalized by a participant. Charsky (2010) described this matter as an 'expressive choice'. While the adjustment of traits might not affect the game, it builds empathy between the player and their character. This should enhance participant engagement, and therefore the likelihood of behavior change (Gee, 2003).

Some games use avatars to represent the player's character. This provides a virtual embodiment of an individual's actions. The implementation of such a technique has been found to increase immersion (Annetta and Holmes, 2006). A less-common approach is to support the customization of this avatar. In a 2008 work, Annetta compared engagement between two personalization approaches (Annetta, 2008). While some avatars could only toggle between genders, others could be specified in detail. When the avatar was customized, this resulted in greater participant satisfaction.

Contextualization. As highlighted earlier, privacy is inherently contextual (Nissenbaum, 2009). An event deemed invasive on a street might not be excessive inside a police station. Serious games can also be influenced by the concept of contextualization. Steiner et al. (2009) considered the importance of grounding content. They assessed a space-travel app and surmised that "*educational game[s] need to appropriately contextualize learning activities ... to retain [the] flow and engagement of the gaming experience*".

Context can also be enhanced through the inclusion of Non-Player Characters (NPCs). These are gameplay avatars that, while interactive, are not controlled by the participant. Maldonado et al. (2005) studied the influence of such entities in a 76-person study. Whereas one of their groups did not interact with an NPC, the other two groups had interactive companions. The authors found that those with NPCs were significantly more retentive in posttest evaluations.

These characters are valuable in providing feedback; another key component of serious games. As highlighted by Johnson et al. (2017), feedback is used to "*direct learners to improve performance, motivation, or learning outcomes*". Ke (2011) conducted a meta-analysis of the approaches underpinning learning games. In their qualitative evaluation of 89 studies, they concluded that "*instructional support features*" are key to such games and should be embedded through "*elaborative feedback*". Feedback is particularly useful when it can correct undesired behavior. By providing 'knowledge of correct results' (Johnson et al., 2017), participants should then be able to better understand their strengths and weaknesses (Benton et al., 2018).

Novelty. As highlighted above, serious games are considered valuable for behavior change (Wouters et al., 2013). However, they have rarely targeted the rapidly-expanding domain of smartwatches.

2.4. Smartwatches and Behavior

Definition. While smartwatches have been prototyped for over a decade (Smith, 2007), they have gained recent consumer popularity. A smartwatch can be defined

as "*an electronic wristwatch that is able to perform many of the functions of a smartphone*" (Collins English Dictionary, 2017).

Behavior. Although few behavioral studies have yet been undertaken, some have explored interactions in detail. Pizza et al. (2016) used wearable cameras to analyse smartwatch use. They monitored 12 participants for 34 days, finding their devices frequently served as timepieces. Notifications also appeared popular, but privacy features were not explored. Jeong et al. (2017) conducted a longitudinal study, collecting data on 50 users over 203 days. Although they analyzed device removal and wear duration, security and privacy were not considered.

Privacy research. While they explored wearables in general, Lee et al. (2016) studied privacy concerns at a large scale. They surveyed over 1,700 individuals to understand the issues that provoked the most anxiety. We scope specifically to smartwatches but use similar privacy scenarios. Ernst and Ernst (2016) investigated how risk perceptions influence smartwatch usage. They surveyed 229 users and found perceived usefulness increased behavioral intention. Perceived risks dissuaded participants from interaction, implying they weighed costs against benefits. This suggests that protection might be encouraged by highlighting risks. Wieneke et al. (2016) came to a similar conclusion in their wearable study. They conducted interviews with 22 users, finding low awareness of the threats. Decision-making was deemed to be irrational, and imparting knowledge might address this issue.

Udoh and Alkharashi (2016) analyzed privacy awareness and smartwatch actions. Through interviews with 10 students, they found that most did little to protect their data. Indeed, Rheingans et al. (2016) opined that behavior is driven by pleasure rather than calculations of utility. Although their survey targeted activity trackers, wearables were highlighted as 'hedonic' technologies.

Due to their advanced functionality, smartwatches can access emails, text messages, contact details and health recordings (Do et al., 2017). Although privacy features are available, such as app permissions, it appears such tools are rarely used (Udoh and Alkharashi, 2016). Our work aims to augment the aforementioned studies, comparing concerns and behavior at a large scale. Furthermore, in addition to analyzing the topic, we seek to encourage protective behavior through our serious game.

Smartwatch games. Smartwatches present a challenging environment, possessing small screens and few buttons. Indeed, Chun et al. (2018) found these devices suffer from several usability issues. Casano et al. (2016) did develop 'Estimate It!': an app which sought to teach measurement and geometry. It was assessed by a pool of experts and a 7-person student sample. While users were

engaged with the gameplay, the interface posed usability challenges. Arroyo et al. (2016) evaluated their watch game through a pretest-posttest design with 96 students. It was found to contribute to a greater enjoyment of mathematics. However, both these applications were used to augment real-life activities. In contrast, we seek to address behavior on the smartwatch. Since the public place data at risk, action must be taken to mitigate the issue.

Critical reflection. While applications might teach protection, games can also place privacy at risk (Cybulski, 2014). Indeed, covert surveillance has been proposed through popular platforms such as World of Warcraft (Hope, 2014). Simulations often track behavior and collect large quantities of data (Hulsey and Reeves, 2014). Players are routinely monitored while the engine anticipates their next move (Whitson and Simon, 2014). We recognize this tension and sought to minimize data collection. Participants were also anonymous and gave informed consent at the start of the study. Through playing our simulation, individuals might develop familiarity with smartwatch apps. Since such apps can pose risks (Schneier, 2015), we accept limitations to our approach. However, we expect the privacy knowledge to outweigh the negative effects.

3. Game Design

3.1. Overview

Narrative. In our game, *(Smart)Watch Out!*, players must navigate their character to a local shop. We selected simple real-world metaphors to align virtual events to physical risks (Garg and Camp, 2012). En route, players collect coins to pay for their purchases. Their town is populated by two types of Non-Player Character (NPC): 'villagers' and 'thieves'. As highlighted in Section 2, these are included to enhance immersion (Maldonado et al., 2005). 'Villagers' ask the player privacy-related questions, while 'thieves' try to violate their privacy. The thieves trigger challenges, and the game ends if settings are not configured in the allotted time. Players win by reaching the shop, having successfully overcome the thieves. Game screenshots can be found in Figure 1.

Mechanics. The online game takes place on a 2D map, similar in style to *Pokémon*. Players control the interface like a smartwatch: they click to tap, drag to swipe and click the side button when required. As previously highlighted, context can be important for encouraging engagement (Steiner et al., 2009). Therefore, we sought to simulate the native environment closely.

Players begin by selecting a three-letter name, before configuring their character. Users can toggle their character's gender, skin color and hair color. This

Figure 1: Smartwatch game: Privacy challenge (left), main gameplay (right)

customization was included to enhance engagement (Annetta, 2010), as discussed in Section 2. Players then navigate their character through the map, interacting with villagers and thieves.

Questions. Villagers ask privacy questions and reward correct answers with coins. For example, a person might be asked, "*How can I protect my data from smartwatch thieves?*", and respond with "*Enable passcode*". Rather than being punished for wrong responses, users are given corrective information. We provided this 'knowledge of correct results' to enhance the learning of our participants (Benton et al., 2018).

Challenges. Thieves trigger privacy challenges, which require players to update the interface's settings. Users navigate the interface and perform configuration, just as on a real smartwatch. Challenges consisted of enabling a screen lock, disabling GPS and restricting SMS permissions. For example, a player might go to *Settings* $\gg$ *Permissions* $\gg$ *BadApp* $\gg$ *Revoke SMS* to complete the final task. All questions and challenges corresponded with the protective features, establish-

ing a relationship between concerns, behavior and gameplay.

Privacy challenges were all time-limited, with health decreasing as this period elapses. This was used to encourage engagement and incentivise the memorization of menu screens. We appreciate that health bars often increase as characters progress through a game. However, we reward success through coins and use health to time-limit our challenges. When users encounter a privacy challenge, they are allowed to decline it. As highlighted in Section 2, incentivization is more influential than coercion (Bada et al., 2015), and most points are gained through challenge completion. The faster participants adjust their settings, the more coins they receive. This positive reinforcement (Kumar, 2013) was included to encourage the memorization of the privacy settings.

3.2. Learning Principles

Behavior change. Our game seeks to impart knowledge, with this additional information leading to behavior change. Therefore, our goals of learning and influence should not be seen as mutually exclusive. As previously mentioned, Sasse et al. (2007) theorized that behavior change should be delivered through three steps. Individuals should be made aware of an issue, then given information, and then granted opportunities to practice new actions. Our interactive game sought to follow this approach. The privacy challenges first highlighted potential issues. Then the questions and feedback provided protective information. Finally, participants practiced watch configuration when undertaking the challenges. Based on task success and failure, a helpful feedback loop was established. By adopting this detailed approach, we hope to encourage protective behavior.

PMT. We adopted Protection Motivation Theory (PMT), partially due to its success in security-themed intervention studies (Albayram, Khan and Jensen, 2017). We seek to emphasize the threat component, highlighting the consequences of violation and the prevalence of risk. By confronting players with privacy challenges, we aim to make issues more salient. The benefits of protection are also explained, as we hope to increase response efficacy. Self-efficacy is crucial, since it concerns the confidence individuals have in their ability to protect themselves (Bandura, 1977). Through experimenting on the simulated interface, users should gain self-belief. Finally, by showing how simple the features are, we aim to reduce perceived response costs. This holistic approach should encourage protection and reduce the Paradox.

Platform. The app is defined as an 'operative game', in that it "*leverages knowledge gained from the study of games or play to exert control upon the world,*

such as encouraging exercise or learning" (Carter et al., 2014). Rather than hosting the app on a single smartwatch, the interface was simulated online. This approach was chosen for several reasons. Firstly, it allowed us to assess the application with a larger and more-diverse sample. Secondly, it also let users experiment and explore without risking their own devices (Beard, 2010). Finally, it ensured that the game was presented consistently, rather than on a range of Wear OS models. While users might feel more engaged on a real smartwatch, our designs can inform native apps.

Principles. We designed our game based on below techniques from Learning Science and HCI. While this work is empirical rather than theoretical, we sought to root the game on sound principles. We implemented the four learning science principles of goal-oriented, challenging, contextual and interactive (Quinn, 2005). This was achieved through our (1) privacy tasks, (2) increased difficulty, (3) simulated interface and (4) game interactivity. We also subscribed to the six principles of educational game design (Annetta, 2010). This was done through named avatars (unique identity), the narrative (immersion), user interaction (interactivity), challenges (increased complexity), scores (informed teaching) and feedback (instructional). As these techniques informed our app, it had every opportunity to be influential. While our primary focus was to encourage protection, we hope our design can inform future smartwatch games.

Game design goals. Through the implementation of PMT principles, our app sought to encourage protective behavior. To achieve this, we aimed to increase self-efficacy (through allowing practice), highlight response efficacy (through providing information), enhance risk perception (through privacy education) and reduce perceived response costs (through demonstrating simplicity). Although these goals were important, they were selected as vehicles of behavior change. If our game can affect these PMT components, protective actions should result.

3.3. Pilot Testing

To refine our application, we conducted pilot testing through the Prolific crowdsourcing service. This platform will be described in greater detail in Section 5.1. Using a distinct sample from the main study (to avoid learning effects), we recruited 30 users. Participants proceeded through the session in the standard manner by watching an introductory video, playing the game and completing an evaluation questionnaire. The materials and procedure is outlined in the next section. Since demographics are solicited in pretest, these details are not available for pilot participants. Based on survey results, 60% enjoyed the game, and 60% found it

usable. This was lower than expected, and fortunately the form solicited qualitative feedback. The most frequent complaints cited difficult controls. In response, we added an on-screen buttonpad to assist touchscreen users. This refined version was used in the final study.

4. Research Context and Hypotheses

4.1. Threat Model

Within this work, we analyse the privacy concerns and actions of smartwatch users. Through this, we explore both behavior change and the Privacy Paradox. Concerns are relative to their particular context, and therefore it is crucial that we outline our scope. While an incident might trigger displeasure, that reaction is irrelevant if the issue lacks feasibility. A threat model describes the risks that are deemed credible within a particular context (Myagmar et al., 2005).

In our model, we assume individuals own a smartwatch with a number of apps. App developers legitimately collect data (constrained by permissions) and often share this (potentially anonymized) data with commercial partners (Schneier, 2015). The watch will have access to GPS, whether natively or synched, allowing the use of navigational features. However, when enabled, this allows real-time identification of the watch's location (Ashbrook and Starner, 2003). Smartwatches are small, valuable and visible. Hence, like most consumer technologies, they face some risk of loss/theft (Matthews, 2016). Indeed, as expressed by Ricci et al. (2018), their "*size and portability makes them easy to steal*". Based on our model, we consider privacy threats from tech companies and petty thieves.

