The role of effort in security and privacy behaviours online

Katarzyna Kinga Krol

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

of

University College London.

Department of Security and Crime Science

University College London

2016

I, Katarzyna Kinga Krol, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

As more and more aspects of users' lives go online, they can interact with each other, access services and purchase goods with unprecedented convenience and speed. However, this also means that users' devices and data become more vulnerable to attacks. As security is often added to tools and services as an after-thought, it tends to be poorly integrated into the processes and part of the effort of securing is often offloaded onto the user. Users are goal-driven and they go online to get things done, protecting their security and privacy might therefore not be a priority.

The six studies described in this dissertation examine the role of effort in users' security and privacy behaviours online. First, two security studies use authentication diaries to examine the user effort required for authentication to organisational and online banking systems respectively. Second, two further studies are laboratory evaluations of proposed mechanisms for authentication and verification. Third, two privacy studies examine the role of effort in users' information disclosure in webforms and evaluate a possible solution that could help users manage how much they disclose.

All studies illustrate the different coping strategies users develop to manage their effort. They show that demanding too much effort can affect productivity, cause frustration and undermine the security these mechanisms were meant to offer. The work stresses the importance of conducting methodologically robust user evaluations of both proposed and deployed mechanisms in order to improve user satisfaction and their security and privacy.

Acknowledgements

I would like to thank my amazing supervisor Angela Sasse for providing me with great support, giving me a lot of freedom to explore and always treating me as an equal.

I am indebted to my examiners, Lynne Coventry and Hervé Borrion, for their helpful comments and suggestions which greatly contributed to the improvement of this dissertation.

Further thanks go to EPSRC for generously funding my doctoral research within the SECReT Doctoral Training Centre (grant number: EP/G037264/1). From the Department of Security and Crime Science, I would like to thank Hervé Borrion, Richard Wortley, Nick Tilley, Vaseem Khan and Gloria Laycock.

During my PhD, I had the chance to work with many great people. I would like to thank my co-authors Alastair Beresford, Dana Chisnell, Emiliano De Cristofaro, Constantinos Papanicolaou, Simon Parkin, Eleni Philippou, Sören Preibusch, Michelle Steves, Mary Theofanos and Alexei Vernitski.

I would like to thank all members of our team, past and present: Ruba Abu-Salma, Adam Beautement, Ingolf Becker, Odette Beris, Sacha Brostoff, Tristan Caulfield, Steve Dodier-Lazaro, Brian Glass, Charlene Jennett, Inka Karppinen, Iacovos Kirlappos, Miguel Malheiros, Granville Moore, Anthony Morton, Chris Porter and Martin Ruskov. Thank you all for sharing your excitement and optimism, and creating a great atmosphere to work in.

Finally, I would like to thank my family and my partner for their personal support.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Real-world problems

As more and more aspects of users' lives go online, individuals can interact with each other, access services and purchase goods with unprecedented convenience and speed. However, the increasing connectedness and availability also mean that users' devices and data become more vulnerable to attacks and exploitation. As security is often added to tools and services as an after-thought, it tends to be poorly integrated into the processes [65]. As a result, the effort of securing is often offloaded onto the user. Users are asked to prove they are human as opposed to a robot, create an account on the websites they visit, preferably with a strong unique password, check URLs, verify SSL certificates, to name a few. These demands ignore the significant user effort that is required. At the same time, users are goal-driven and they go online to accomplish their tasks and protecting their security and privacy are often not their priority. Users' effort does not only comprise the time consumed and the cognitive and physical demand but also the emotional costs of frustration and embarrassment. Online security and privacy can be considered enabling tasks because they make it possible to move services known from the physical world (e.g., news, banking, shopping) to the online world providing speed and convenience. However, the way in which these systems are often implemented means there are hurdles users have to overcome in order to access them. These hurdles include logging in, being verified as human, completing forms etc. Although humans are goal-oriented and dislike being interrupted, the industry still relies on users to perform security and privacy tasks that are disruptive to their workflow and keep them away from their primary task. At the same time, attackers are often excellent psychologists who know human attitudes towards effort and exploit the heuristics that guide human behaviour to their own advantage [169, 240].

Today, one of the main selling points of online services and products is how fast they are and how seamlessly they can fit into users' daily activities. In crass discrepancy to that are the security compliance demands the industry puts on users. As Herley stresses, *"[security] advice almost always ignores the cost of user effort"* [112, p. 141]. At the same time, the environment is hostile. Online vendors are incentivised to elicit as much information from users as they can [176]. Attackers are lurking trying to infect users' computers with viruses and draft them into their botnets. Attackers and security experts are not the only ones in competition for user attention, the online environment is full other different stimuli (e.g., adverts, notifications) that try to attract users' attention as well [113].

1.2 Research gap

Research so far has failed to take a holistic view of users and the effort that they need to expend for security and privacy tasks. The amount of effort required to perform a task is a simple decision making rule or heuristic. Yet, security and privacy research has not given it due consideration. Studies have too often just focused on assessing if using a new authentication mechanism or following a procedure is doable and they do so in an isolated laboratory setting, away from other tasks and demands the user faces in the real world. In other words, evaluations are often only conducted pre-deployment under perfect conditions assessing if the user "can do it" and not if they want to do it or how adding this security-related effort would affect their productivity and actual security in a long term.

1.3 Research question

The study of effort in security and privacy behaviours online is important as it can improve the design of future processes and systems, making a difference for both the users and industry. Through six studies using a mix of quantitative and qualitative research methods, this doctoral dissertation aims to answer the research question: **What is the role of effort in users' security and privacy behaviours online?** To scope this question, the following sub-questions were devised:

1. For security, how does effort influence users' behaviours with regard to authentication and verification?

2. For privacy, how does effort influence users' behaviours with regard to information disclosure?
3. Are there differences between the role of effort in security and privacy behaviours?
4. What are users' coping strategies for managing security and privacy effort?
5. What are the factors that moderate how much effort users expend?
6. What is the relationship between actual and perceived effort?

1.3.1 Research scope

The focus of the research is on understanding the role of effort in users' security and privacy behaviours online. The six studies described in this dissertation investigate user behaviours for authentication, verification and information disclosure and explore their perceptions of effort and the actual effort (measured as time) involved.

The first two studies aim to learn about users' adoption of already deployed authentication solutions and explore the different strategies they had developed to cope with authentication effort. Further two studies assess mechanisms pre-adoption and aim to learn about the perceived effort and actual time required when used on two different devices, as well as making participants reflect on possible contexts of use. These two studies can only be treated as preliminary assessments, part of a series of evaluations as the technologies go through their development cycles. Finally, the last two studies explore how users disclose information through webforms. After learning about actual disclosure, its relationship with effort and motivations in the first study, an intervention is designed and evaluated in a follow-up study.

The focus of all six studies is not on assessing security but on learning about the user effort involved in behaviours related to security and privacy. However, all participant perceptions relating to security and privacy are recorded and analysed since they have an impact on user decisions and behaviours. When participants had questions regarding security, the researcher answered them using information provided by the creators of the technologies. This dissertation makes no attempt to evaluate the security of the proposed mechanisms or to assess what level of security would be appropriate for the participants.

1.3.2 Effort, security and privacy defined

Throughout the dissertation, 'effort' and 'workload' are used interchangeably. Both terms refer to the amount of work that has to be completed to achieve a certain goal. 'Effort' is more of an abstract concept while 'workload' can be seen as quantifiable and also related to creating work in an organisational context. The concept of effort is further described in Section 3.3. The term 'security' refers to a state where a system and/or data is free from compromise. 'Privacy' describes an individual's right and/or ability to control how their data is being shared. In terms of the difference between security and privacy, in this dissertation, the simplified view is adopted that while users would not like some of their data to be compromised (security), they have the desire to share personal details about themselves depending on context (privacy). The Glossary Section (p. 213-214) contains more detailed definitions of these and other key terms.

1.4 Publications

Below is a summary of the author's publications during her PhD outlining her contribution to the six studies and the resulting publications.

Authentication in an organisation

This study was conducted by Angela Sasse and collaborators from NIST, Michelle Steves and Mary Theofanos, as well as Dana Chisnell from UsabilityWorks. The study was designed and carried out before the commencement of the author's PhD and she was invited to help with the analysis. Two main publications came out of the work: a NIST Report (#1) and an article at HCI International 2014 (#2). For the report (#1), the author's analyses directly contributed to Chapter 6 of the report and the author wrote a first draft of this chapter. The publication went through several rounds of edits and revisions with Hannah Wald from Booz Allen Hamilton being a technical writer on the report. The author of this dissertation wrote up article #2 based on the study and presented it at HCI International. The publication focused on the qualitative results of the study and it was mostly based on her analyses. Angela Sasse gave the article structure and provided ideas for the solutions that could be implemented to address the problems discovered. Chapter 4.2 contains elements of this article but it was substantially re-written to include the author's other analyses in line with the focus of this dissertation.

1. Steves, M., Chisnell, D., Sasse, A., Krol, K., Theofanos, M. & Wald, H. (2014). Report: Authentication Diary Study. NISTIR 7983.
2. Sasse, M.A., Steves, M., Krol, K. & Chisnell, D. (2014). The Great Authentication Fatigue – And How to Overcome It. HCI International 2014, 6th International Conference on Cross-Cultural Design.

Two-factor authentication in online banking

The study was part of Eleni Philippou's MSc project in 2014. It was jointly designed by Emiliano De Cristofaro, Eleni Philippou and the author. Eleni Philippou ran the lab sessions. The author analysed the quantitative and qualitative data and wrote up the article (#3) with Emiliano De Cristofaro's edits and comments. It was published at the NDSS Workshop on Usable Security (USEC 2015). Section 4.3 contains an edited version of this publication.

3. Krol, K., Philippou, E., De Cristofaro, E. & Sasse, M.A. (2015). "They brought in the horrible key ring thing" Analysing the usability of two-factor authentication in UK online banking. NDSS Workshop on Usable Security (USEC 2015).

Vernitski Authentication Grid

Constantinos Papanicolaou conducted a laboratory evaluation of the Vernitski Authentication Grid as part of his MSc project in 2013; he created a working prototype of the grid and ran the study sessions. The author designed the study, analysed the quantitative and qualitative data and wrote up the article (#4) published at HCI International 2015. Section 5.2 contains an edited version of this publication.

4. Krol, K., Papanicolaou, C., Vernitski, A. & Sasse, M.A. (2015). "Too taxing on the mind!" Authentication grids are not for everyone. HCI International 2015, 3rd International Conference on Human Aspects of Information Security, Privacy and Trust.

Human verification mechanisms

The study went through several cycles of design with feedback from other members of the team. The author ran the sessions with all participants in the study and analysed the quantitative data. Simon Parkin and the author jointly coded and analysed the qualitative data. The author wrote up the initial draft of the article (#5) published at the NDSS Workshop on Usable Security (USEC 2016). The second publication

based on the study (#6) was an article at the IEEE International Conference on Identity, Security and Behavior Analysis (ISBA 2016). This publication focuses on participants' perceptions of biometrics using the example of NoBot. The article received a Special Contribution Award at ISBA 2016.

5. Krol, K., Parkin, S. & Sasse, M.A. (2016). Better the devil you know: A user study of two CAPTCHAs and a possible replacement technology. NDSS Workshop on Usable Security (USEC 2016).
6. Krol, K., Parkin, S. & Sasse, M.A. (2016). "I don't like putting my face on the Internet!" An acceptance study of face biometrics as a CAPTCHA replacement. IEEE International Conference on Identity, Security and Behavior Analysis (ISBA 2016).

Quantifying disclosure

Sören Preibusch designed and implemented the study on Amazon Mechanical Turk. The author was responsible for coding the quantitative and qualitative data. While Sören analysed the quantitative data, the author analysed the qualitative data. The article (#7) published at the Workshop on the Economics of Information Security (WEIS 2012) had an economics focus. Section 6.2 briefly summarises the relevant findings and presents additional analyses in line with the focus of the thesis that were not part of the publication.

7. Preibusch, S., Krol, K. & Beresford, A.R. (2012). The privacy economics of voluntary over-disclosure in Web forms. Workshop on the Economics of Information Security (WEIS 2012).

Informing disclosure

The study followed from the previous one that focused on quantifying disclosure. Sören Preibusch and the author jointly worked on the design of the study. The author coded and analysed the qualitative data while Sören analysed the quantitative data that the author had previously coded. Article #8 is a position paper that the author wrote based on the findings from both disclosure in webforms studies. Some ideas from this article are contained in Chapter 6. The author also wrote up article #9 which was published at the Workshop on Privacy in the Electronic Society (WPES 2016). Chapter 6.3 is an edited version of this publication.

8. Krol, K. (2014). "Wait: That's optional!" Designing helpful over-disclosure alerts. Designing Human Technologies (DHT) 2.0.
9. Krol, K. & Preibusch, S. (2016). Control versus Effort in Privacy Warnings for Webforms. Workshop on Privacy in the Electronic Society (WPES 2016).

Publications spanning multiple studies

During her PhD, the author also published the following articles which draw upon the results of more than one study presented in this dissertation.

10. Krol, K., & Preibusch, S. (2015). Effortless Privacy Negotiations. IEEE Security & Privacy, 13(3), 88-91.

This magazine article focuses on offline and online privacy decisions and suggests lessons learned from the physical world can help develop better solutions for privacy in the online world. It contains a high-level summary of both disclosure in webforms studies.

11. Parkin, S. & Krol, K. (2015). Appropriation of security technologies in the workplace. Workshop on Experiences of Technology Appropriation: Unanticipated Users, Usage, Circumstances, and Design at ECSCW 2015.

This is a position paper for a workshop that discusses the impact of security on appropriation of technologies in the workplace. The author of the thesis contributed examples from the authentication diary studies.

12. Parkin, S., Krol, K., Becker, I., & Sasse, M.A. (2016). Applying Cognitive Control Modes to Identify Security Fatigue Hotspots. Security Fatigue Workshop at the Symposium on Usable Privacy and Security (SOUPS 2016).

This is a position paper discussing security fatigue. To support the arguments made, the author used the findings from both authentication diary studies and the study on verification mechanisms.

13. Krol, K., Spring, J.M., Parkin, S., & Sasse, M.A. (2016). Towards robust experimental design for user studies in security and privacy, The LASER Workshop: Learning from Authoritative Security Experiment Results.

This article proposes five principles for conducting methodologically robust experiments in usable security and privacy. The author of the thesis devised the five principles and the studies in this dissertation are used as examples.

1.5 Academic service

During her PhD, the author acted as a reviewer for the European Conference on Information Systems (ECIS) 2013, the Privacy Enhancing Technologies Symposium (PETS) 2014, the Information Systems journal (2014), the IEEE Security and Privacy magazine (2015) and the SIGCHI Conference on Human Factors in Computing Systems 2016. The author was also a volunteer at the HCI International 2015 conference.

1.6 Research contribution

The research presented in this dissertation makes several substantive and methodological contributions towards understanding the role of effort in users' security and privacy behaviours online.

1.6.1 Substantive contributions

The diary studies on authentication propose a taxonomy of coping strategies. The studies show that too much security-related effort can lead to abandonment of provisioned services and devices, having an impact on individual and organisational productivity. Users replace secondary security tasks with other tasks that have an acceptable level of effort. Their willingness to expend effort is associated with the level of security they believe they require.

There is a paucity of studies that would evaluate security mechanisms post-adoption. This dissertation fills the gap that studies often only consider user performance with proposed mechanisms rather than how deployed mechanisms are appropriated and how the effort they require fits in with existing tasks. In this dissertation, several mechanisms are evaluated post-adoption: various authentication mechanisms in an organisation, two-factor authentication for online banking and reCAPTCHA. Evaluating mechanisms post-adoption provides insight into how users have incorporated security technologies and processes into their every-day lives. Two studies in this dissertation also examine proposed mechanisms never evaluated before: the Vernitski Authentication Grid and NoBot. The results show that even when presented with a mechanism in the lab, participants see them very contextually reflecting in what situations they would be willing to use them and why.

Two studies on information disclosure benchmark the amount of information users provide in webforms. They explore user motivations behind expending the effort

to disclose and assess the role of mandatoriness and incentives in the process. The second of the studies evaluates the wording and options for warning messages. It shows that rather than deleting all optional information indiscriminately, users would like to have control over how much they share.

1.6.2 Methodological contributions

The studies in this dissertation use robust research methodologies often involving multiple stages of collecting both quantitative and qualitative data. Actual behaviours are the primary focus and the studies do not solely rely on self-reports as these suffer from various shortcomings (as discussed in Chapter 3). Nevertheless, post-experiment interviews and feedback help interpret participants' reactions, and discrepancies between actual and perceived effort are thematised as well.

1.7 Overview of studies

Table 1.1 provides an overview of the studies presented in this dissertation.

Study	Authentication diaries		Auth. and verification mechanisms		Disclosure in webforms	
	Auth. in an organisation	**2FA in online banking**	**Vernitski Auth. Grid**	**Verification mechanisms**	**Quantifying disclosure**	**Informing disclosure**
Described in	Section 4.2	Section 4.3	Section 5.2	Section 5.3	Section 6.2	Section 6.3
Research methods used	Auth. diary, interview	Auth. diary, two interviews	Laboratory evaluation, interview	Laboratory experiment, interview	mTurk experiment, questionnaire	mTurk experiment, questionnaire
Sample size	23	21	36	87	1,500	4,620
Effort captured through	Self-reports, authentication events	Self-reports, authentication events	Timing, self-reports	Timing, self-reports, NASA TLX	Timing, self-reports	Timing, self-reports, ratings
Data collection	July – August 2011	August 2014	August 2013	October 2014 – February 2015	January – March 2012	October – December 2012
Main publications	Steves et al. (2014) [242], Sasse et al. (2014) [225]	Krol et al. (2015) [145]	Krol et al. (2015) [142]	Krol et al. (2016) [144], Krol et al. (2016) [143]	Preibusch et al. (2012) [193]	Krol (2014) [140], Krol and Preibusch (2016) [147]

Table 1.1: Overview of the six studies presented in the dissertation.