4.2. Smartwatch Privacy Features

Since we analyse behavior, it is also wise to highlight the protective features available. We scoped our focus to Wear OS[3] devices, as these products offer broad privacy functionality. Furthermore, as the smartwatch market is varied, it was wise to consider a defined ecosystem. Android watches possessed three relevant features: *screen locks*, (disabling) *GPS* and *app permissions*. Rather than mandating protective behavior, we sought to encourage users to appreciate the benefits of privacy. It was crucial to avoid paternalism, and to accept that tools are

[3]Wear OS, known formally as 'Wear OS by Google', was previously branded as 'Android Wear'. All these devices support the discussed privacy features. GPS can be native or through a synched smartphone.

not necessary in every situation. As will be justified in Section 5.2, behavior was reported on a Likert Scale from Always to Never.

Screen locks. Wear OS watches currently support screen locks through graphical patterns. While they pose a delay, they reduce the risk a lost device is accessed. Smartwatches are small, popular, consumer-focused and valuable: the perfect target for thieves. Screen locks are recommended on phones to mitigate the theft risk (Consumer Reports, 2014). Since the Android OS is similar on watches, we expect this threat to also be reduced. Criminals might steal a device primarily for its hardware value. However, if they can access personal data (e.g., text messages, contacts, emails), they could cause additional harm. Therefore, locks are a first line of defense in keeping data private.

GPS. GPS is beneficial for navigation and sporting activities. But, by definition, it allows a watch's position to be identified in real time (Ashbrook and Starner, 2003). Smartwatch tracking might be highly personal, since the sensor is attached to its owner's wrist. If an individual spends eight hours at a particular location, it might be inferred that this is their home or workplace. By combining pieces of data, a company could build a detailed view of a user's life (Creese et al., 2012; Aktypi et al., 2017). We do not call for the complete avoidance of GPS; it is a key feature of smartwatches. However, by disabling it when not required, privacy can be protected.

App permissions. Smartwatches can contain sensitive information, from text messages to contact details (Do et al., 2017). While apps might access these records legitimately, data is often shared with third-parties (Eadicicco, 2014). Fortunately, as on smartphones, Wear OS allows users to toggle permissions. If apps cannot access watch data, they cannot read details or share them further. Although this can constrain functionality, many users accept permissions without considering the risks (Felt, Ha, Egelman, Haney, Chin and Wagner, 2012).

4.3. Concerns and Behavior

In this study, we analyse both privacy concern and behavior. However, to assess the Paradox appropriately, correspondence should exist between our questions. If a feature does little to assuage a concern, then its usage is irrelevant to our study. Therefore, to support the *principle of compatibility* (Ajzen, 1988), we considered incidents in which tools could reduce a risk. When evaluating privacy concerns, we solicited responses to privacy scenarios. This popular approach has been adopted in previous studies of wearable concern (Lee et al., 2016). It also avoids mentioning the topic, and hence should mitigate priming (Braunstein et al., 2011). As will be justified in Section 5.2, participants indicated their response

from Very Pleased to Very Displeased. We now outline the scenarios (*data access*, *location tracking* and *data sharing*) and highlight their correspondence with the protective features.

Data access. As previously mentioned, screen locks should mitigate the risk of watch theft. Therefore, to gauge concern, participants considered a stranger accessing their device's data. These concerns were compared against how consistently a lock was enabled. Users might be opposed to any form of criminality. However, if data is accessible, further harm can be caused. If a person fears this incident and yet neglects protection, their behavior might not align with concern.

Location tracking. While GPS offers helpful functionality, it can allow companies to monitor a watch's location (Troiano, 2016). We asked respondents to consider this eventuality and provide their reaction. To assess behavior, we solicited how often 'location services' were used. This term was preferred to 'GPS' for comprehensibility. Since usage can place privacy risk (rather than protecting it), user responses were reverse-coded. Again, we do not call for these services to be wholly avoided. However, if a person uses GPS while opposing monitoring, a concern-behavior disparity might exist.

Data sharing. For apps to function, companies legitimately collect watch data. These details can then be shared with third parties, as is common in the digital economy (Schneier, 2015). Permission settings are one of the few means of restricting data collection (Eadicicco, 2014). As a scenario, we asked participants to consider their personal details being shared. Such concerns were compared against how often they adjusted their permissions. If a person fears sharing yet rarely acts, a disparity might be present.

Correspondence. Through our scenarios and protective features, we establish a conceptual link between concern and behavior. This connection is illustrated in Figure 2, with concerns denoted by Cs and behavior frequencies denoted by Bs. The arrows highlight how behavior influences risk, which can then influence concerns. A negative scenario (e.g., location tracking) might generate concern, with that concern partially informed by probability and severity. However, actions might be available (e.g., disabling GPS) which can reduce either the likelihood or the damage. If a person often uses this protection, it should mitigate the risk. To examine the Paradox, we compare the degree of concern (C) with the regularity of protection (B). This approach seeks to ensure the *principle of compatibility* is respected (Ajzen, 1988).

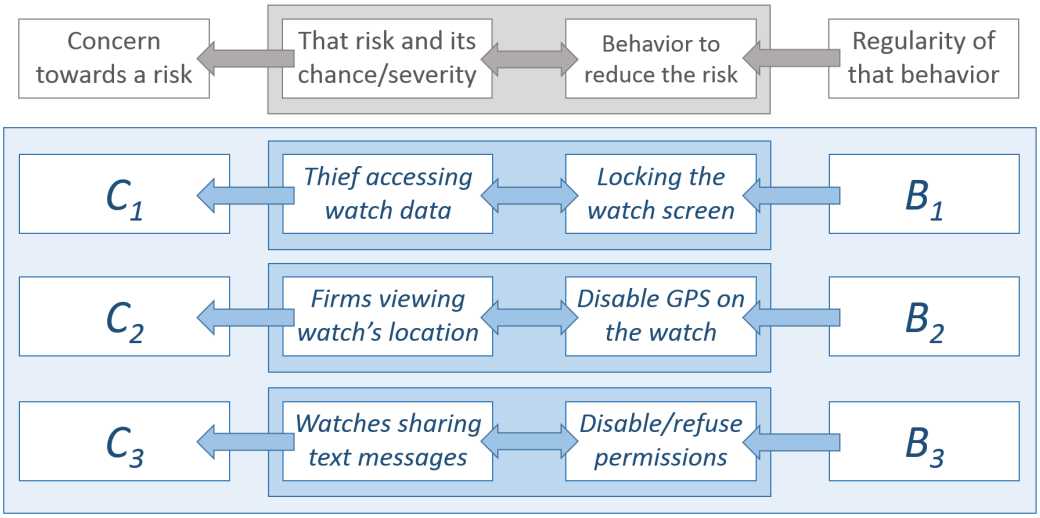

Figure 2: Relationship between privacy concern, risk and behavior

4.4. Hypotheses

As will be outlined in Section 5.3, we analysed the concerns and behavior of a treatment group (who played the smartwatch game) and a control group (who did not). They complete questionnaires both prior to the intervention (pretest) and after this stage (posttest). Ordinal responses were translated into concern and behavior scores, as outlined in Section 5.2. Through this, we explored the following six hypotheses.

Concerns. Although data protection is an important topic, it often lacks salience (Williams et al., 2016*a*). By highlighting the risks of wearable devices, privacy threats might appear more tangible. Therefore, we asserted that our treatment group would express greater concerns after gameplay. To ensure the validity of our study, changes should owe to the game rather than external factors. Thus, we also assessed our control group responses. We posited that their concerns should not increase from pretest to posttest. Null hypothesis significance testing is a matter of continued debate (Nickerson, 2000). However, we do not test for equivalence; we only care that differences do not reach the significance threshold. Concerns are measured by mean score, appropriate since Likert responses were aggregated over related questions (Norman, 2010; Carifio and Perla, 2008).

Behavior. Previous work suggests that risk awareness can promote protective actions (Deuker, 2009). Through imparting knowledge, we should encourage users to guard their data. By exploring the frequency of such actions, we can as-

sess whether behavior has improved. Since behavior change is challenging, and underexplored on smartwatches, we deem this to be our primary contribution. We posited that our treatment group would employ protective behaviors more often after gameplay. We also asserted that this frequency would not increase for our control group. Behavior is measured by mean score, calculated in a similar manner to the concerns.

Paradox. Since a lack of awareness can contribute to the Privacy Paradox (Deuker, 2009), information and practice should mitigate the issue. We defined the Paradox as a 2-point disparity (/5) between the concern and behavior scores. We did not believe differences of 1 point signaled a dissonance, but thought a 3-point definition was extreme. While we expected concerns to be awoken by gameplay, we predicted they would increase less than behavior scores. Therefore, we posited that the disparity would be less prevalent in posttest. If so, this might suggest that the application can help reduce the Privacy Paradox. We also asserted that prevalence would not decrease in our control group. This would imply that our questionnaires did not introduce significant bias.

Research questions. While not falsifiable hypotheses, we conducted qualitative analysis driven by two research questions. Firstly, we wished to explore the privacy rationale of smartwatch owners. By understanding the decision-making processes of these individuals, we could explore the impact of our gameplay principles. Secondly, since it is unlikely that all participants would be influenced, we wish to assess why users do not guard their privacy. Through this information, we seek to inform the development of future interventions.

Summary. In summary, our hypotheses and research questions are:

H1a. The posttest mean concern score of the treatment group will be significantly higher than its pretest score.

H1b. The posttest mean concern score of the control group will *not* be significantly higher than its pretest score.

H2a. The posttest mean behavior score of the treatment group will be significantly higher than its pretest score.

H2b. The posttest mean behavior score of the control group will *not* be significantly higher than its pretest score.

H3a. The posttest Paradox prevalence of the treatment group will be significantly lower than its pretest prevalence.

H3b. The posttest Paradox prevalence of the control group will *not* be significantly lower than its pretest prevalence.

RQ1. What is the privacy rationale of smartwatch owners?

RQ2. Why do some smartwatch users not protect their smartwatch privacy?

5. Method

5.1. Participants

Demographics. We screened for adults owning smartwatches, as they could report their behavior. 504 participants completed the necessary stages, with the total divided equally between the treatment group and control group. While treatment participants took part in all three stages (including the game session), the control group completed the pretest and posttest questionnaires. Of the 504 individuals, 56% were male, 43% were female and 1% identified as non-binary. 73% were under 36, with an estimated mean age of 31.7. This aligns well with our target population, since smartwatch users have been noted as disproportionally male and young (NPD Connected Intelligence, 2014). The sample was also well-educated, since 63% had at least a degree. This enhances external validity, since smartwatch owners tend to have higher qualifications (Desarnauts, 2016).

Randomization. One week after the first questionnaire, a randomized half of our sample participated in the game session. To ensure our randomization was sufficient, we compared our treatment and control groups. This was important, as differences in responses could owe to demographics. In this case, it would be challenging to determine the influence of the game.

We found no significant differences in gender ($X^2(2) = 2.53$, $p = .283$), age ($U = 30046$, $p = .262$) or education level ($U = 31479.5$, $p = .860$). There was also no difference between the pretest concern scores ($U = 28988$, $p = .070$), behavior scores ($U = 30156.5$, $p = .325$) or disparity prevalence ($X^2(1) = 1.16$, $p = .281$). We therefore conclude that our groups were sufficiently randomized, and this should support fair posttest comparisons.

Recruitment. Recruitment was undertaken through the Prolific crowdsourcing service[4]. We screened for adult smartwatch owners with an approval rating of at least 90% (the highest filter available). Such filtering is deemed good practice to maximize the reliability of responses (Peer et al., 2014). While this sample is not wholly representative of smartwatch users, Prolific has a more-diverse participant pool than MTurk (Peer et al., 2017). Prolific members have also been found to be less dishonest (Peer et al., 2017), reducing the risk of false reporting. There may be self-selection bias, since individuals chose whether to participate.

[4] https://prolific.ac/

But since the surveys did not ostensibly concern privacy, owners should not have been dissuaded by such issues. At the start, a study link was made visible to the pool of ∼1,000 valid users. Those who joined (i.e. the participants) completed the pretest survey and gave informed consent. One week later, half the participants were randomly allocated to the treatment group. They took part in the game session. After a further week, all respondents were sent the final survey. Individuals were fully informed of their compensation: $1.80 for stage one, $5.20 for stage two and $1.30 for stage three.

5.2. Materials

Instruments. To evaluate concerns and behavior, we made use of online questionnaires. This approach was chosen for its suitability to large-sample remote studies (Evans and Mathur, 2005).