Chapter 2

Literature review

This chapter reviews relevant research in the areas of psychology, human-computer interaction (HCI) and usable security and privacy. First, it synthesises the literature on effort and decision making and justifies the theoretical stance taken in this dissertation – the theory of bounded rationality and heuristic decision making. Second, it describes models developed to conceptualise security and privacy decisions. Third, the chapter clarifies the research approach taken in this dissertation that embraces the user with their capabilities and needs rather than forcing them to comply. Fourth, it discusses the principles of security and privacy behaviours as uncovered by research to date. Fifth, the literature review focuses on several mechanisms and the research conducted on them as they will be the focal point of the studies presented in this dissertation. The areas described are: authentication, two-factor authentication, CAPTCHAs as well as security and privacy indicators.

2.1 Effort and decision making

The principle of least effort was first discussed by Ferrero in 1894 [76]. According to this principle, humans choose the path of least resistance or effort to achieve their goals. The principle has been applied to many fields since – ranging from the ways people choose words [284] to the ways they seek information [79]. More than a century ago, the need to reduce effort led to the development of early computing machinery deployed at scale [258].

2.1.1 Effort in economic theories

Humans invest effort in order to obtain something they want in return. For that reason, it is plausible to look for a theory that would govern human behaviour in economics. Amongst the earliest models to explain humans' trade-offs in weighing the costs and

benefits of personal data disclosure is the *privacy calculus*. Culnan and Armstrong [59] take the model of a fully rational utility maximiser for whom the risks of disclosure must not exceed the benefits thereby unlocked (e.g., completion of a purchase or sign-up for a service free of charge) and who considers if *"their personal information will subsequently be used fairly and they will not suffer negative consequences"* [59, p. 106]. Whilst the original concept interprets the negative consequences narrowly (e.g., the monetary losses following data breaches or price discrimination), they can be understood more broadly to encompass all costs associated with disclosure, including the physical and cognitive effort of recalling and entering information. Revealing details about oneself thus becomes an input to a production function in the classic interpretation of economics [108]: a range of inputs (that is personal data items) would yield a desirable output, such as the completion of an online transaction. Users achieve efficiency by minimising their inputs for a given output, that is they seek to reduce their effort.

Although the interpretation of users balancing the pros and cons of disclosure appeals by its simplicity, it hinges on the classic concept of the *homo oeconomicus*, the idea that individuals would make judgements fully informed and fully rationally. Mill summarised that this approach sees the individual as *"a being who inevitably does that by which he may obtain the greatest amount of necessaries, conveniences, and luxuries, with the smallest quantity of labour and physical self-denial with which they can be obtained."* [168, p. 56]. This model has been found to be no accurate depiction of human behaviour and laboratory experiments into the behavioural economics of privacy have demonstrated that consumers' privacy decisions contravene their preferences (e.g., [235, 25]). This model also ignores the social drivers of disclosure, by which the individual would create their image in society [274, 163] and accumulate social capital [72]. Economic anthropologists such as Sahlins [214] and Mauss [164] pointed at tribal societies where economic exchange is based on reciprocity. Although contemporary societies tend to be market economies rather than gift economies, privacy research still asks whether reciprocity could explain information disclosure and trust [66].

2.1.2 Effort and mental shortcuts

Heuristics

One critique of the privacy calculus also highlights the sheer impossibility of making a rational decision about data disclosure. On the one hand, users lack insight into companies' data processing practices and are thus unable to assess the risks associated with disclosing their details. On the other hand, even if they had all relevant details, the amount of information would be too large to process – in particular for privacy-related decisions that occur during habitual web use. For example, McDonald and Cranor [165] estimated that reading all privacy statements a typical web user is confronted with would involve costs exceeding 201 hours a year, worth about $3,534 a year per American Internet user.

Users therefore resolve to *heuristics*, mental shortcuts that may not deliver the best solution, but are a good enough approximation. Simon [234] introduced the notion of *satisficing*, implying that individuals make choices that are not optimal but sufficient in the given circumstances. Simon suggested that humans use simple rules of thumb or heuristics to make decisions. They do not optimise because it is impossible to in each situation fully understand all variables and outcomes and estimate the expected utility of each possible course of action.

Contemporary heuristics frameworks, such as by Tversky and Kahneman [255] and Gigerenzer [91], build upon Simon's ideas and emphasise that heuristics help people in taking decision in a fast and frugal way. They are fast as humans are goal-oriented and a long process of decision making is holding them back from reaching their goal and frugal because the decisions rely on a limited amount of information. Heuristic thinking has been applied to the field of security and human behaviour by Stajano and Wilson [240] who proposed a range of seven decision principles individuals employ to make fast and frugal decisions and that are commonly exploited by criminals. For example, the *social compliance principle* refers to the fact that in society, people are conditioned to follow the instructions given by those in a position of authority. Online criminals have been exploiting this by posing as representatives of authoritative bodies and asking individuals to comply with their requests. Another principle the researchers identified is *time pressure*, where victims are compelled to act fast which results in them taking decisions without much deliberation. Both principles

have been exploited for example in phishing emails posing to be from the tax office urging individuals to pay taxes as soon as possible to avoid fines [110].

Effort has also been considered to be a heuristic. Kurger et al. [150] conducted a series of experiments to assess if effort is a heuristic for the quality of artwork. The authors conducted three experiments where participants were asked to assess the quality of a poem, a painting and a suit of armour, respectively. In each experiment, participants were provided with various pieces of information about the artwork including the name of the artist, the materials used, the weight of the object which varied depending on the artwork. Also in each experiment, participants were given the information on the time it took the artist to complete the artwork which the researchers treated as a proxy for effort. The results showed that indeed participants used effort as a heuristic for quality. They assessed the quality of the work to be higher the more time the artist spent on creating it. The effect of effort was moderated by ambiguity as participants' perceptions of quality were more influenced by effort when it was more difficult to judge the quality of the piece of art.

Thinking *fast and slow*

According to a theory by Kahnemann [128], heuristic decision-making could be categorised as "thinking fast". The idea that human brains sometimes think "fast" and at other times "slow" has its roots in a critique of expected utility theory. As mentioned above, utility theory treats humans as fully rational agents and does not account for cognitive biases. In response, Kahneman and Tversky proposed prospect theory [129] according to which when making decisions humans assign values to gains and losses as opposed to final assets. The value function tends to be concave for gains and convex for losses, that is it is steeper for losses than for gains. Instead of probabilities, the authors spoke of decision weights. The decision-making process comprises of two stages, termed *editing* and *evaluating*. The editing phase is a preliminary evaluation of decisions and outcomes where individuals create a simple representation of different prospects. In the evaluation phase, the prospects that resulted from editing are evaluated and the ones with the highest value are selected.

The idea behind thinking "fast and slow" is that humans can exhibit two modes of thought: "System 1" is a fast, impulsive and emotional way of thinking. It is often unconscious, requires little effort but is prone to biases. At the same time, "System 2"

is slow, deliberate and more rational, it is more likely to be employed for complex decisions. Kahneman [128] described a range of experiments that show that System 1 and System 2 thinking can often lead to different decision outcomes even when the inputs are the same. System 1 thinking is easier because it interprets new situations through the lens of existing patterns even if the new situation has little in common with previous ones. System 1 can be influenced by different biases such as anchoring, optimism and loss aversion, framing and sunk cost. *Anchoring* refers to the human tendency to be influenced by irrelevant numbers. If primed with an unrelated number, participants in studies have been shown to give higher or lower numbers as answers to questions depending on the value they were previously primed with (e.g., [255]). The *optimism bias* is the human tendency to think that they are less at risk of experiencing adversity and harm than others [231]. The *sunk cost fallacy* or *escalation of commitments* [241] refer to the idea that individuals are likely to continue with their predefined course of action despite negative outcomes rather than to abandon the entire endeavour.

In terms of effort, System 1 requires less effort and enables fast decision making, System 2 however, requires deliberation and time. A similar distinction between low and high effort decisions is reflected in Rasmussen's framework describing levels of performance which will be summarised in the next section.

Nudge

Thaler and Sunstein [249] built on the research conducted in the area of heuristics and proposed a differentiation between the automatic and reflective system in individual decision making. The *automatic system* corresponds to System 1 as described by Kahneman [128]; it is responsible for decisions taken fast and instinctively. The *reflective system* corresponds to System 2 which is activated for decisions needing careful deliberation, for example, concerning future life plans. Thaler and Sunstein [249] argue that since every-day decisions are the result of the autonomous system, *nudging* can help people take decisions that would benefit them in terms of health, finances and overall well-being. *Nudging* means changing the choice architecture so that a beneficial option is more likely to be chosen, without forbidding other options. A classic example is displaying fruit at eye-level in a shop without removing access to junk food. Section 2.4.6 provides examples of how the nudge theory has been applied to security and privacy behaviours.

Levels of performance

Rasmussen [194] defined three levels of performance which Reason [197] then later applied to sources of human error. These theories predominantly stem from the study of safety and are based on an industrial setting, often in aviation, where the goal is to avoid accidents. However, they have also been successfully applied to other domains such as healthcare and security. Rasmussen [194] specified three levels of performance: knowledge-based, skill-based and rule-based performance. Humans take a *knowledge-based approach* often in novel situations where there are no clear solutions or guidance. The individual has to apply conscious attention and adapt to the situation making this approach effortful, tiring and error-prone. *Rule-based performance* refers to situations where an individual pursues a goal in a conscious manner but the actions they have to take can be performed automatically. In this way, tasks can be performed in a near-unconscious manner which requires reduced effort. Finally, *skill-based performance* applies to situations where there is a match between the specific circumstances and the solutions that the individual was prepared or trained for in a work context. To illustrate these approaches in a security context, if a novel situation appears that an individual is not prepared for, they will have to expend effort and take a knowledge-based approach. With time, they might develop a workaround that undermines security but saves effort and will apply this solution in future cases to avoid fatigue and achieve their primary task.

The described theories have in common is that they emphasise that certain decisions are taken unconsciously in order to minimise the effort that is needed for deciding on a course of action. Similarly, Gardner [88] suggests that human brains have two systems for interpreting risk, one primitive and unconscious while the other one rational and conscious. If these two systems arrive at different conclusions, they can make humans worry about statistically rare risks such as child abduction and terrorism but ignore more frequent risks such as obesity and addiction.

Osman [178] argues that although the dominance of the unconscious is a popular idea and individuals might feel encouraged to follow their "gut feeling" [89], there is a general lack of evidence that unconscious decisions lead to better outcomes. Based on research conducted by Richardson [206], Osman argues that individuals can improve their performance through the mental practice of tasks. This corresponds to the levels

of performance specified by Rasmussen which meant that there are problems that individuals can be trained for. Taking a step back, regardless of who is right in the debate, the outcome is that humans either take fast and effortless decisions that are subject to biases or they can be trained through repeated effortful practice to react in desirable ways. In either case, this is not conducive to informed security and privacy decisions. In the case of effortless and unconscious thinking that minimises effort, as mentioned above, Stajano and Wilson demonstrated that heuristics can lead to detrimental security choices. In the case of providing training, it could be an option for organisations wanting to train their employees in security [20]. However, even in a workplace situation, training employees is only possible until a certain point [20] as employees' main focus is to meet their production goals. When it comes to individual users, it would be hardly feasible to train all Internet users in security and privacy. Taking the example of phishing, attackers' techniques are ever changing and currently becoming more sophisticated and any teaching would need to be updated on a regular basis making it effortful for the users to keep up with [190]. Faced with these dilemmas, this dissertation takes the stance that the design of security and privacy-enabling technologies should be conducive to promoting an informed security and privacy choice, minimising the effort that is involved.

2.1.3 Models explaining human protection behaviour

There are three classic psychology models that aim to capture what factors impact human behaviour and how humans are persuaded to take a particular course of action, in particular protective behaviours: theory of reasoned action [78, 7], theory of planned behaviour [6] and the health belief model [117]. They will now be discussed in turn from the point of view of how they thematise effort.

Theory of reasoned action

The theory of reasoned action focuses on the role of attitudes and behaviours. It suggests that the decision to take a particular action is influenced by the expected results that the individual is hoping to achieve. The theory proposes that two factors influence intention – the individual's attitudes as well as subjective norms. The theory has been used to predict how humans will behave based on their pre-existing attitudes and intentions. In terms of effort, the theory of reasoned action implies that a stronger intent

towards a particular behaviour would lead to an individual being ready to invest more effort in achieving it.

Theory of planned behaviour

The theory of planned behaviour was proposed by Ajzen [6] to expand on the theory of reasoned action by adding the concept of behavioural control. The theory states that the intention to perform a behaviour can be predicted based on three factors: one's attitude towards an action, subjective norms and perceived behavioural control. The term *attitude* refers to one's belief if a certain behaviour would make a positive or negative influence on their lives. *Subjective norms* refer to an individual's perception if their immediate environment (e.g., colleagues, friends) would approve of this behaviour. *Perceived behavioural control* relates to an individual's perception of how easy or how difficult it is to perform an action. This last factor is related to self-efficacy [15] that is the belief in one's ability to successfully accomplish a task. Thus, self-efficacy has an impact on how much effort an individual is ready to expend to achieve a goal. Ajzen states that *"holding intention constant, the effort expended to bring a course of behavior to a successful conclusion is likely to increase with perceived behavioral control"* [6, p. 184]. This means that an individual will be willing to expend more effort to achieve a goal if they believe they will be successful.

In the field of information security, Herath and Raghav Rao [111] applied the theory of planned behaviour to compliance with security policies in organisations. They found that while perceptions about the severity of a security incident, response efficacy and self-efficacy positively impacted attitudes towards security policy, the cost associated with a response had a negative effect.

The health belief model

The health belief model [117] is a theory aiming to explain what influences individuals' likelihood to adopt health-promoting behaviours. The theory suggests that three factors explain individual differences between taking health-promoting actions: perceived severity, perceived susceptibility of developing a health problem and perceived benefits and barriers. Later, self-efficacy was also added to the model [209] to emphasise that also a person's belief that they can successfully achieve a desired outcome plays a role. The model thematises effort as a factor when mentioning perceived barriers say-

ing that individuals weigh the perceived costs of taking an action against the benefits. The barriers may include inconvenience, time, pain or monetary cost necessary to take a preventive action.

Protection motivation theory

Protection motivation theory [207] was developed to explain the influence of *fear appeals* on decision making. Fear appeals refer to the fact that individuals can be persuaded to take a particular action, be it to buy a product or to cast a vote, by arousing fear. Protection motivation theory suggests that individuals take the decision to protect themselves based on the perceived severity of a threatening event, the perceived likelihood of it happening (e.g., their own susceptibility), the efficacy of the preventive action and their perception of their own self-efficacy. Confronted with a choice to take a protective action, individuals can cope through adaption or maladaption. Individuals will respond adaptively if they believe that *"the recommended coping response is effective [...] and that one can successfully perform the coping response"* [208]. Maladaptive responses emerge if individuals do not perceive a coping strategy to be effective and/or they do not believe they can enact a preventive behaviour successfully. Figure 2.1 summarises the cognitive processes behind adaptive and maladaptive responses. In terms of effort, Rogers introduces the notion of "response costs" to mean *"inconvenience, expense, unpleasantness, difficulty, complexity, side effects, disruption of daily life, and overcoming habit strength"* [208, p. 104].

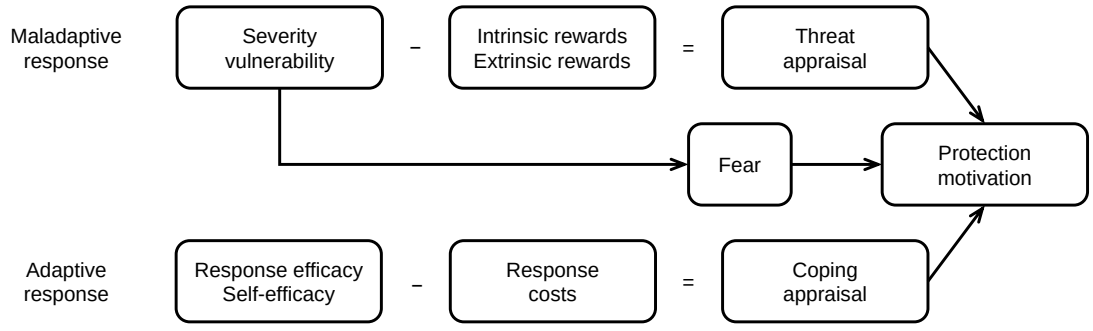

Figure 2.1: The cognitive process in protection motivation theory as proposed by Rogers [208].

Protection motivation theory has been applied to many areas such as healthcare and marketing; it has been also popular in information security research. While most researchers applied the theory to study the adherence of staff to organisational security

policy (e.g., [121, 189]), Jenkins et al. [125] applied the theory to authentication. They used fear appeals to prompt users to create unique passwords. The researchers used keystroke dynamics to detect password reuse and provided just-in-time fear appeals to encourage participants to choose a unique password for a website they were creating a password for. They found that around 88.4% of those who received a fear appeal registered a unique password. In comparison, only around 4.5% of those who were not confronted with a fear appeal created a unique password. While it is an interesting study in terms of detecting password reuse and using fear appeals, applying it to prevent password reuse seems rather flawed. In a piece of research that was published in the same year, Florêncio and Herley [82] demonstrated the sheer impossibility of creating a strong unique password for every system. The work by Jenkins et al. [125] seems to ignore this and fails to consider the consequences of making users create a new password for every system. For example, participants might have resorted to writing the newly created password down, which, depending on the threat model, might lower the security of that password. This is not to mention other costs associated with using just-in-time fear appeals such as: disruption to the primary task, task switching costs, frustration etc. Although the authors acknowledge that *"[r]arely is the act of "being secure" the primary purpose of using a computer"* [125, p. 208], they feel entitled to shift the user focus to a security task without considering if it is justified and what the cost could be. Following Herley [113], it is important to stress that human attention should be treated as a precious resource and used very wisely.