When designing the experiment, we considered using existing instruments. For example, the Concern for Information Privacy (CFIP) scale (Smith et al., 1996) has been adopted in many privacy works (Culnan and Armstrong, 1999; Chellappa and Sin, 2005). Of greatest popularity is the Internet User's Information Privacy Concern (IUIPC) questionnaire (Malhotra et al., 2004). This instrument, which considers control, collection and awareness, seeks to gauge broad opinions. However, existing scales did not sufficiently capture smartwatch privacy. Lee et al. (2016) developed a rare form for wearable concerns. Although their scale was generic, their scenarios focused on headworn devices. Wear OS behavior is platform-specific and requires specialized questions. Therefore, we constructed a smartwatch instrument based on Lee et al.'s (2016) work. We also selected privacy features that were consistently found across these devices. Lee et al.'s measure was itself adapted from a highly-cited evaluation of smartphone concerns (Felt, Egelman and Wagner, 2012). With the scale being suitable for mobile devices, it seemed apt to extend it to smartwatches.

Design. It was important that instruments did not introduce response biases. To reduce this risk, the queries were carefully refined over several weeks. Furthermore, the same forms were used in pretest and posttest, enabling a fair comparison. Responses can be biased when privacy is primed (Braunstein et al., 2011). Therefore, we included six decoy questions and shuffled the query sequence. This should also reduce the risk of response fabrication, as users should not gauge what answers are expected.

Validation. Once the form was drafted, it received face validation (Collingridge, 2014). This consisted of review by a privacy expert and a psychometrician. The

former certified that the topic was addressed, while the latter suggested further refinements. Finally, we conducted a pilot study with four smartwatch users. They completed the questionnaire and then discussed their interpretation of the queries. Since the measure appeared to be understandable, it was then retained for the real study. This validation process followed the approach of Felt et al.'s (2012) well-cited concern evaluation.

Overview. The 3-minute form (used in pretest and posttest) was comprised of demographics, concerns and behavior. These questions are shown below in Table 1 (numbers denote correspondence and privacy queries are italicized).

Table 1: Pretest/Posttest Questionnaire

#	Demographics
1	With which gender do you most identify?
[Male, Female, Other]	
2	What is your age?
[18-25, 26-35, 36-45, 46-55, 56-65, 66+]	
3	What is your highest level of education?
[School/GCSE, A-Level/College, Degree, Masters, PhD]	
#	Concern Scenarios: Very Pleased, Pleased, Neutral, Displeased, Very Displeased
C1	How would you feel if your watch's screen became dirty?
C2	*How would you feel if a thief viewed all the data on your watch?*
C3	How would you feel if your watch's battery died during the day?
C4	*How would you feel if other companies were spying on your watch's location?*
C5	How would you feel if your watch's apps slowed down?
C6	*How would you feel if your watch shared your text messages with another company?*
#	Behavior Questions: Always, Often, Occasionally, Rarely, Never
B5	How often do you install your watch's updates?
B6	*How often do you disable/refuse the permissions of watch apps?*
B1	How often do you clean your watch?
B4	*How often do you use location services (GPS) on your watch?*
B3	How often do you charge your watch overnight?
B2	*How often do you use a passcode/pattern lock on your watch?*

Concerns. We solicited concerns by gauging reactions to hypothetical scenar-

ios. This avoided coarse ratings of 'privacy', since the complex topic is not suited to such evaluations (Paine et al., 2007). Incidents were drawn from the potential consequences of lax protection. As highlighted in Section 4.3, these were: *data access*, *location tracking* and *data sharing*. These scenarios were chosen since such events appeared common in the literature. Responses were made on a five-point Likert Scale, from Very Pleased to Very Displeased via Neutral. Although we considered beginning our scale at Neutral, we did not wish to anchor negative responses. This scale also aligns with the approach taken by Lee et al. (2016).

Behavior. These questions corresponded both with the protective features and the concern queries. Our chosen actions (enabling *screen lock*, *disabling GPS* and *adjusting permissions*) should mitigate the privacy risks. In this manner, they seek to address user concerns. Behavior replies were made on a five-point frequency scale, ranging from Always to Never. This avoids a coarse binary distinction and accepts that action might not be consistent. The correspondence between concern and behavior is key, as studies have been criticized for improper comparisons (Trepte et al., 2014). With all queries contextualized around an owned device, we avoided comparing the abstract with the practical.

Rationale. The pretest survey consisted of the aforementioned questions. In posttest, we also included an extra section of qualitative queries. These questions were designed and refined in the same manner as above. Again, we searched for existing measures that could extract privacy rationale. However, none appeared relevant for the smartwatch context. For example, while health contexts are frequently targeted (Boer and Mashamba, 2005; MacDonell et al., 2013), wearables are rarely explored. Our queries, found in Appendix A, appeared on the last page of the online questionnaire. Since page navigation was disabled, earlier responses should not be affected. This ensures pretest and posttest can be fairly compared.

In this section, we explored why participants might (not) be influenced by the game. The questions were structured on Protection Motivation Theory, enabling an analysis of privacy rationale. They also concerned common justifications for lax behavior. These included users not being informed (unaware), having low self-efficacy (unconfident), not recognizing the benefits (unconvinced) and not sensing risk (unconcerned). The rationale questions therefore considered perceived threat, self-efficacy and response efficacy. If participants are uninfluenced, this might suggest that gameplay principles have lacked success.

5.3. Procedure

Study stages. Since we wished to compare privacy behavior before and after the game, we selected a pretest-posttest structure. The approach is summa-

rized below in Figure 3. It was adapted from the methodology of Albayram et al. (2017), who successfully encouraged smartphone locking behavior. We used a similar experimental structure, but evaluated our smartwatch game. While we considered a crossover design (Watson and Marks, 1971), we had concerns with this approach. In such repeated-measure experiments, each participant receives a series of treatments. For example, while one half might receive Intervention A then Intervention B, the other half would get B before A. Since we sought to inform users, learning effects might have biased our later stages (Cleophas, 1990).

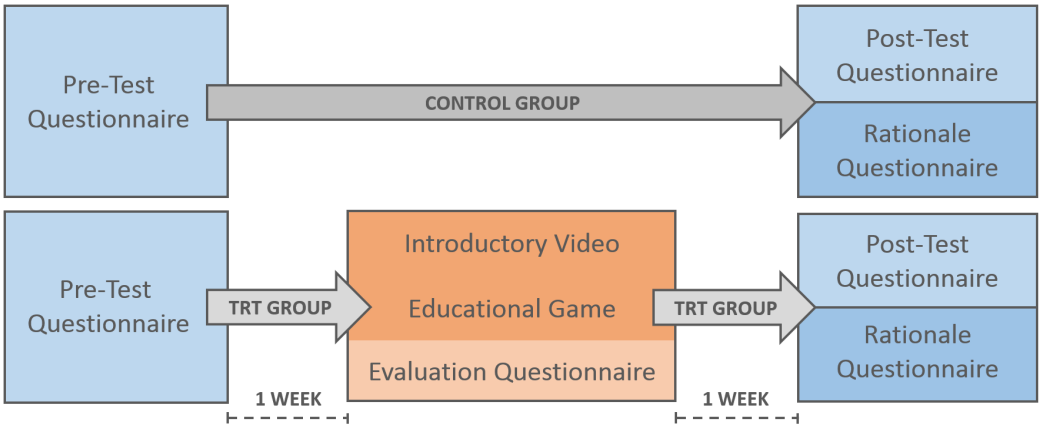

Figure 3: The three-stage experimental structure

The pretest questionnaire was followed by the game session, followed by the posttest questionnaire. Participants were not informed of the second survey, in an attempt to mitigate demand characteristics (Nichols and Maner, 2008). The game session also included a brief questionnaire, enabling the collection of qualitative evaluations. We imposed a one-week gap after this stage, as is common in comparable studies (Albayram, Khan and Jensen, 2017; Albayram, Khan and Fagan, 2017) and security research (Wiedenbeck et al., 2005; DeWitt and Kuljis, 2006; Kumaraguru et al., 2009). This was undertaken to explore whether behavioral lessons were retained. It also sought to reduce priming (Braunstein et al., 2011) and give participants time to put knowledge into action. Since we screened for smartwatch owners, users should be able to continue their experimentation. Study purpose was disguised, participation was anonymous and compensation was allocated equally. Therefore, we mitigated many incentives to falsify behavior. The study also received ethical approval from our institution's IRB prior to conduction.

We used treatment and control groups, with this design known to combat confounding variables (Shuttleworth, 2010). The treatment condition played the

game, while the control group took no action. Both groups completed all pretest and posttest questionnaires, enabling the comparison of responses. The average completion times were 3.45, 37.63 and 6.13 minutes for stages one, two and three, respectively. Since these means are in excess of the compensated durations (3, 25 and 4), we deemed participants to have invested sufficient effort.

Game session. In stage two, we did not seek to dictate actions or reduce a participant's agency. Instead, we provided users with information and an opportunity to trial new behavior. The session consisted of three components. Participants were first shown a brief video. It introduced the game and outlined the mechanics of the privacy challenges. It also demonstrated how to use the interface, seeking to minimize usability issues. The video page can be found in Figure 4. Participants then played the serious game for five minutes. This allowed them to learn new skills and practice protective techniques. Finally, individuals completed a brief questionnaire to evaluate the game. In the interest of brevity, these results are not discussed. There was a risk that participants would complete the session questionnaire while skipping the other components. To combat this, visual codes were embedded within the video and the game. To demonstrate their participation, individuals were required to report these codes.

5.4. Analysis

Quantitative. Since our Likert data was not normally distributed, we opted for non-parametric techniques. We used Spearman's Rank (r_s) to assess correlation between our ordinal variables (Spearman, 1904). For significance testing independent samples, we performed the Mann-Whitney U Test (Mann and Whitney, 1947). This was selected when our dependent variables were ordinal and the independent variables were nominal. For testing two related samples, the Wilcoxon Signed-Rank Test was preferred instead (Wilcoxon, 1945). When the dependent variables were nominal, we made use of the Chi-Squared Test (X^2) (Pearson, 1900). For repeated measures with such nominal variables, we opted for the McNemar Test (McNemar, 1947). As convention, $\bar{x}$ denotes means and σ indicates standard deviation.

Qualitative. Through our posttest questionnaire, we received a rich set of qualitative responses. For such data, we conducted an inductive process of framework analysis (Ritchie et al., 2003). This approach bears similarity to thematic analysis (Braun and Clarke, 2006), but is "*not bound by a particular epistemological position, giving it freedom and flexibility*" (Parkinson et al., 2016). We opted for an inductive process as we sought to avoid bias from latent assumptions (Boy-

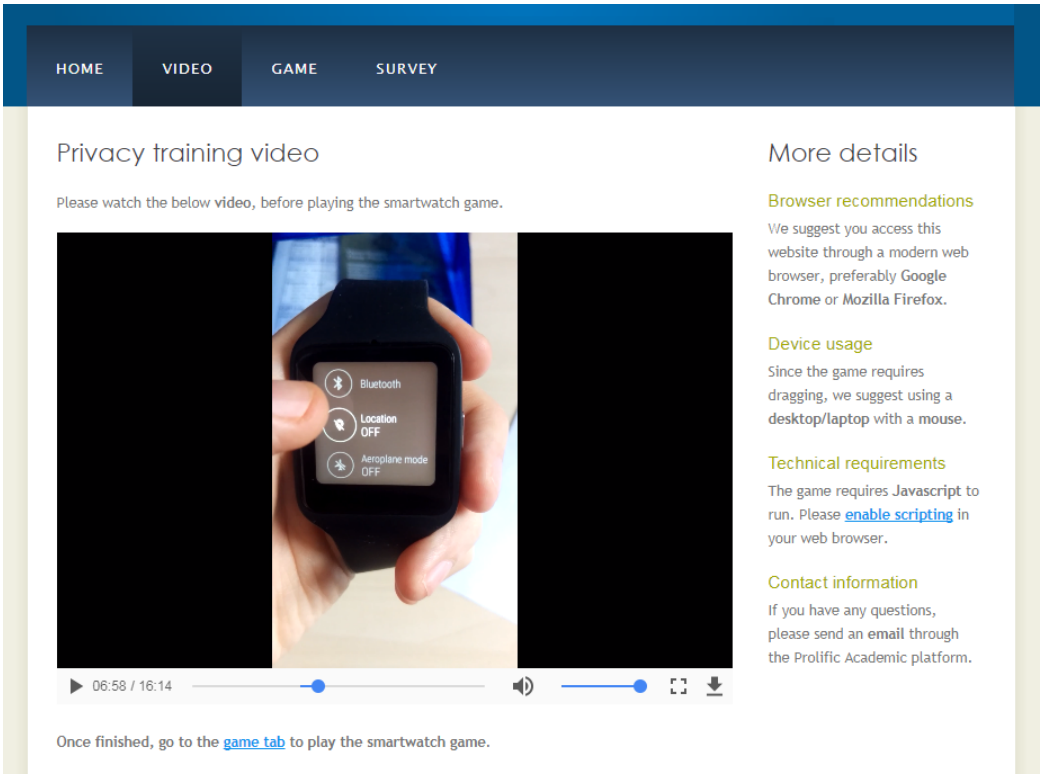

Figure 4: Introductory smartwatch video

atzis, 1998). Although we considered our findings in line with PMT, we ensured coding was data-driven and not moulded to this model.