2.2 Conceptualising security and privacy decisions

2.2.1 Models of security decisions

Safety and security decisions

Rasmussen [195] developed a model showing the factors that influence employees' safety behaviour. He identified three main forces that impact individual behaviour: the security management of users, management pressure towards efficiency and gradient towards least effort. His theory specifies that individual actors adapt their behaviour based on criteria such as workload, cost effectiveness, risk of failure and joy of exploration. Albrechtsen [9] adapted Rasmussen's model to describe information security behaviours. As shown in Figure 2.2, there are three forces that influence performance:

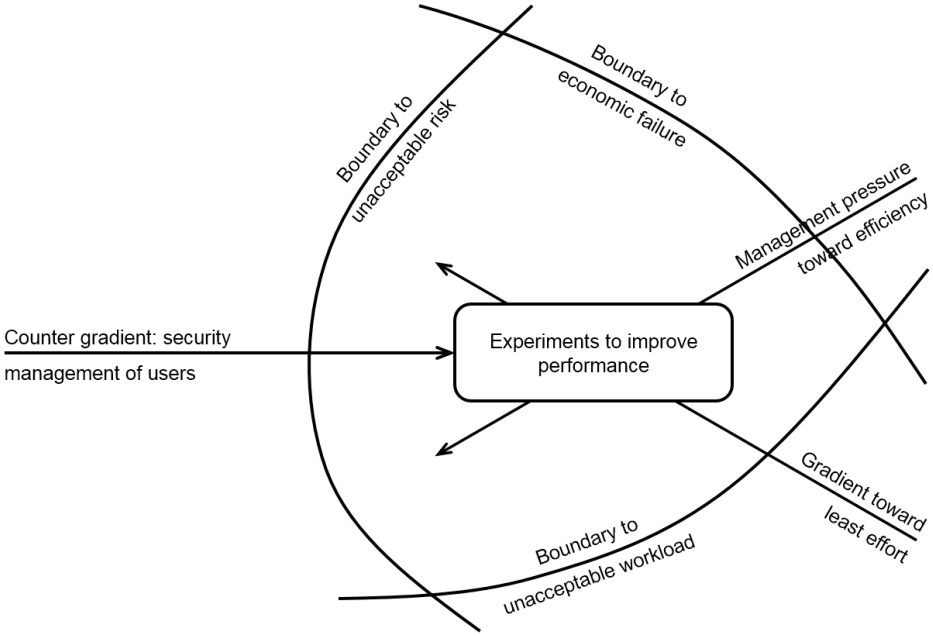

Figure 2.2: Albrechtsen's [9] adaptation of Rasmussen's [195] model for information security.

the management exerts pressure towards more efficiency while staff exhibit a gradient towards least effort. At the same time, information security management provides a counter-gradient pushing the movement away from the boundary of unacceptable risk. Albrechtsen stresses that *"high information security workload creates a conflict of interest between functionality and information security"* [p. 276]. According to Rasmussen [195], the fact that employees violate policy appears rational given their actual workload and time constraints they are under. While Albrechtsen and Rasmussen emphasise the important role of workload and individuals' predisposition to apply least effort, other prominent frameworks do not give it much weight.

Communication-Human Information Processing (C-HIP) model

The Communication-Human Information Processing (C-HIP) model by Wogalter [278] traces the signal from a security warning through a series of steps until it reaches the receiver. The path the signal goes through is shown in Figure 2.3. Effort is represented there only indirectly in the middle three factors – attention switch, attention maintenance and memory comprehension. Wogalter's model is problematic, it assumes conscious deliberation on the side of the user and seems to ignore habit. The human-in-the-loop security framework by Cranor [57] builds upon Wogalter's work. Although it does mention habituation as a factor in users' behaviour, it does not correct his generous assumptions of a conscious decision.

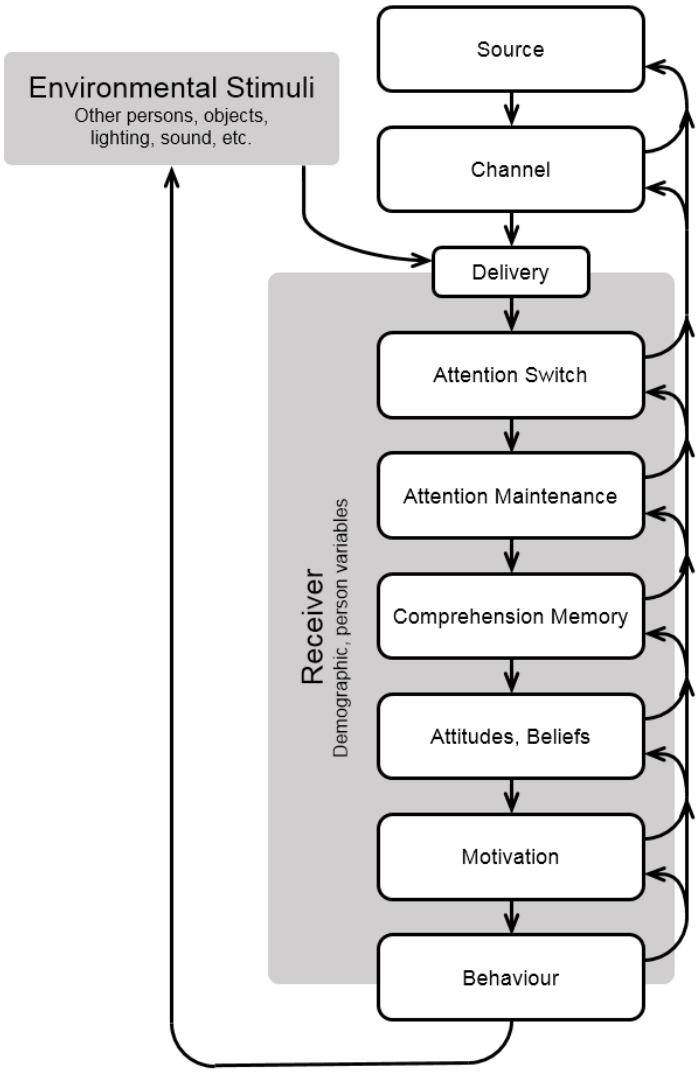

Figure 2.3: The Communication-Human Information Processing (C-HIP) model by Wogalter [278].

The models by Wogalter and Cranor both present the human as having vast resources in terms of time and effort that they can expend on making a decision. Their decision maker resembles the maximiser who weighs all the options to make the optimal choice. However, as described in Section 2.1.2, researchers in the areas of psychology and usable security have been identifying principles of human behaviour that are simple heuristics that enable fast decision making. Cialdini identified principles of influence [54], Gragg [98] spoke of psychological triggers, and Stajano and Wilson described the principles used in scams [240]. Recently, Ferreira et al. [75] made an attempt to combine these principles and proposed a unified taxonomy. In an early study [141], the author of this dissertation herself observed that participants were far from being maximisers, they took the decision to heed or ignore a warning quickly and

normally just relied on one cue, one heuristic, to make a choice. Trust was an important cue, it was either trust towards the anti-virus software, trust towards the university where the study took place or the lack of trust towards the warning and its issuer.

Compliance budget

Beautement et al. [20] proposed the idea that in an organisation employees have a *compliance budget* meaning there is a certain amount of effort they can invest in taking precautions to stay secure. Once this budget is exhausted, they become unwilling to comply. The authors emphasise that employee effort has to be tracked and managed wisely in order to *"focus the effort available on key security tasks to maximize their return on investment and avoid wasteful expenditure on less critical measures"* [20, p. 49].

Technology Acceptance Model

Although only peripherally thematised in usable security theory, effort has been named as a factor in other related areas. The central idea in technology acceptance literature is that perceived ease of use and perceived usefulness are key drivers in an individual's choice to use a piece of technology [62]. Davis defined 'perceived ease of use' as *"the degree to which a person believes that using a particular system would be free of effort"* [62, p. 320]. Later theories in the field, for example by Venkatesh et al. [261], suggested that four constructs have an impact on user acceptance of technology: performance expectancy, effort expectancy, social influence and facilitating conditions. They defined 'effort expectancy' as *"the degree of ease associated with the use of the system"* [261, p. 450].

2.2.2 Models of privacy decisions

Privacy model by Adams and Sasse

Adams and Sasse's [4] created a privacy model which stresses that privacy is contextual. This means individuals determine what they disclose based on the situation they find themselves in. To give an example, they talk about their strengths and weakness differently when talking to a future employer and when confiding in their best friend. As shown in Figure 2.4, Adams and Sasse [4] specified that users make their decision about disclosure based on how sensitive the piece of information is, who the information receiver is, and how they will use this information.

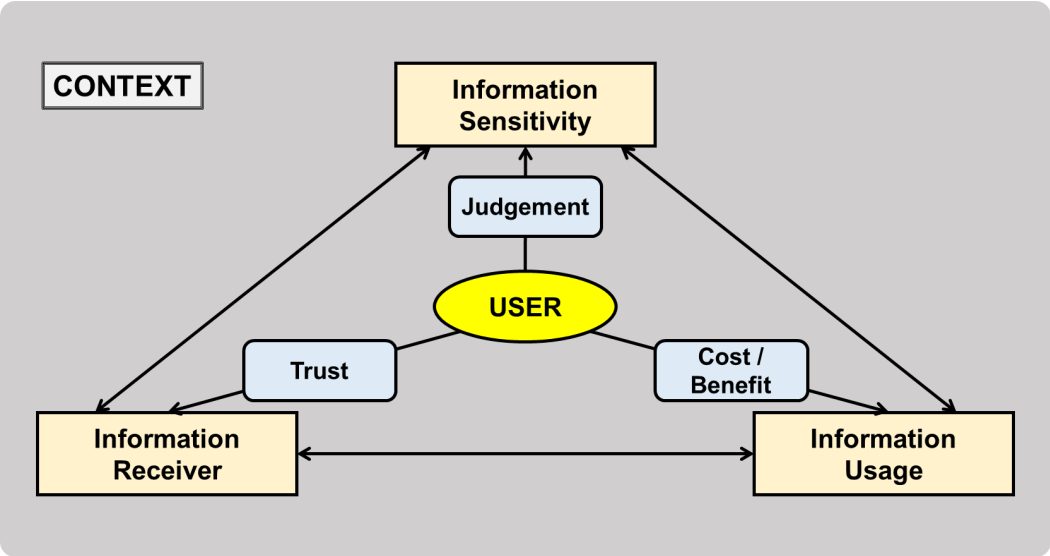

Figure 2.4: A model of factors and issues in privacy decisions by Adams and Sasse [4].

The first factor described by Adams and Sasse [4] was *information sensitivity*. The authors stressed that individuals perceive the sensitivity of their information on a spectrum rather than thinking of it as a binary – private versus non-private. Perceived sensitivity depends on context and the authors emphasised that information users are comfortable sharing in one context, they might be not fine with sharing in another. Regardless of context, research has shown that there are general information categories that users might perceive to be more sensitive than others – such as medical [2] and financial information [184]. Information sensitivity and the need to adequately protect it is also addressed by the law. The UK Data Protection Act (DPA) [44] specifies the following data items are sensitive: race, ethnicity, politics, religion, trade union status, health, sex life and criminal record.

Research has found that item sensitivity can be a determinant of disclosure. In a study by Metzger [166], disclosure and falsification were positively correlated with participants' perceived sensitivity of data items. Participants who disclosed information in the study were more likely to falsify sensitive information, in the given scenario this was data that could be linked to financial or identity records. Similarly, Braunstein et al. [42] demonstrated that users are more likely to limit the flow of personal data items when these are more sensitive. Additionally, studies have also shown that users are less likely to share information which would show them in a negative light [4].

The second factor described by Adams and Sasse [4] was *information receiver*. Users' decision as to whether to disclose a piece of information depends on who is

going to receive it. Research has shown that users are comfortable with sharing information with organisations they have a trusted relationship with [2].

Trust towards the information receiver is an important consideration when designing privacy-related studies. In order to remove the trust bias that participants would have towards university-based researchers collecting the information, researchers have gone to great lengths in constructing a story around who the information receiver is. Examples include pretending to be a credit card company [162] or an online retailer (e.g., [235, 25]).

The third factor described by Adams and Sasse [4] was *information usage*. Users base their decision whether to share their information on its expected use by the receiver. Relevance of the information in the given context matters. If a person considers a piece of information not to be necessary in the given context, they will deem the request for it to be intrusive [115]. In the context of a credit card application, Malheiros et al. [162] found that fairness played a role in participants' decision to disclose information. Fairness is understood as the users' belief that the data is being collected and used ethically and for the stated purpose.

Other studies have shown similar results. In a study by Metzger [166], participants were found to provide truthful information if they thought the information query was relevant in the given context. For example, they provided the correct shipping address since they expected this data would be used to successfully deliver a product to them. If users do not perceive the benefits of a disclosure to outweigh the costs, they can engage in privacy protection behaviours by either denying answers or falsifying them (e.g., [232, 58, 158, 173]).

Privacy concern

Research has repeatedly demonstrated a discrepancy between privacy preferences and actual behaviours (e.g., [126, 24, 55]). This has led researchers and the media to speak of a *privacy paradox* in that users say they are concerned about their privacy but share information about themselves online. This discrepancy could be caused by several factors, two possible explanations might be that (i) users seek immediate gratification or (ii) that these phenomena are studied using different research methods. The explanation related to immediate gratification could be framed in terms of "System 1" and "System 2" thinking as described by Kahneman [128] (summarised in Section 2.1.2).

An explanation for the privacy paradox could be that System 1 thinking is employed in the actual decisions on the spot and System 2 thinking when participants reflect about their values and attitudes when responding to surveys. What links to this is the fact that human behaviour is goal-oriented and when immersed in a task, users tend to ignore other concerns [141]. This will be further discussed in Section 2.4.1.

Expected benefit

Another fundamental motivator for the disclosure of information is the expected benefit. Users disclose information based on the benefits they perceive they will receive in return. The benefits can be considered as both (i) the service or goods they receive or (ii) the compensation or discount they obtain. (i) Malheiros and Preibusch [161] conducted a large-scale observational study of user behaviour in registration forms for Google, Microsoft and Yahoo. They found that users were more likely to invest time and effort in completing a registration process for websites that were offering them more value, for example product download and activation services. (ii) Privacy research has also attempted to put a price on the value of personal information. In a field study selling cinema tickets [127], one third of users were willing to pay one Euro in order not to provide their phone number. Studies have also compensated users for disclosing information. In an experiment at the London School of Economics and Political Science (LSE) [139], student participants received chocolate in exchange for providing items of personal information. A total of 372 students completed the study: 91% disclosed their LSE username and 14% their LSE password.

Effort in privacy behaviours

For privacy, effort has been examined in the context of answering data requests. Hui et al. [120] conducted a field experiment examining the impact of privacy statements and seals on disclosure. Their findings showed that there was a negative association between disclosure rates and the number of items requested. Other studies confirm this – if a high number of data items is requested, users tend to perceive the transaction as effortful and costly [13, 152]. Also the number of steps in providing information matters. When a service provider includes a step of verifying a phone number as part of their registration process, they can experience lower account creation rates than services that do not have this step [161].

Malheiros et al. [162] asked participants to rate the effort involved in answering data requests. They found that gender, children count and marital status were considered the most effortless to answer. The items of information considered to be the hardest were: weekly spending, childhood deaths and monthly income. The authors speculate that this might be due to the fact that the answers potentially required calculations and explicit recall of information.

2.3 Research attitudes towards users

Information security has considered users as part of their work from its early days. In 1883, Auguste Kerckhoffs devised six principles for military cryptography. The final principle stresses that systems *"must be easy to use and must neither require stress of mind nor the knowledge of a long series of rules"* [131]. What he spoke about can be interpreted as reducing cognitive effort that users have to expend in order to successfully protect a system. Similarly, in their seminal work defining principles of information security, Saltzer and Schroeder stressed "psychological acceptability" is necessary for the functioning of security [215].

What these early works stressed was both respect for users as well as the acknowledgement that systems cannot be secure if individuals cannot cope with the demands placed on them. However, since the early days of usable security as a research area, strands of academic work have been treating users in a condescending way speaking of humans as the "weakest link" [230]. In one of the first papers in the discipline, Whitten and Tygar [276] suggested users need to be educated about cryptography in order to be able to use encrypted email. This suggestion ignores the fact that it is neither feasible nor necessary for users to become security experts as it is neither necessary nor feasible for every driver to understand in detail how exactly a car works. However, such research approaches persist and more recently studies have worked on how to make users memorise randomly-assigned 56-bit codes [38] or how to force them to pay attention to warnings [10].

The research in this dissertation takes a different stance, it takes an approach focused on protecting the human rather than on preserving the security of information and assets. Security and privacy exist foremostly to protect and support humans in their primary tasks. Humans are goal-oriented and prefer not spend time performing secu-

rity and privacy tasks since they do not perceive them as enablers but as hurdles. Sasse and Fléchais [221] argued that often security mechanisms are chosen without considering their impact on users' actual tasks which negatively affects the user and undermines security. The research presented here embraces the user rather than "scaring and bullying" them into security [216]. It speaks not of human limitations and frailty [203] but of their capabilities and needs [23]. Following Herley [113], it stresses the fact that human effort is a precious and limited resource and therefore it needs to be used with great care.