As a first stage, we familiarized ourselves with the data. The text was read and re-read, enabling initial patterns to be noted. To ensure the richest of analyses, this approach was undertaken for all 504 participants. For the second phase, we annotated responses with high-level labels. Similar codes were then collated, resulting in general themes (stage three). For example, '*Not aware*' and '*Low risk awareness*' were coalesced into '*Lack of awareness*'. This sought to establish consistent categories across an entire dataset. As a fourth stage, these themes were developed into coding indices. Each index sought to structure responses in a hierarchical and consistent manner. For categorization, these indices then formed our coding frames (stage five). To increase the robustness of this process, we developed strict definitions for each theme. As a final stage, responses were categorized and used to inform our conclusions. We also chose vivid examples to serve as in-

teresting extracts (Braun and Clarke, 2006). This allowed us to showcase those participant remarks which exemplify a theme.

6. Results and Discussion

6.1. Concerns, Behavior and the Privacy Paradox

In the interest of brevity, we primarily focus on final responses and pretest-posttest comparisons. This highlights our central contribution: evaluating whether protective behavior can be encouraged. One week after the game session, we distributed the posttest questionnaires to both groups. It should also be noted that current behavior was requested rather than behavioral intent. This enabled exploration of whether protective actions had increased in frequency.

Concerns. Across both groups, 93% expressed posttest opposition to the *data access* scenario. This was represented by a mean response of 4.60 (σ = 0.843), with the scale ranging from Very Pleased (1) to Very Displeased (5). Based on these metrics, users seem to value their watches. Concern appeared to significantly increase in the treatment group (Z = -2.04, p = .042, d = 0.13). Since the game highlighted the potential consequences, individuals might now recognize the risks.

Location tracking also proved unpopular, with 90.7% signaling their displeasure ($\bar{x}$ = 4.50, σ = 0.855). Participants appear to reject the notion of technological monitoring. This provides an interesting contrast to the commonality of location-based services. Despite the gameplay session, responses failed to significantly differ from the pretest percentage (92%, $\bar{x}$ = 4.56, σ = 0.780). When apprehension is near-maximal, it is challenging to enhance concerns.

Grievances over *data sharing* appeared to slightly reduce, though the change was not significant. 96% were (at least) displeased in pretest ($\bar{x}$ = 4.75, σ = 0.667), suggesting this was the most-concerning incident. Again, these results are in tension with the realities of the online economy (Schneier, 2015). This concern decreased to 93.1% after the game session ($\bar{x}$ = 4.65, σ = 0.834). Since the treatment group learned protective techniques, perhaps some felt better prepared.

We then examined our first hypothesis (*H1a*), analysing whether the treatment group's concern score had significantly increased. We found no significant difference between pretest and posttest metrics (H1a: Reject, p = .544). This was counter to our expectations and might indicate that worries are near a maximum. We then explored *H1b*, positing that the control concern score would not increase. Since they did not differ (H1b: Accept, p = .814), our questionnaires should not have introduced significant bias.

Behavior. We now consider behavior, assessing whether our game proved influential. Prior to gameplay, *screen locks* were used often by 57%. This resulted in a sample-wide mean of 3.37 (σ = 1.694), with the scale ranging from Never (1) to Always (5). This suggested that many users did little to protect their smartwatches. Encouragingly, usage significantly increased (Z = -3.54, $p < .001$, d = 0.22) in the treatment group (64%, $\bar{x}$ = 3.66, σ = 1.632). If individuals now recognize the risk, they might appreciate the protection.

By *disabling GPS*, individuals can limit the threat of location monitoring (Rahman, 2013). Although the tracking scenario was strongly opposed, participants seem to enjoy the service. In pretest, only 23% of our sample used GPS rarely ($\bar{x}$ = 2.64, σ = 1.197). This was compared to the 51% who enabled it often. The treatment group then had opportunities to practice protective behavior. However, their GPS usage failed to significantly change ($\bar{x}$ = 2.68, σ = 1.094). Individuals might recognize the risks but still opt for functionality benefits (Lee et al., 2013). This is a valid response, provided users make informed decisions.

Finally, we considered the usage of smartwatch *app permissions*. Before the game session, these settings were often checked by 24% of participants ($\bar{x}$ = 2.82, σ = 0.975). Since loose permissions allow functionality, sharing might have been common. After playing the game, treatment-group usage significantly increased (Z = -3.65, $p < .001$, d = 0.23) to 32% ($\bar{x}$ = 3.11, σ = 1.027). While this remains a minority, it is promising that protection was encouraged. If the risks of sharing can be demonstrated, perhaps users will adjust their behavior.

Our third hypothesis, *H2a*, asserted that the treatment group's behavior score would significantly increase. This was found to be true, with these participants taking greater action (H2a: Accept, $\bar{x}$ = 3.32, σ = 0.83, Z = -4.40, $p < .001$, d = 0.28). Such results support our central contribution: that our game might encourage protective behavior. This might encourage the further development of privacy apps. *H2b* was then examined, which posited that the control behavior score would not increase. Since it failed to differ significantly (H2b: Accept, p = .150), the questionnaires should not have biased responses.

Pretest Paradox. To assess the Privacy Paradox, we compared participants' concerns with their behavior. Before users played the game, a disparity appeared to be prevalent. For *screen lock* protection, respondents' behavior scores were significantly less than their concern scores (Z = -12.43, $p < .001$, d = 0.85). While they might want protection, the action might be deemed inconvenient. When considering *location tracking*, behavior scores were also lower than concern scores (Z = -17.39, $p < .001$, d = 1.31). Again, users might have appreciated the functionality without gauging the risks. Finally, the same concern-behavior disparity

was found for *permissions* (Z = -18.10, $p < .001$, d = 1.39). Many individuals might genuinely oppose this scenario. However, they might have little awareness of when data is accessed.

Across the three features, individuals acted significantly less privacy-conscious than their concerns (Z = -18.60, $p < .001$, d = 1.45). This large effect size suggests a disparity might be present. A concern-behavior heatmap is found below in Figure 5, with the white line denoting the Paradox. It illustrates how participants were largely concerned, yet protection was inconsistent. Based on our 2-point definition, 43.3% (218/504) were potentially susceptible to the issue. This echoes Williams et al. (2017) in suggesting the Paradox is prevalent for smartwatches.

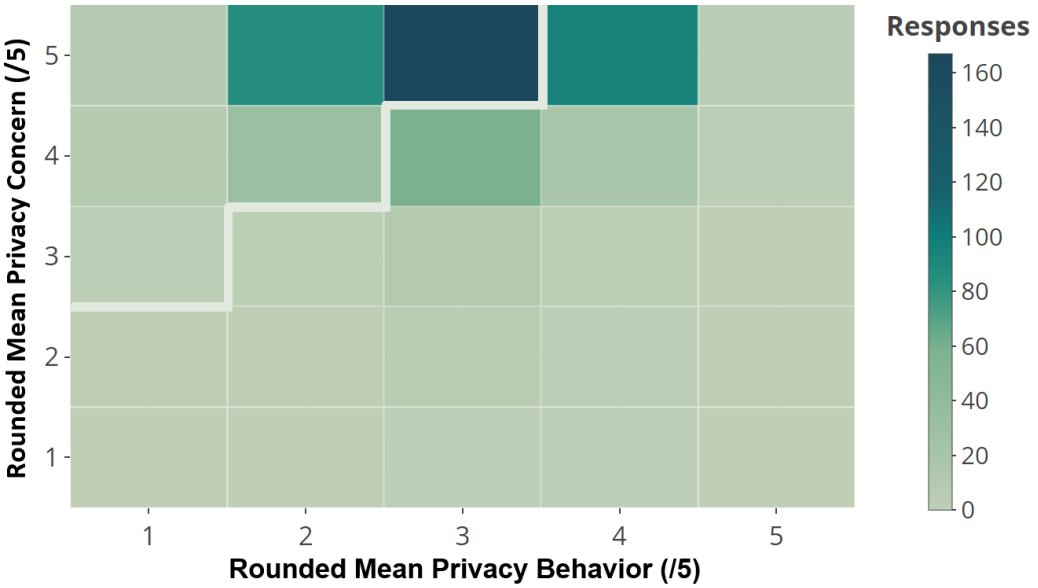

Figure 5: Distribution of pretest privacy concerns and behavior

Posttest Paradox. We can assess the Paradox through two techniques. In the first, we consider prevalence: the quantity of (at least) 2-point disparities. This analyzes how common the issue is in a group of individuals. In the second, we measure magnitude: the gap between concern and behavior scores. This assesses the extent to which the components diverge.

To explore *H3a*, we analyzed whether the Privacy Paradox decreased in the treatment group. The Paradox was indeed both less prevalent (H3a: Accept, $X^2(2)$ = 17.58, $p < .001$, V = 0.19) and reduced in magnitude (Z = -3.58, $p < .001$, d =

0.23). While 45.6% were prone in pretest, this decreased to only 29.4%. While the game is not a panacea, it appears to successfully mitigate the issue.

We then considered *H3b*, which posited that Paradox prevalence would not decrease in the control group. In this case, neither the prevalence nor magnitude significantly decreased (H3b: Accept, $p = .801$ and $p = .345$). 40.9% were prone in pretest, and 39.7% continued to be afterwards. The control group served its purpose, since extraneous variables appeared to have little influence. Therefore, our game might have mitigated the Privacy Paradox. Even if partially effective, such apps could provide a complement to awareness campaigns. In Figure 6, we present the prevalence across both groups. While the treatment group (randomly) began with more instances, it removed far more through the study.

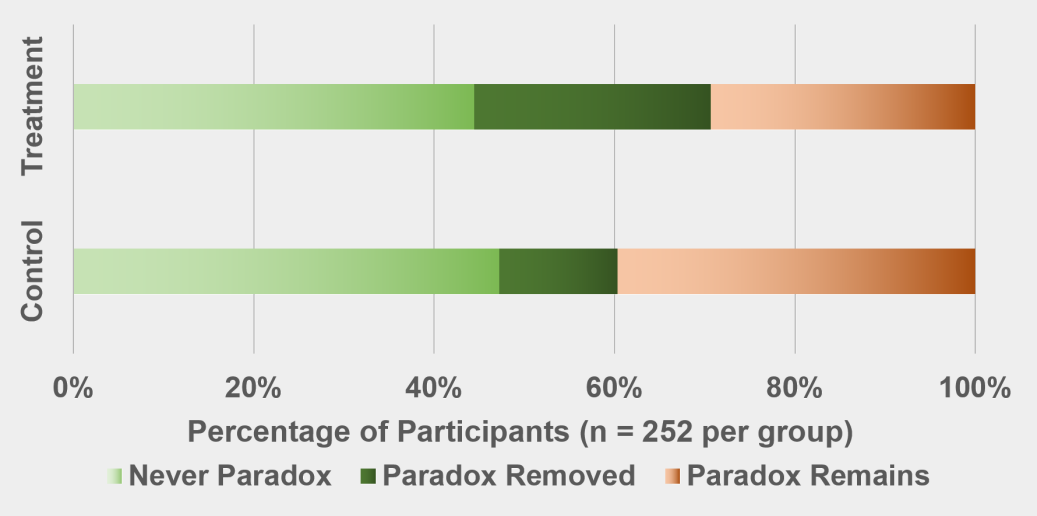

Figure 6: Privacy Paradox prevalence throughout the study

While our quantitative findings are enlightening, we move on to consider privacy rationale. With few studies conducted of smartwatch owners, this grants a rare opportunity to examine their views.

6.2. Behavioral Rationale

The posttest questionnaire also included qualitative queries, enabling a dissection of rationale. Participant responses were coded through inductive analysis, ensuring our findings were data-driven (Boyatzis, 1998). Following coding, we considered the response ratios through Protection Motivation Theory (PMT). This was used to explore why participants did (or did not) change their behavior. By

understanding the game's strengths and weaknesses, we can assess the influence of our principles. We can also then inform the development of future applications.

Self-efficacy. We first asked users if they believed they understood protection, testing whether they felt informed. This assessed both self-efficacy and the learning aspects of our game. 75% of the treatment group answered in the affirmative, significantly more than in the control group ($X^2(2) = 42.00$, $p < .001$, $V = 0.29$). This suggests that the app was successful in imparting knowledge. When asked why, 26% of treatment participants said they had now taken action. Of those who did not understand, 44% felt under-informed and 11% deemed it too complex. This indicates that existing approaches are not a panacea for increasing skills. Our game did not include large quantities of information (to keep gameplay engaging), but the content could be increased if necessary. Representative quotes, along with the participant number, are shown below:

> "*I work in IT and have a good knowledge of data protection*" (#47, Do you understand? - Yes).