2.4 Principles guiding security and privacy behaviours

Over the last two decades, research into usable security and privacy has identified several principles that influence user behaviours. The main principles can be summarised as follows.

2.4.1 Human behaviour is goal-oriented

Since human behaviour is goal-oriented (e.g., [46, 100]), asking users to perform a security task diverts them from their intended workflow [217]. Staying secure online can not only disrupt the interaction with a system or the use of a service but make it difficult or even impossible to accomplish the primary task.

2.4.2 Security and privacy are secondary tasks

Users are online to accomplish a certain goal and security and privacy choices are just a secondary task for them. Sasse et al. [220] explained that if something gets in the user's way of completing their actual task, they will attempt to circumvent it and they will gradually develop an aversion to it: *"the extra effort required will foster resentment in users, and feed the perception that security is 'not sensible' because it interferes with real work"* [p. 128].

2.4.3 Security measures may cause more risky behaviour

Adams [5] introduced the concept of the *risk thermostat* stating that the introduction of a security or safety measure is likely to produce more risky behaviour. One example of this is the introduction of the mandatory wearing of seat belts in cars. Instead of lowering the number of deaths caused by car accidents, it has been shown that drivers felt more protected, drove faster, and this had the knock-on effect of creating more

danger on the roads. This phenomenon has been described in economics under the term *moral hazard*. A classic example is when companies engage in more risky behaviour once they have purchased an insurance cover. In usable security research, a number of participants in a study by Krol et al. [141] emphasised that they dared to engage in more risky online behaviour because they believed their anti-virus would protect them.

2.4.4 No tangible reward for secure and privacy-preserving behaviour

Security and privacy are abstract concepts. There is no immediate reward for an action in accordance with security procedures, the reward is that nothing bad happens [273]. While accessing a service or downloading a file gives immediate gratification, not having done it because of security and privacy concerns gives no tangible reward and can affect the completion of the primary task. Additionally, human behaviour is shaped by positive and negative reinforcement. If someone is rewarded for 'correct' behaviour, they are likely to repeat it. If they are punished for 'wrong' behaviour, they are less likely to repeat it. In the context of security and privacy, the reward is problematic. The reward for 'correct' security behaviour is the absence of negative consequences, there is nothing additionally rewarding happening. At the same time, adverse consequences hardly ever come immediately after a user's action. They may come in days, weeks, months or might never be visible to the user.

2.4.5 Involving users in decision making

There has been a debate around in how far users should be involved in security decisions. Cranor [57] proposed a framework aiming to help system designers reason about user involvement in securing these systems. The author proposed that humans should be generally kept "out of the loop" but acknowledged that this was not always feasible. She defined the following three strategies for creating user-friendly security systems. First, it is important to create systems that do not require user involvement in security-relevant decisions. Cranor stressed that it is best to keep users out of the loop because they often lack the knowledge and motivation to perform security tasks. Second, if some level of human involvement is required, designers should find ways to make security tasks easy and intuitive to complete. Third, if human involvement is unavoidable, users have to be taught to successfully protect systems through training.

Following Cranor, Egelman et al. [67] distinguished between three types of security systems: first, mitigations that are not visible to the user; second, mitigations that do something on behalf of the user and third, mitigations that communicate with the user and suggest a course of action. In terms of effort, mitigations of the second and third type require user attention and consume their effort, therefore, keeping the human "out of the loop" would be more desirable. The authors conducted a study into users' tolerance of security-related delays. Their results demonstrated that users were more likely to cheat and disengage if they were given an explanation not related to security or only mentioning security vaguely. The authors suggested that if a mitigation cannot be made invisible, users should be provided with an explanation in order to ensure their cooperation.

Wash et al.'s [268] conducted a range of studies into users' perceptions and management of Windows updates. They collected qualitative and quantitative data through interviews, surveys and actual logs which allowed them to compare perceptions of how participants thought their computers were running updates with how they were actually running updates. The researchers found that their participants had a range of misunderstandings around updates; not understanding the settings for automatic updates and when their computer was being updated. The authors speculated that keeping users out of the loop might have resulted in them not being able to understand the process and configure their settings according to their preferences.

2.4.6 Impact of defaults and guidance

Despite extensive privacy controls, studies have shown that 87% of users have default or permissive settings [97]. Research has demonstrated that existing defaults do not lead users to secure choices [85]. To address this, more recent studies have made attempts to guide or 'nudge' users towards secure choices. Turland et al. [253] created and evaluated an application that aimed to lead users towards more secure Wi-Fi networks. They found that the use of colour and ordering impacted which network their participants chose. Nudging has also been used in the field of privacy to guide users away from privacy-invasive applications [53] or influence how they disclose information on social media [265].

2.4.7 Habituation and desensitisation

Confronted with security alarms that are clearly false positives, users gradually become desensitised and develop the habit of 'swatting away' warnings. Krol et al. [141] showed that almost 82% or participants ignored a security warning that stood in their way of downloading a PDF file. Nearly half of the participants later explained they did not believe the warning was accurate and saw it as a "routine thing". Similarly, Akhawe and Felt [8] showed that users who click through 47% of SSL warnings made their decision within 1.5 seconds. Herley [112] emphasised the great effort users have to put into staying secure online and stated that adopting unsafe online practices is a rational rejection of security advice. He stressed that the effort is immense and the benefits are small since the likelihood of being attacked remains relatively low. Faced with ever-increasing warning fatigue, Sunshine et al. [247] suggested to block unsafe connections and reduce the use of warnings in benign situations in order to prevent desensitisation.

2.5 Specific areas of focus

2.5.1 Authentication

Today, most web users encounter security as authentication. Authentication refers to the process of verifying a user has the necessary information, token or biometric characteristic to prove they have the right to access a system. Traditionally, the user can be verified based on what they know (e.g., password or PIN), what they have (a token) and who they are (biometric characteristic).

Knowledge-based credentials

Knowledge-based credentials such as passwords and PINs have a long history and are very widely deployed nowadays. From the point of view of effort, passwords require unaided recall that is the user has to remember them exactly and be able to retrieve them from memory when asked. Having to create and memorise high numbers of passwords makes users re-use them for multiple accounts [81], write them down [3] or choose passwords that are easy to remember. Studies have demonstrated that people do not use the entire space of password creation possibilities but instead concentrate around a relatively small number of predictable choices [33]. This has been demonstrated also for graphical passwords [60, 260], PINs [37] and personal knowledge questions [35].

Several password database hacks and leaks have shown that the most common passwords tend to be 'password' and '123456' (e.g., [272, 33]).

With the emergence of touchscreen devices, there has been a shift away from a traditional screen and keyboard set-up towards virtual keyboards. Password entry is even more difficult on touchscreens because of screen size, the lack of familiar haptic feedback known from physical keyboards, different key layouts and screen depths used [99]. Previous studies have demonstrated that password entry on touchscreens can be significantly more difficult and time-consuming [16, 99]. Schaub et al. [226] investigated the usability and security of six types of virtual keyboards. They showed that the keyboards differed in the usability of password entry (entry time, accuracy) and susceptibility to shoulder-surfing. They found that keyboard designs with poor usability were more resistant to shoulder-surfing. Moreover, research has demonstrated that the entry method affects users' password choice and security as users choose passwords that are easier to type [16, 282].

At the same time, passwords are becoming easier to crack as computing power increases. In 1999, Schneier stressed that *"password crackers can now break anything that you can reasonably expect a user to memorize"* [229]. More than 16 years later, we largely still rely on users to memorise their passwords.

More and more websites offering services and goods ask users to create passwords for accessing their products. Bonneau and Preibusch [36] consider the demand for users to expend cognitive effort in terms of a *tragedy of the commons*. By asking users to create passwords, the service providers are depleting the common good of human memory. Also, being asked to create a password on every site makes users re-use strong passwords which results in lowering the strength of that password as some sites do not offer adequate protection. The work uncovers that the industry's standards are far from ideal, with many sites storing passwords in the clear.

Why are passwords still so common despite all their shortcomings? They tend to be easy to deploy, no explanation or training is required on how to use them as their use is wide-spread [34]. This has made passwords an easy one-size-fits-all solution that is considered to be applicable in all sorts of scenarios, for all types of systems [36, 114].

In order to ease memorability, many alternative password schemes have been proposed. One group of them is graphical passwords that seek to ease unaided recall by

substituting it with recognition. Although they rely on a simplified understanding of psychology [182], such schemes aim to tap into human visual memory, exploring the so called "Picture Superiority Effect". One such password scheme is called PassFaces[1] and it aims to tap into humans' unique ability to memorise and recognise faces. It falls under the category of *cognometrics* where the user has to choose an image from a group of distractors. For PassFaces, users are asked to remember an individual from a grid of 3×3 faces in a sequence of three screens of faces. User studies of PassFaces have shown the scheme is neither effortless nor secure. Brostoff and Sasse [46] found that login frequencies decreased when PassFaces were used. Participants logged in with one third of the frequency when they authenticated using a grid of PassFaces rather than passwords, because the former took significantly longer. Thus, users took the decision not to go through a login process that cost them time, cognitive and physical effort if they were to spend only little time on the system. Also, having multiple passwords for this system would be difficult as they would likely interfere as demonstrated in a study be Everitt et al. [73]. The security of PassFaces is also questionable since guessability has been shown to be high. In a study by Davis et al. [60], participants tended to choose faces based on their own ethnic background and the attractiveness of the pictured individuals.