> "*I've never kept myself informed about it so I don't think I know how*" (#196, Do you understand? - No).

To further assess self-efficacy, we asked users whether they felt confident they could protect themselves. If so, this indicates that they feel prepared, even if this confidence is misplaced. 64% of the treatment group responded 'Yes', again significantly more than in the control group ($X^2(2) = 21.78$, $p < .001$, $V = 0.21$). As expected, this suggests the game encouraged greater confidence. Of those answering in the affirmative, 29% said they had now taken action. Of those responding 'No', 36% reported insufficient knowledge and 24% thought adversaries were too skilled. Such views could lead to 'privacy fatalism', with users failing to invest the effort. In response, it might help to showcase the limits to hackers' capabilities.

> "*I passcode the watch and disable certain features*" (#335, Are you confident that you can protect your smartwatch? - Yes).

> "*I think that there are much smarter hackers out there than me*" (#206, Are you confident that you can protect your smartwatch? - No).

Response costs. Individuals have previously doubted whether privacy is worth the effort (Williams et al., 2017). While protection is desirable, often it is traded for convenience or short-term necessity. Therefore, we asked our sample whether

they believed the benefits outweighed the bother. This sought to gauge the response costs of protection; another PMT component. 86% of the treatment group agreed, compared to 82% of control participants. The proportions failed to differ, and this component appears in little need of attention. Curiously, those answering 'Yes' were significantly less likely to display the Paradox ($X^2(2) = 6.05$, $p = .049$, $V = 0.11$). This suggests that we should advertise privacy benefits, particularly since its intangibility can limit salience (Jackson et al., 2005). Of those who judged little benefit, 40% believed their data was innocuous. Therefore, revised applications could focus on the inferences possible from apparently-trivial data (Creese et al., 2012).

> "*Protection always better than cure, so it is well worth the effort*" (#19, Is privacy worth the effort? - Yes).

> "*I don't see why anyone would care about my personal smartwatch data*" (#226, Is privacy worth the effort? - No).

Perceived threat. To gauge the perceived threat, one half of the PMT model, we asked users whether they felt at risk. Only 32% of the control group agreed, significantly less than the 49% of treatment participants ($X^2(2) = 15.27$, $p < .001$, $V = 0.18$). Again, this suggests that the game was influential in adjusting behavioral rationale. However, still over half of our treatment group doubted the threat. In refined versions, threats will be given greater emphasis. Of those who perceived an issue, 20% blamed sophisticated hackers. Of those doubting the threat, 31% thought their smartwatch was innocuous. Since many undervalue their data, this encourages the signposting of inference capabilities.

> "*People can steal information that others wouldn't think is important*" (#181, Are you under threat? - Yes).

> "*I've got nothing worth spying on*" (#268, Are you under threat? - No).

6.3. Suggested Further Remediation

For those participants who played the game, the Privacy Paradox became less prevalent. However, many users continued to act less privately than their concerns might warrant. Therefore, we analyzed the justifications given by those who still displayed a disparity. Through examining responses to our PMT queries, we refined recommendations for future approaches.

Protective feature demonstration. Of those displaying a disparity, 56.3% still claimed to understand protection. The most frequent justification (30.6%) for this was that they possessed technical expertise. This suggests that many are aware of protective features, but choose not to act. In these cases, they likely fail to perceive a privacy risk. While these users might be hard to convince, at least they possess technical understanding. If their risk exposure can be highlighted, it might trigger protective behavior. Of those who did not understand, 62.0% expressed that they were insufficiently informed. Some of these would have played the game, indicating that apps are not a panacea. For such users, feature demonstrations might be a wise approach.

Only 48.3% of disparity-prone users were confident in protecting themselves. 61.1% of these blamed a lack of information, implying that further content must be provided. If doubts stem from insufficient knowledge, we recommend demonstrations to showcase privacy-protecting features (e.g., app permissions). These videos could be embedded within our serious game, centralising the information in a single location. By watching demonstrations of smartwatch configuration, unsure individuals should gain self-efficacy. The success of such measures could be ascertained through follow-up challenges and questions. If users can navigate menus after viewing this content, the games might have imparted knowledge.

Highlight risk exposure. Even of those expressing the Paradox, 79.9% claimed privacy was worth the effort. Over 30% of those valued its protection, while most negative respondents doubted sensitivity. It seems that individuals will voice their support for privacy, even if they fail to act. This supports the view that social norms contribute to the disparity (Blank et al., 2014). If we highlight the actual risk, users might follow up on their claims.

When asked if they felt under threat, 39.7% of disparity-prone users responded 'Yes'. This small proportion was expected, as individuals are unlikely to act if they feel secure. Again, the most-popular justification was a perceived lack of data sensitivity (37.1%). Of those who did feel under threat, over 40% blamed hackers and their techniques. This creates an intriguing contrast: some users think their data is innocuous, while others believe they are vulnerable. Again, this suggests that we should inform individuals of criminals' capabilities (Nurse, 2018).

Our prototype game was web-based and hence could not assess installed applications. In updated versions, we could incorporate risk evaluations. The game would sit directly on the smartwatch, enabling the monitoring of apps and behavior. It might then distinguish sensitive data from trivial details, helping users manage the threat. Based on current settings, the game could deduce the inferences and include these within privacy challenges. Such an approach might both

demonstrate capabilities and enhance the realism of the game. To examine our success, watch settings could be assessed over time. If individuals begin restricting their permissions, they might have learned to value protection.

6.4. Results Summary

Findings. Prior to gameplay, privacy-protective behavior was far from consistent. In pretest, less than 60% used passwords often, while permissions were accessed rarely. Despite this lack of action, participants claimed to be concerned about their privacy. After playing our game, treatment-group behavior became more protective. Indeed, their usage of both screen locks and permissions significantly increased. This appeared to mitigate the Privacy Paradox. In contrast, the behavior of the control group did not significantly adjust. This suggests that our intervention game might have proved influential.

Principles. Through qualitative questions, we then assessed the efficacy of our principles. We sought to enhance self-efficacy by providing participants with protective knowledge. This appeared successful, since our treatment group were both more confident and claimed greater understanding (than the control group). Our game also aimed to highlight the ease of protection, since response costs are a key factor. Again, treatment participants were more likely to believe privacy was worth the effort. Finally, we assessed whether individuals felt their data was under threat. As the treatment group were more convinced, our game might have enhanced their threat perception.

Enhancements. While many participants adjusted their behavior, others did not. This suggests that the game could benefit from further development. Of those still prone to the Privacy Paradox, a large proportion believed they were insufficiently informed. Many also lacked confidence in protection, again justified by a lack of knowledge. To address this, the game could be enhanced through an embedded video. Through demonstrations of behavior, users should gain familiarity with guarding their watches.

Although most believed that privacy was worth it, many deemed their data innocuous. Similarly, when queried on their threat perception, some participants thought the risk was low. It is possible that these users are not aware of advanced inference techniques. To enhance our game, we could adapt gameplay based on installed applications. By highlighting each participant's risk exposure, they might be persuaded to update their settings.

7. Conclusions

Contributions in context. Our work suggests that the Privacy Paradox can be actively mitigated. This has great implications for both research and industry. Previous studies have highlighted the importance of increasing awareness (Pötzsch, 2009; Deuker, 2009). However, since public campaigns can lack efficacy (Sasse et al., 2007), it was unclear whether mitigation was feasible. Through imparting knowledge and incentivising action, the Paradox was successfully reduced. Therefore, while the issue appears to defy rationality (Acquisti and Grossklags, 2005), it is far from intractable. Our research has also demonstrated the potential of smartwatch games. Few wearable apps have yet targeted this environment (Casano et al., 2016; Arroyo et al., 2016). However, the interface appears to be sufficiently rich. We adapted techniques from Learning Science, and we offer *(Smart)Watch Out!* as a foundation for future applications.

The technology industry might also be affected by Paradox mitigation. Many companies currently base their business models on data collection (Schneier, 2015). Until now, even when consumers express privacy concerns, they fail to use protection. This allowed vendors to defend their practices as tacitly accepted (Bösch et al., 2015). With the growth of the Internet-of-Things, data collection appears set to increase (Perera et al., 2015). Smart devices are often cheap and obsolescent (Schneier, 2015), reducing incentives for security protection. This environment is considered likely to foster the Paradox (Williams et al., 2016*b*). However, if learning techniques can realign behavior with concern, users might gain greater agency. Conversely, research provides several opportunities for tech companies. Smartwatch interfaces continue to be constrained and unfamiliar (Chun et al., 2018). But games appear successful in teaching platform navigation. Future apps could target non-privacy topics, such as installing apps or configuring settings.

By actively mitigating the Paradox, our work offers a novel contribution. Groom and Calo (2011) proposed a study to realign behavior with concerns. However, it is unclear whether the experiment was ever conducted. Jackson and Wang (2018) did mitigate the issue in their smartphone nudging work. But whereas they realigned behavior on a single date, we considered actions over a two-week period. Furthermore, the Paradox has not been previously explored within smartwatch environments. As these devices grow in sophistication and popularity, privacy research will become increasingly important.

Our results compare favourably with non-game privacy interventions. LaRose and Rifon (2007) used warnings and seals to highlight the risk of interactions. When the issues were communicated, participants were significantly less likely

to disclose. However, since their sample was smaller and student-composed, it has limitations in external validity. Similar to our work, Shillair et al. (2015*b*) developed an intervention based on Protection Motivation Theory. They assessed the influence that salient responsibility would have on behavior. While their approach increased self-efficacy, it did not result in enhanced privacy. Our game had similar success to 'Cyberheroes': an interactive e-book developed by Zhang-Kennedy et al. (2017). The application highlighted the risks of digital footprints and online interactions. Although protection was encouraged for one week, it was only evaluated by 44 participants. Our work was trialled in detail by 504 smartwatch users. Finally, Archer et al. (2014) assessed the influence of instructional videos. Through their 80-person study, they found the intervention led to settings configuration. While our research had similar success, we analysed behavior over a greater period of time.

Despite our promising findings, we accept that games do not provide a panacea. Hays (2005) conducted an extensive literature review of instructional applications. While games had the capacity to impart information, there was no evidence they were the preferred solution. In future work, we wish to compare our app's influence with that of other techniques. Wouters et al. (2013) highlighted that games are often assumed to be motivational. However, they found enthusiasm failed to differ between this approach and traditional instruction. If games are to achieve success, they must be appreciated by their users. Of particular concern is the relative scarcity of empirical evaluations (Tobias et al., 2011). While many apps are developed, few are assessed for their behavioral influence. We call for the conduction of intervention studies, much like our own research.

Summary. As our primary contribution, we explored whether protective behavior could be encouraged through a smartwatch game. Individuals completed privacy challenges on a simulated interface, directly practicing protective actions. The application was refined in a pilot study and directly informed by Learning Science principles.

To assess its influence, we solicited the concerns and behavior of 504 smartwatch owners. Half of these users played the game (treatment group), while the other half did not (control group). Protective behavior became significantly more frequent in the former group, with the Privacy Paradox significantly reduced. The control group did not differ throughout the study, implying that extraneous variables were minimized. These results suggest that privacy protection can be encouraged, and provide support to future privacy apps.

To gain insight into participants' rationale, we also solicited qualitative responses. Behavior deviated from concerns for several reasons, including data

sensitivity, social norms and ideology. We concluded our work by distilling recommendations for future privacy interventions. These approaches were informed by Protection Motivation Theory. Tools should be demonstrated, risk exposure should be highlighted and 'user-based penetration testing' should be considered. Through these techniques, behavior might be realigned with concerns.

Limitations and future work. While we stand by our findings, we accept several limitations. Firstly, our interface was simulated rather than hosted on a smartwatch. This might have reduced its usability and potential for influence. However, this remote approach enabled large-scale analysis with a varied demographic. We are developing serious Wear OS apps, and they will be informed by these findings. Secondly, due to the online nature of our study, behavior was reported. While empirical data is preferable, top-tier studies frequently adopt this approach (Albayram, Khan and Jensen, 2017; Coventry et al., 2014; Staddon et al., 2012). We sought to limit the risk of fabrication through ensuring anonymity, mitigating priming and allocating compensation equally. Participants from Prolific have also been found to be less dishonest than those on competing platforms (Peer et al., 2017). Since privacy was disguised in the surveys, this further mitigated the Hawthorne Effect (Adair, 1984). In future research we wish to conduct field studies, informed by our findings.

Thirdly, although our privacy measures were carefully devised, they did not receive statistical validation. Informed by our study's results, we plan to refine and validate the privacy questionnaires. Since these measures were reviewed and piloted before our experiment, we believe they still proved valuable. Finally, while one-week pauses are common in security research (Albayram, Khan and Jensen, 2017; Zhang-Kennedy et al., 2017; Wiedenbeck et al., 2005; DeWitt and Kuljis, 2006; Kumaraguru et al., 2009), we accept that retention was not extensively assessed. Individuals might retain knowledge and behavior for a month, before returning to their old habits. As further work, longitudinal studies should explore whether games are influential over a greater period of time.

Despite these limitations, we developed the first game to concern smartwatch privacy. Through a large-scale evaluation, it appeared to encourage user protection. We also achieved the rare feat of mitigating the Privacy Paradox. As smart devices proliferate through our societies, it is crucial that privacy is supported.

Acknowledgements

Meredydd Williams receives funding through an EPSRC Doctoral Studentship in Cyber Security (Ref:1513964). This study did not receive any specific grant

from funding agencies in the public, commercial, or not-for-profit sectors.

Declarations of Interest

Declarations of interest: none.

References

Acquisti, A. and Gross, R. (2006), 'Imagined communities: Awareness, information sharing, and privacy on the Facebook', *Privacy Enhancing Technologies in Lecture Notes in Computer Science* **4258**, 36–58.

Acquisti, A. and Grossklags, J. (2005), 'Privacy and rationality in individual decision making', *IEEE Security & Privacy* (1), 26–33.

Adair, J. G. (1984), 'The Hawthorne effect: A reconsideration of the methodological artifact', *Journal of Applied Psychology* **69**(2), 334–345.

Ajzen, I. (1988), *Attitudes, personality, and behavior*, Dorsey Press.

Ajzen, I. (1991), 'The theory of planned behavior', *Organizational Behavior and Human Decision Processes* **50**(2), 179–211.

Aktypi, A., Nurse, J. R. C. and Goldsmith, M. (2017), Unwinding Ariadne's identity thread: Privacy risks with fitness trackers and online social networks, *in* 'Proceedings of the 2017 on Multimedia Privacy and Security at the 24th ACM Conference on Computer and Communication Security (CCS)'.

Albayram, Y., Khan, M. M. H. and Fagan, M. (2017), 'A study on designing video tutorials for promoting security features: A case study in the context of Two-Factor Authentication (2FA)', *International Journal of Human-Computer Interaction* **33**(11), 1–16.

Albayram, Y., Khan, M. M. H. and Jensen, T. (2017), "...better to use a lock screen than to worry about saving a few seconds of time": Effect of fear appeal in the context of smartphone locking behavior, *in* 'Proceedings of the 13th Symposium on Usable Privacy and Security', pp. 49–63.

Anderson, C. L. and Agarwal, R. (2010), 'Practicing safe computing: A multimedia empirical examination of home computer user security behavioral intentions', *MIS Quarterly* **34**(3), 613–643.

Annetta, L. A. (2008), 'Why and how video games should be used in education', *Theory Into Practice* **47**(3), 229–239.

Annetta, L. A. (2010), 'The "I's" have it: A framework for serious educational game design', *Review of General Psychology* **14**(2), 105–112.

Annetta, L. A. and Holmes, S. (2006), 'Creating presence and community in a synchronous virtual learning environment using avatars', *International Journal of Instructional Technology and Distance Learning* **3**(8), 27–43.

Archer, K., Wood, E., Nosko, A., De Pasquale, D., Molema, S. and Christofides, E. (2014), 'Disclosure and privacy settings on social networking sites: Evaluating an instructional intervention designed to promote informed information sharing', *International Journal of Cyber Behavior, Psychology and Learning* **4**(2), 1–19.

Arroyo, I., Liu, Y., Wixon, N. and Schultz, S. (2016), 'Toward embodied game-based intelligent tutoring systems', *Intelligent Tutoring Systems in Lecture Notes in Computer Science* **9684**, 488–490.

Ashbrook, D. and Starner, T. (2003), 'Using GPS to learn significant locations and predict movement across multiple users', *Personal and Ubiquitous Computing* **7**(5), 275–286.

Bada, M., Sasse, A. and Nurse, J. R. C. (2015), Cyber security awareness campaigns: Why do they fail to change behaviour?, *in* 'Proceedings of the International Conference on Cyber Security for Sustainable Society', pp. 118–131.

Bandara, R., Fernando, M. and Akter, S. (2018), Is the Privacy Paradox a matter of psychological distance? An exploratory study of the Privacy Paradox from a Construal Level Theory perspective, *in* 'Proceedings of the 51st Hawaii International Conference on System Sciences', pp. 3678–3687.

Bandura, A. (1977), 'Self-efficacy: Toward a unifying theory of behavioral change', *Psychological Review* **84**(2), 191–215.

Barnard-Wills, D. and Ashenden, D. (2015), 'Playing with privacy: Games for education and communication in the politics of online privacy', *Political Studies* **63**(1), 142–160.

Barnes, S. B. (2006), 'A privacy paradox: Social networking in the United States', *First Monday* **11**(9).

Beard, C. (2010), *The experiential learning toolkit: Blending practice with concepts*, Kogan Page Publishers.

Benton, L., Vasalou, A., Berkling, K., Barendregt, W. and Mavrikis, M. (2018), A critical examination of feedback in early reading games, *in* 'Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems'.

Blank, G., Bolsover, G. and Dubois, E. (2014), A new privacy paradox, Technical report, Global Cyber Security Capacity Centre.

Boer, H. and Mashamba, M. T. (2005), 'Psychosocial correlates of HIV protection motivation among black adolescents in Venda, South Africa', *AIDS Education & Prevention* **17**(6), 590–602.

Bonneau, J. and Preibusch, S. (2010), 'The privacy jungle: On the market for data protection in social networks', *Economics of Information Security and Privacy* pp. 121–167.

Bösch, C., Erb, B., Kargl, F., Kopp, H. and Pfattheicher, S. (2015), 'Tales from the dark side: Privacy dark strategies and privacy dark patterns', *Proceedings on Privacy Enhancing Technologies* **2016**(4), 237–254.

Boyatzis, R. E. (1998), *Transforming qualitative information: Thematic analysis and code development*, Sage.

Braun, V. and Clarke, V. (2006), 'Using thematic analysis in psychology', *Qualitative Research in Psychology* **3**(2), 77–101.

Braunstein, A., Granka, L. and Staddon, J. (2011), Indirect content privacy surveys: Measuring privacy without asking about it, *in* 'Proceedings of the 7th Symposium on Usable Privacy and Security'.

Breuer, J. and Bente, G. (2010), 'Why so serious? on the relation of serious games and learning', *Journal for Computer Game Culture* **4**, 7–24.

Briggs, P., Jeske, D. and Coventry, L. (2017), Behavior change interventions for cybersecurity, *in* 'Behavior Change Research and Theory', Academic Press, pp. 115–136.

Bruyneel, S. and Dewitte, S. (2016), Health nudges: How behavioural engineering can reduce chocolate consumption, *in* 'The Economics of Chocolate', pp. 157–170.

Canova, G., Volkamer, M., Bergmann, C. and Borza, R. (2014), NoPhish: An anti-phishing education app, *in* 'Proceedings of the International Workshop on Security and Trust Management', pp. 188–192.

Carifio, J. and Perla, R. (2008), 'Resolving the 50-year debate around using and misusing Likert scales', *Medical Education* **42**(12), 1150–1152.

Carter, M., Downs, J., Nansen, B. and Harrop, M. (2014), Paradigms of games research in HCI: A review of 10 years of research at CHI, *in* 'Proceedings of the 1st ACM SIGCHI Annual Symposium on Computer-Human Interaction in Play', pp. 27–36.

Casano, J., Tee, H., Agapito, J., Arroyo, I. and Rodrigo, M. M. T. (2016), Migration and evaluation of a framework for developing embodied cognition learning games, *in* 'Proceedings of the 3rd Asia-Europe Symposium on Simulation & Serious Gaming', pp. 199–203.

Charsky, D. (2010), 'From edutainment to serious games: A change in the use of game characteristics', *Games and Culture* **5**(2), 177–198.

Chellappa, R. K. and Sin, R. G. (2005), 'Personalization versus privacy: An empirical examination of the online consumer's dilemma', *Information Technology and Management* **6**(2-3), 181–202.

Chenoweth, T., Minch, R. and Gattiker, T. (2009), Application of Protection Motivation Theory to adoption of protective technologies, *in* 'Proceedings of the 42nd Hawaii International Conference on System Sciences'.

Chun, J., Dey, A., Lee, K. and Kim, S. (2018), 'A qualitative study of smartwatch usage and its usability', *Human Factors and Ergonomics in Manufacturing and Service Industries* **28**(4), 186–199.

Clarke, R. (1999), Introduction to dataveillance and information privacy, and definitions of terms, Technical report. Last accessed on 22nd February 2019.
URL: *http://www.qatar.cmu.edu/iliano/courses/10F-CMU-CS349/slides/privacy.pdf*

Cleophas, T. J. M. (1990), 'A simple method for the estimation of interaction bias in crossover studies', *The Journal of Clinical Pharmacology* **30**(11), 1036–1040.

Collingridge, D. (2014), 'Validating a questionnaire'. Last accessed on 22nd February 2019.
URL: *www.methodspace.com/validating-a-questionnaire/*

Collins English Dictionary (2017), 'Smartwatch definition and meaning'. Last accessed on 22nd February 2019.
URL: *https://www.collinsdictionary.com/dictionary/english/smartwatch*

Connolly, T. M., Boyle, E. A., MacArthur, E. and Hainey, T. (2012), 'A systematic literature review of empirical evidence on computer games and serious games', *Computers & Education* **59**(2), 661–686.

Consumer Reports (2014), '5 steps to protect your smart phone from theft or loss'. Last accessed on 22nd February 2019.
URL: *https://www.consumerreports.org/cro/2014/04/5-steps-to-protect-your-smart-phone-against-theft-or-loss/index.htm*

Coventry, L., Jeske, D. and Briggs, P. (2014), Perceptions and actions: Combining privacy and risk perceptions to better understand user behaviour, *in* 'Proceedings of the 10th Symposium on Usable Privacy and Security'.

Creese, S., Goldsmith, M., Nurse, J. R. C. and Phillips, E. (2012), A data-reachability model for elucidating privacy and security risks related to the use of online social networks, *in* 'Proceedings of the 11th International Conference on Trust, Security and Privacy in Computing and Communications', pp. 1124–1131.

Culnan, M. J. and Armstrong, P. K. (1999), 'Information privacy concerns, procedural fairness, and impersonal trust: An empirical investigation', *Organization Science* **10**(1), 104–115.

Cybulski, A. D. (2014), 'Enclosures at play: Surveillance in the code and culture of videogames', *Surveillance & Society* **12**(3), 427–432.

Dai, J., Duan, Y. and Yang, F. (2016), An optimized privacy inference attack based on smartwatch motion sensors, *in* 'Proceedings of the 2016 International Conference on iThings-GreenCom-CPSCom-SmartData', pp. 474–479.

Deriso, D., Susskind, J., Krieger, L. and Bartlett, M. (2012), Emotion mirror: A novel intervention for autism based on real-time expression recognition, *in* 'European Conference on Computer Vision', pp. 671–674.

Desarnauts, B. (2016), 'Wristly Insider's report #45'. Last accessed on 22nd February 2019.
URL: *https://medium.com/wristly-thoughts/one-year-in-and-only-now-are-we-getting-to-know-apple-watch-owners-db60d565d041*

Deuker, A. (2009), 'Addressing the privacy paradox by expanded privacy awareness - The example of context-aware services', *IFIP Advances in Information and Communication Technology* **320**, 275–283.

DeWitt, A. J. and Kuljis, J. (2006), Aligning usability and security: A usability study of Polaris, *in* 'Proceedings of the 2nd Symposium on Usable Privacy and Security', pp. 1–7.

Do, Q., Martini, B. and Choo, K.-K. R. (2017), 'Is the data on your wearable device secure? An Android Wear smartwatch case study', *Software: Practice and Experience* **47**(3), 391–403.

Dolan, P., Hallsworth, M., Halpern, D., King, D. and Vlaev, I. (2010), MINDSPACE: Influencing behaviour for public policy, Technical report, Institute of Government, London School of Economics and Political Science.

Eadicicco, L. (2014), 'A new wave of gadgets can collect your personal information like never before'. Last accessed on 22nd February 2019.
URL: *http://www.businessinsider.com/privacy-fitness-trackers-smartwatches-2014-10*

Egelman, S., Tsai, J., Cranor, L. and Acquisti, A. (2009), Timing is everything?: The effects of timing and placement of online privacy indicators, *in* 'Proceedings of the 2009 SIGCHI Conference on Human Factors in Computing Systems', pp. 319–328.

Ernst, C. P. H. and Ernst, A. W. (2016), The influence of privacy risk on smartwatch usage, *in* 'Proceedings of the 2016 Americas' Conference on Information Systems'.

Evans, J. R. and Mathur, A. (2005), 'The value of online surveys', *Internet Research* **15**(2), 195–219.

Fazio, R. H. and Roskos-Ewoldsen, D. R. (2005), Acting as we feel: When and how attitudes guide behavior, *in* 'Persuasion: Psychological Insights and Perspectives', Allyn & Bacon, pp. 41–62.

Felt, A. P., Egelman, S. and Wagner, D. (2012), "I've got 99 problems, but vibration ain't one": A survey of smartphone users' concerns, *in* 'Proceedings of the Second ACM Workshop on Security and Privacy in Smartphones and Mobile Devices', pp. 33–44.

Felt, A. P., Ha, E., Egelman, S., Haney, A., Chin, E. and Wagner, D. (2012), Android permissions: User attention, comprehension, and behavior, *in* 'Proceedings of the 8th Symposium on Usable Privacy and Security'.

Fishbein, M. (1979), 'A theory of reasoned action: Some applications and implications', *Nebraska Symposium on Motivation* **27**, 65–116.

Floyd, D. L., Prentice-Dunn, S. and Rogers, R. W. (2000), 'A meta-analysis of research on protection motivation theory', *Journal of Applied Social Psychology* **30**(2), 407–429.

Flushman, T., Gondree, M. and Peterson, Z. N. (2015), This is not a game: Early observations on using alternate reality games for teaching security concepts to first-year undergraduates, *in* 'Proceedings of the 8th Workshop on Cyber Security Experimentation and Test'.

Ford, M. and Palmer, W. (2018), 'Alexa, are you listening to me? An analysis of Alexa voice service network traffic', *Personal and Ubiquitous Computing* . [Currently online; volume/number awaiting physical publication].

Furman, S., Theofanos, M. F., Choong, Y. and Stanton, B. (2011), 'Basing cybersecurity training on user perceptions', *IEEE Security & Privacy* **10**(2), 40–49.

Gambino, A., Kim, J., Sundar, S., Ge, J. and Rosson, M. B. (2016), User disbelief in privacy paradox: Heuristics that determine disclosure, *in* 'Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems', pp. 2837–2843.

Garg, V. and Camp, J. (2012), End user perception of online risk under uncertainty, *in* 'Proceedings of the 45th Hawaii International Conference on System Science', pp. 3278–3287.

Gee, J. P. (2003), 'What video games have to teach us about learning and literacy', *Computers in Entertainment* **1**(1), 20–23.

Ghiglieri, M., Volkamer, M. and Renaud, K. (2017), 'Exploring consumers' attitudes of smart TV related privacy risks', *Human Aspects of Information Security, Privacy and Trust in Lecture Notes in Computer Science* **10292**, 656–674.

Groom, V. and Calo, R. (2011), Reversing the privacy paradox: An experimental study, *in* 'Proceedings of the Research Conference on Communications, Information and Internet Policy'.

Hale, M. L., Gamble, R. F. and Gamble, P. (2015), CyberPhishing: A game-based platform for phishing awareness testing, *in* 'The 48th Hawaii International Conference on System Sciences', pp. 5260–5269.

Hallam, C. and Zanella, G. (2017), 'Online self-disclosure: The privacy paradox explained as a temporally discounted balance between concerns and rewards', *Computers in Human Behavior* **68**, 217–227.

Hays, R. T. (2005), The effectiveness of instructional games: A literature review and discussion, Technical report, Naval Air Warfare Center, Training Systems Division.

Hope, A. (2014), World of Spycraft: Video games, gamification and surveillance creep, *in* 'Surveillance Futures: Social and Ethical Implications of New Technologies for Children and Young People', pp. 174–185.

Huizenga, J., Admiraal, W., Akkerman, S. and Dam, G. t. (2009), 'Mobile game-based learning in secondary education: Engagement, motivation and learning in a mobile city game', *Journal of Computer Assisted Learning* **25**(4), 332–344.

Hulsey, N. and Reeves, J. (2014), 'The gift that keeps on giving: Google, Ingress, and the gift of surveillance', *Surveillance & Society* **12**(3), 389–400.

Irvine, C. E., Thompson, M. F. and Allen, K. (2005), 'CyberCIEGE: Gaming for information assurance', *IEEE Security & Privacy* **3**(3), 61–64.

Jackson, C. B. and Wang, Y. (2018), 'Addressing the Privacy Paradox through personalized privacy notifications', *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* **2**(2).

Jackson, J., Allum, N. and Gaskell, G. (2005), Perceptions of risk in cyberspace, *in* 'Trust and Crime in Information Societies', Edward Elgar, pp. 245–281.

Jeong, H., Kim, H., Kim, R., Lee, U. and Jeong, Y. (2017), 'Smartwatch wearing behavior analysis: A longitudinal study', *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* **1**(3), 60.

Johnson, C. I., Bailey, S. K. T. and Van Buskirk, W. L. (2017), Designing effective feedback messages in serious games and simulations: A research review, *in* 'Instructional Techniques to Facilitate Learning and Motivation of Serious Games', pp. 119–140.

Kato, P. M., Cole, S. W., Bradlyn, A. S. and Pollock, B. H. (2008), 'A video game improves behavioral outcomes in adolescents and young adults with cancer: A randomized trial', *Pediatrics* **122**(2), e305–e317.

Ke, F. (2011), A qualitative meta-analysis of computer games as learning tools, *in* 'Gaming and Simulations: Concepts, Methodologies, Tools and Applications', pp. 1619–1665.

Keith, M. J., Thompson, S. C., Hale, J., Lowry, P. B. and Chapman, G. (2013), 'Information disclosure on mobile devices: Re-examining privacy calculus with actual user behavior', *International Journal of Human-Computer Studies* **71**(12), 1163–1173.

Kelley, P. G., Cranor, L. F. and Sadeh, N. (2013), Privacy as part of the app decision-making process, *in* 'Proceedings of the 2013 SIGCHI Conference on Human Factors in Computing Systems', pp. 3393–3402.

Kirlappos, I. and Sasse, M. A. (2011), 'Security education against phishing: A modest proposal for a major rethink', *IEEE Security & Privacy* **10**(2), 24–32.

Kokolakis, S. (2017), 'Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon', *Computers & Security* **64**.

Kumar, J. (2013), 'Gamification at work: Designing engaging business software', *Design, User Experience and Usability in Lecture Notes in Computer Science* **8013**, 528–537.

Kumar, P., Vitak, J., Chetty, M., Clegg, T. L., Yang, J., McNally, B. and Bonsignore, E. (2018), Co-designing online privacy-related games and stories with children, *in* 'Proceedings of the 17th ACM Conference on Interaction Design and Children', pp. 67–79.

Kumaraguru, P., Cranshaw, J., Acquisti, A., Cranor, L. F., Hong, J., Blair, M. A. and Pham, T. (2009), School of phish: A real-world evaluation of anti-phishing training, *in* 'Proceedings of the 5th Symposium on Usable Privacy and Security'.

LaRose, R. and Rifon, N. (2007), 'Promoting i-safety: Effects of privacy warnings and privacy seals on risk assessment and online privacy behavior', *Journal of Consumer Affairs* **41**(1), 127–149.

Lee, H., Park, H. and Kim, J. (2013), 'Why do people share their context information on social network services? A qualitative study and an experimental study on users' behavior of balancing perceived benefit and risk', *International Journal of Human-Computer Studies* **71**(9), 862–877.

Lee, L., Lee, J. H., Egelman, S. and Wagner, D. (2016), Information disclosure concerns in the age of wearable computing, *in* 'Proceedings of the 2016 Workshop on Usable Security'.

Lin, J. J., Mamykina, L., Lindtner, S., Delajoux, G. and Strub, H. B. (2006), Fish'n'Steps: Encouraging physical activity with an interactive computer game, *in* 'International Conference on Ubiquitous Computing', pp. 261–278.

Linehan, C., Kirman, B., Lawson, S. and Chan, G. (2011), Practical, appropriate, empirically-validated guidelines for designing educational games, *in* 'Proceedings of the SIGCHI Conference on Human Factors in Computing Systems', pp. 1979–1988.

Lovio, R., Halttunen, A., Lyytinen, H., Näätänen, R. and Kujala, T. (2012), 'Reading skill and neural processing accuracy improvement after a 3-hour intervention in preschoolers with difficulties in reading-related skills', *Brain Research* **1448**, 42–55.

MacDonell, K., Chen, X., Yan, Y., Li, F., Gong, J., Sun, H., Li, X. and Stanton, B. (2013), 'A Protection Motivation Theory-based scale for tobacco research among Chinese youth', *Journal of Addiction Research & Therapy* **4**, 154.

Maldonado, H., Lee, J. R. L., Brave, S., Nass, C., Nakajima, H., Yamada, R., Iwamura, K. and Morishima, Y. (2005), We learn better together: Enhancing eLearning with emotional characters, *in* 'Proceedings of the 2005 Conference on Computer Support for Collaborative Learning', pp. 408–417.

Malhotra, N. K., Kim, S. S. and Agarwal, J. (2004), 'Internet users' information privacy concerns (IUIPC): The construct, the scale, and a causal model', *Information Systems Research* **15**(4), 336–355.

Mann, H. B. and Whitney, D. R. (1947), 'On a test of whether one of two random variables is stochastically larger than the other', *The Annals of Mathematical Statistics* **18**(1), 50–60.

Matthews, S. (2016), 'Smartwatch dangers - Are you the target?', *War on Identity Theft* . Last accessed on 22nd February 2019.
URL: *http://www.waronidtheft.org/smartwatch-dangers-are-you-the-target/*

McDonald, A. M. and Cranor, L. F. (2008), 'The cost of reading privacy policies', *I/S: A Journal of Law and Policy for the Information Society* **4**(22), 543–568.

McNemar, Q. (1947), 'Note on the sampling error of the difference between correlated proportions or percentages', *Psychometrika* **12**(2), 153–157.

Morar Consulting (2016), The dangers of our digital lives, Technical report. Last accessed on 22nd February 2019.
URL: *https://static2.hidemyass.com/20170623/web/o/docs/hma-survey-summary-2-5-16.pdf*

Myagmar, S., Lee, A. J. and Yurcik, W. (2005), Threat modeling as a basis for security requirements, *in* 'Proceedings on the Symposium on Requirements Engineering for Information Security'.

Nichols, A. L. and Maner, J. K. (2008), 'The good-subject effect: Investigating participant demand characteristics', *The Journal of General Psychology* **135**(2), 151–166.

Nickerson, R. S. (2000), 'Null hypothesis significance testing: A review of an old and continuing controversy', *Psychological Methods* **5**(2), 241–301.

Nissenbaum, H. (2009), *Privacy in context: Technology, policy, and the integrity of social life*, Stanford University Press.

Norman, G. (2010), 'Likert scales, levels of measurement and the "laws" of statistics', *Advances in Health Sciences Education* **15**(5), 625–632.

Norman, P. and Conner, M. (1996), The role of social cognition models in predicting health behaviours: Future directions, *in* 'Predicting Health Behaviour: Research and Practice with Social Cognition Models', Open University Press, pp. 197–225.

NPD Connected Intelligence (2014), Consumers and wearables report, Technical report. Last accessed on 22nd February 2019.
URL: *https://www.npd.com/wps/portal/npd/us/news/press-releases/2015/the-demographic-divide-fitness-trackers-and-smartwatches-attracting-very-different-segments-of-the-market-according-to-the-npd-group/*

Nurse, J. R. C. (2018), Cybercrime and you: How criminals attack and the human factors that they seek to exploit, *in* 'The Oxford Handbook of Cyberpsychology', Oxford University Press.

Oetzel, M. C. and Gonja, T. (2011), The online privacy paradox: A social representations perspective, *in* 'CHI'11 Extended Abstracts on Human Factors in Computing Systems', pp. 2107–2112.

Paine, C., Reips, U. D., Stieger, S., Joinson, A. and Buchanan, T. (2007), 'Internet users' perceptions of 'privacy concerns' and 'privacy actions'', *International Journal of Human-Computer Studies* **65**(6), 526–536.

Palen, L. and Dourish, P. (2003), Unpacking privacy for a networked world, *in* 'Proceedings of the 2003 SIGCHI Conference on Human Factors in Computing Systems', pp. 129–136.

Parkinson, S., Eatough, V., Holmes, J., Stapley, E. and Midgley, N. (2016), 'Framework analysis: A worked example of a study exploring young people's experiences of depression', *Qualitative Research in Psychology* **13**(2), 109–129.

Pearson, K. (1900), 'On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling', *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **50**(302), 157–175.

Peer, E., Brandimarte, L., Samat, S. and Acquisti, A. (2017), 'Beyond the Turk: Alternative platforms for crowdsourcing behavioral research', *Journal of Experimental Social Psychology* **70**, 153–163.

Peer, E., Vosgerau, J. and Acquisti, A. (2014), 'Reputation as a sufficient condition for data quality on Amazon Mechanical Turk', *Behavior Research Methods* **46**(4), 1023–1031.

Perera, C., Ranjan, R., Wang, L., Khan, S. U. and Zomaya, A. Y. (2015), 'Big data privacy in the Internet of Things era', *IT Professional* **17**(3), 32–39.

Phelan, C., Lampe, C. and Resnick, P. (2016), It's creepy, but it doesn't bother me, *in* 'Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems', pp. 5240–5251.

Pike, S., Kelledy, M. and Gelnaw, A. (2017), Measuring U.S. privacy sentiment: An IDC special report, Technical report. Last accessed on 22nd February 2019. **URL:** *http://www.idc.com/getdoc.jsp?containerId=prUS42253017*

Pizza, S., Brown, B., McMillan, D. and Lampinen, A. (2016), Smartwatch in vivo, *in* 'Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16', pp. 5456–5469.

Pötzsch, S. (2009), Privacy awareness: A means to solve the privacy paradox?, *in* 'IFIP Advances in Information and Communication Technology', Vol. 298, pp. 226–236.

Quinn, C. N. (2005), *Engaging learning: Designing e-learning simulation games*, John Wiley & Sons.

Rahman, N. H. A. (2013), 'Privacy disclosure risk: Smartphone user guide', *International Journal of Mobile Network Design and Innovation* **5**(1), 2–8.

Rainie, L., Kiesler, S., Kang, R., Madden, M., Duggan, M., Brown, S. and Dabbish, L. (2013), 'Anonymity, privacy, and security online', *Pew Internet & American Life Project* .

Rajivan, P. and Camp, J. (2016), Influence of privacy attitude and privacy cue framing on android app choices, *in* 'Proceedings of the 12th Symposium on Usable Privacy and Security'.

Raman, R., Lal, A. and Achuthan, K. (2014), Serious games based approach to cyber security concept learning: Indian context, *in* 'Proceedings of the 2014 International Conference on Green Computing Communication and Electrical Engineering'.

Raynes-Goldie, K. and Allen, M. (2014), 'Gaming privacy: A Canadian case study of a children's co-created privacy literacy game', *Surveillance & Society* **12**(3), 414–426.

Renaud, K., Volkamer, M. and Renkema-Padmos, A. (2014), 'Why doesn't Jane protect her privacy?', *Privacy Enhancing Technologies in Lecture Notes in Computer Science* **8555**, 244–262.

Rheingans, F., Cikit, B. and Ernst, C. P. H. (2016), The potential influence of privacy risk on activity tracker usage: A study, *in* 'The Drivers of Wearable Device Usage', Springer, pp. 25–35.

Ricci, J., Baggili, I. and Breitinger, F. (2018), Watch what you wear: Smartwatches and sluggish security, *in* 'Wearable Technologies: Concepts, Methodologies, Tools, and Applications', IGI Global, pp. 1458–1478.

Richter, G., Raban, D. R. and Rafaeli, S. (2015), Studying gamification: The effect of rewards and incentives on motivation, *in* 'Gamification in education and business', pp. 21–46.

Ritchie, J., Spencer, L. and O'Connor, W. (2003), Carrying out qualitative analysis, *in* 'Qualitative Research Practice: A Guide for Social Science Students and Researchers', Sage Publications, pp. 219–262.

Ritterfeld, U., Cody, M. and Vorderer, P. (2009), *Serious games: Mechanisms and effects*, Routledge.

Rogers, R. W. (1983), Cognitive and psychological processes in fear appeals and attitude change: A revised theory of protection motivation, *in* 'Social Psychophysiology: A Sourcebook', Guilford Publications, pp. 153–176.

Ryan, R. M. and Deci, E. L. (2000), 'Intrinsic and extrinsic motivations: Classic definitions and new directions', *Contemporary Educational Psychology* **25**(1), 54–67.

Sasse, M. A., Ashenden, D., Lawrence, D., Coles-Kemp, L., Fléchais, I. and Kearney, P. (2007), Human vulnerabilities in security systems, Technical report. Last accessed on 22nd February 2019.
URL: *http://www.cs.ucl.ac.uk/fileadmin/sec/publications/HFWG_White_Paper_final.pdf*

Schneier, B. (2015), *Data and Goliath*, W. W. Norton & Company.

Sheng, S., Magnien, B. and Kumaraguru, P. (2007), Anti-phishing Phil: The design and evaluation of a game that teaches people not to fall for phish, *in* 'Proceedings of the 3rd Symposium on Usable Privacy and Security', pp. 88–99.

Shillair, R., Cotten, S. R., Tsai, H. S., Alhabash, S., LaRose, R. and Rifon, N. J. (2015*a*), 'Online safety begins with you and me: Convincing Internet users to protect themselves', *Computers in Human Behavior* **48**, 199–207.

Shillair, R., Cotten, S. R., Tsai, H. Y. S., Alhabash, S., LaRose, R. and Rifon, N. J. (2015*b*), 'Online safety begins with you and me: Convincing Internet users to protect themselves', *Computers in Human Behavior* **48**, 199–207.

Shuttleworth, M. (2010), 'Experimental research'. Last accessed on 22nd February 2019.
URL: *https://explorable.com/scientific-control-group*

Skinner, B. F. (1953), *Science and human behavior*, number 92904, Simon and Schuster.

Smith, H. J., Milberg, S. J. and Burke, S. J. (1996), 'Information privacy: Measuring individuals' concerns about organizational practices', *MIS Quarterly* **20**(2), 167–196.

Smith, M. T. (2007), 'Reconciling ICT and wearable design: Ten lessons from working with Swatch', *The Role of Design in Wearable Computing* pp. 16–20.

Spearman, C. (1904), 'The proof and measurement of association between two things', *The American Journal of Psychology* **15**(1), 72–101.

Staddon, J., Huffaker, D., Brown, L. and Sedley, A. (2012), Are privacy concerns a turn-off?, *in* 'Proceedings of the 8th Symposium on Usable Privacy and Security'.

Steiner, C. M., Kickmeier-Rust, M. D., Mattheiss, E. and Albert, D. (2009), Undercover: Non-invasive, adaptive interventions in educational games, *in* 'Proceedings of 80Days' 1st International Open Workshop on Intelligent Personalisation and Adaptation in Digital Educational Games', pp. 55–65.

Suknot, A., Chavez, T., Rackley, N. and Kelley, P. G. (2014), Immaculacy: A game of privacy, *in* 'Proceedings of the 1st ACM SIGCHI Annual Symposium on Computer-Human Interaction in Play', pp. 383–386.

Tobias, S., Fletcher, J. D., Dai, D. Y. and Wind, A. P. (2011), Review of research on computer games, *in* 'Computer Games and Instruction', pp. 127–222.

Trepte, S., Dienlin, T. and Reinecke, L. (2014), Risky behaviors: How online experiences influence privacy behaviors, *in* 'From the Gutenberg Galaxy to the Google Galaxy. Surveying Old and New Frontiers after 50 Years of DGPuK', pp. 225–244.

Troiano, A. (2016), 'Wearables and personal health data: Putting a premium on your privacy', *Brooklyn Law Review* **82**(4), 1715–1754.

Udoh, E. S. and Alkharashi, A. (2016), Privacy risk awareness and the behavior of smartwatch users: A case study of Indiana University students, *in* 'Proceedings of the 2016 Future Technologies Conference', pp. 926–931.

Voyer, B. (2015), ''Nudging' behaviours in healthcare: Insights from behavioural economics', *British Journal of Healthcare Management* **21**(3), 130–135.

Wang, N., Zhang, B., Liu, B. and Jin, H. (2015), Investigating effects of control and ads awareness on Android users' privacy behaviors and perceptions, *in* 'Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services', pp. 373–382.

Wang, Y., Leon, P. G., Acquisti, A., Cranor, L. F., Forget, A. and Sadeh, N. (2014), A field trial of privacy nudges for Facebook, *in* 'Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems', pp. 2367–2376.

Watson, J. P. and Marks, I. M. (1971), 'Relevant and irrelevant fear in flooding - A crossover study of phobic patients', *Behavior Therapy* **2**(3), 275–293.

Whitson, J. R. and Simon, B. (2014), 'Game studies meets surveillance studies at the edge of digital culture: An introduction to a special issue on surveillance, games and play', *Surveillance & Society* **12**(3), 309–319.

Wiedenbeck, S., Waters, J., Birget, J. C., Brodskiy, A. and Memon, N. (2005), Authentication using graphical passwords: Effects of tolerance and image choice, *in* 'Proceedings of the 1st Symposium on Usable Privacy and Security', pp. 1–12.

Wieneke, A., Lehrer, C. and Zeder, R. (2016), Privacy-related decision-making in the context of wearable use, *in* 'Proceedings of the 20th Pacific Asia Conference on Information Systems'.

Wilcoxon, F. (1945), 'Individual comparisons by ranking methods', *Biometrics Bulletin* **1**(6), 80–83.

Williams, M., Nurse, J. R. C. and Creese, S. (2016*a*), 'Privacy salience: Taxonomies and research opportunities', *IFIP Advances in Information and Communication Technology* **498**, 263–278.

Williams, M., Nurse, J. R. C. and Creese, S. (2016*b*), The perfect storm: The Privacy Paradox and the Internet-of-Things, *in* 'Proceedings of the 2016 11th International Conference on Availability, Reliability and Security (ARES)', pp. 644–652.

Williams, M., Nurse, J. R. C. and Creese, S. (2017), "Privacy is the boring bit": User perceptions and behaviour in the Internet-of-Things, *in* 'Proceedings of the 15th International Conference on Privacy, Security and Trust (PST)'.

Wisniewski, P. J., Knijnenburg, B. P. and Lipford, H. R. (2016), 'Making privacy personal: Profiling social network users to inform privacy education and nudging', *International Journal of Human-Computer Studies* **98**, 95–108.

Wouters, P., Van Nimwegen, C., Van Oostendorp, H. and Van Der Spek, E. D. (2013), 'A meta-analysis of the cognitive and motivational effects of serious games', *Journal of Educational Psychology* **105**(2), 249–265.

Youn, S. (2009), 'Determinants of online privacy concern and its influence on privacy protection behaviors among young adolescents', *Journal of Consumer Affairs* **43**(3), 389–418.

Zhang-Kennedy, L., Abdelaziz, Y. and Chiasson, S. (2017), 'Cyberheroes: The design and evaluation of an interactive ebook to educate children about online privacy', *International Journal of Child-Computer Interaction* **13**, 10–18.

Biographies

Meredydd Williams is a Cyber Security PhD student at the University of Oxford. He received his MPhil Advanced Computer Science from the University of Cambridge in 2014. Prior to this, he completed a BEng Software Engineering at Aberystwyth University. His research interests include online privacy, behavior change, and the security aspects of the Internet-of-Things.

Jason R. C. Nurse is an Assistant Professor in Cyber Security in the School of Computing at the University of Kent. Prior to this he was a Senior Research Fellow in the Department of Computer Science at the University of Oxford and a fellow at Wolfson College, Oxford. His research interests include the Internet-of-Things, risks to identity security and privacy in cyberspace, and human factors. Nurse is a Visiting Academic at the University of Oxford, and a Visiting Fellow in Defence & Security at Cranfield University. He was selected as a Rising Star for his research into cybersecurity and privacy, as a part of the Engineering and Physical Sciences Research Council's Recognising Inspirational Scientists and Engineers (RISE) awards campaign.

Sadie Creese is Professor of Cyber Security in the Department of Computer Science at the University of Oxford. She was founding director of the Global Cyber Security Capacity Centre at the Oxford Martin School and a member of the Coordinating Committee for the Cyber Security Oxford network. She is engaged in a broad portfolio of cybersecurity research spanning situational awareness, visual analytics, risk propagation and communication, threat modelling and detection, network defence, dependability and resilience and privacy.

Appendix A. Posttest Rationale Questions

Table A.2: Posttest Rationale Questions

#	Question
1	In which way do you believe you learn most effectively?
[Visually, Aurally, Verbally, Physically, Unsure]	
2	Do you think you understand how to protect your smartwatch data? Why?
[Yes, No, Unsure]	
3	Do you feel confident that you can protect your smartwatch data? Why?
[Yes, No, Unsure]	
4	Do you think taking action to protect your smartwatch data is worth the effort? Why?
[Yes, No, Unsure]	
5	Do you think your smartwatch data faces a realistic threat? Why?
[Yes, No, Unsure]	