GrIDsure[2] is another example of a graphical password and falls under the category of *drawmetrics* that require the user to recreate a shape or a simple drawing. For GrIDsure, users are asked to memorise a pattern and then when authenticating, they receive a grid filled with digits and need to enter the digits that correspond to their pattern. Brostoff et al. [45] studied the usability of GrIDsure and found that in nearly 18% of usages, participants were trying to enter the PIN on the grid directly instead of typing it. This undermines the security property offered by the grid, namely resistance to shoulder-surfing. Researcher have also studied the Android Pattern-Lock on mobile devices. Uellenbeck et al. [256] found that users commonly start their patterns in the upper left corner and choose a three-node straight line. Similarly, Andriotis et al. [12] showed that users choose sequences of points that are easy to predict and over 50% of participants started their pattern from the top left.

[1] http://passfaces.com/

[2] http://www.gridsure-security.co.uk/

Token-based credentials

A common example of a token-based credential is an access badge that the user has to swipe through or hold next to a reader in order to access a location. Other examples include RSA tokens (SecureID)[3] that generate a one-time password (OTP) or the proposed replacement for credentials – Pico[4]. While these are dedicated devices called *hardware tokens*, there are also *software tokens* – programmes that can generate a one-time password on a device such as a computer or a mobile phone.

In terms of user effort, possession-based credentials are considered to require less cognitive effort as the user does need to create, memorise and recall a secret, but they might be demanding physically as the user has to operate a token. However, from a cognition point of view, the user might not need to remember a password but they need to remember to have a token with them which might introduce mental strain leading to frustration.

Token-based credentials have often been evaluated as a second factor in two-factor authentication (e.g., [270, 271]) and there are only few studies that assessed them on their own. Payne et al. [183] evaluated Pico which utilises a main device, Pico, and multiple wearables, Picosiblings, to authenticate users to systems. They conducted 20 semi-structured interviews asking participants to handle prototypes created using plasticine and Polymorph. Three main factors played a role in participants' acceptance of Pico: inconvenience, risk perception and responsibility. Inconvenience *"was intimately related to the effort that participants imagined they might have to expend on using and remembering to carry a dedicated Pico device"* [183, p. 7]. For this reason, some participants suggested integrating Pico with another device or creating a Pico smartphone app. An app would reduce the cognitive burden of remembering to take a dedicated Pico device with them and the physical burden of carrying it. Apart from the potential annoyance of having to carry around Pico devices, participants also felt anxiety because of perceived risks. They worried about what they would need to do to ensure Pico's security was maintained and with this came a sense of responsibility. Since Pico is a physical token, participants felt they would need to be responsible for both keeping it secure and available (e.g., on their person when they needed it).

[3] https://www.rsa.com/en-us/products/identity-and-access-management
[4] http://mypico.org/

Biometric-based credentials

Biometric-based systems authenticate individuals based on measuring an aspect of their anatomy (e.g., facial features, fingerprint), some ingrained skill or behaviour (e.g., typing rhythm, handwritten signature) or a combination of the two (e.g., voice) [11].

In terms of user effort, biometrics have held the promise to solve many usability issues since the user has to neither remember a password nor carry a token with them. However, experiences of real-life deployments of biometric systems [218] have shown that current biometric systems are not always accurate, take time to authenticate the user and their use might be physically demanding due to poor design of user interfaces. One of the problems identified by Sasse [218] was that enrolment and verification were performed on different equipment which can be confusing for the user. Coventry et al. [56] also stressed that the enrolment stage should be used as an educational opportunity to familiarise the user with the technology as they showed that exposure influences the acceptance of biometrics.

It is common to ask users to make adjustments rather than to adjust the technology behind biometric interfaces. For example, users have to be instructed to provide their fingerprints in a way that the scanner can read them. A study by Theofanos et al. [250] showed that instructions helped improve the quality of fingerprints that participants gave. Older women were able to give the best prints despite the fingerprints being the least pronounced in this demographic. Although the authors find this result encouraging, one could ask why users should compensate for the shortcomings of technologies with their time, effort and frustration. In the case of unsuccessful authentication, biometric systems can make users feel bad about themselves sending a message that they are not within the range of a societal 'norm' [198]. For example, older users can feel stigmatised because biometric features change over time and for fingerprints, the papillary ridges on human fingers fade with age which makes it difficult for technologies to capture them.

Trewin et al. [251] explored the demands placed on users – in terms of time, effort, error and task disruption – of voice, face and gesture, in a within-subjects study. Although biometrics were recorded as being faster than using passwords, none of the mechanisms were considered usable. This is important as although a mechanism can be regarded as better than others, its 'absolute' measure of usability may be poor.

A recent study by Bhagavatula et al. [27] explored usability and participant experience of both face and fingerprint-based verification for mobile devices, through a lab study and structured survey. Both mechanisms were compared side-by-side with more traditional PINs within subjects. The study scenarios were designed with physical settings in mind, such as using the mechanisms while in motion or in a dark room, aiming to assess user and technology performance. The results show that participants preferred unlocking devices with a fingerprint more than a PIN. Participants had more problems with face unlock and many abandoned using it.

Heckle et al. [109] asked participants to role-play purchases in an online bookstore with role-played use of fingerprint technologies using simulated occurrence of errors. Based on the results, the authors suggested that application contexts with more apparent user benefits were seen as more usable and acceptable.

Evaluating authentication

Although several attempts have been made to assess different authentication schemes (e.g., [177, 200, 43, 28]), the work by Bonneau et al. [34] is considered to be the most comprehensive. The researchers evaluated 35 password-replacement schemes, comparing passwords with OAuth and OpenID, hardware and phone-based tokens. They did not conduct a user study, but relied on a set of 25 subjective metrics falling under three categories: security, deployability and usability. The following eight usability properties were considered: Memorywise-Effortless (U1), Scalable-for-Users (U2), Nothing-to-Carry (U3), Physically-Effortless (U4), Easy-to-Learn (U5), Efficient-to-Use (U6), Infrequent-Errors (U7) and Easy-Recovery-from-Loss (U8). These usability factors can be mapped onto different aspects of workload. While U1, U2 and U5 mostly relate to cognitive effort, U3 and U4 are associated with physical effort. U6 is related to the relationship between time and effort while U7 and U8 refer to recovery from errors and loss. Although the framework is an important staring point, the ratings the schemes received need to be verified empirically through evaluations with actual users. As the studies presented in Chapters 4 and 5 of this dissertation will show, not all authentication events are the same and all these usability factors are contextual depending on both the user and the circumstances they are in; for example, frequency of use, time pressure, value of the account and the device used.

2.5.2 Two-factor authentication

A combination of two or more authentication factors has been referred to as *two-factor authentication* (2FA). The goal of 2FA is to increase security by adding a one-time password to prevent password database breaches or guessing attacks.

In 2006, Braz and Robert [43] compared 14 different authentication methods based on their security and usability. They were among the first to discuss the usability of 2FA and suggested that, by increasing redundancy, 2FA strengthens security but may negatively impact the user experience. They considered biometrics as an example of a second factor and address the problems of accessibility (i.e., iris recognition might not work for every person), social acceptability and the effort involved in enrolment. Although the authors drew on usability guidance, a significant drawback of their work is that they did not conduct a study with actual users of 2FA.

Strouble et al. [245] analysed the effect on productivity of the Common Access Card (CAC), a smart card and photo ID used by the employees of the US Department of Defense (DoD). Out of 313 survey respondents, 35.5% answered that accessing their emails while off-site was more difficult. One of their participants explained: *"The introduction of the CAC card for home use has decimated the communications channels that our reserve unit has spent years developing. We are now looking at going back to paper bulletins with stamps"* [245, p. 198]. During the switch to the new system, 2/3 of users forgot their card in their readers. The researchers estimated that this equalled to the productivity loss of approximately 260+ person-years and the DoD spent $10.4M on time lost.

Gunson et al. [102] investigated perceptions of security and usability in the context of automated telephone banking. The researchers juxtaposed single and two-factor authentication methods and asked 62 participants to rate their experience on a 22-item questionnaire. When second factors of authentication were enforced, users felt more secure than when using only passwords or PINs, but at the expense of usability. Participants considered single-factor authentication to be more convenient and easier to use.

While the studies mentioned so far compared 2FA with passwords, another line of work presented comparative studies of different 2FA mechanisms. Weir et al. [270] investigated usability perceptions of three 2FA technologies (a push-button token, a card-activated token, and a PIN-activated token). They asked 50 participants to use

each option, share what they liked and disliked about them, suggest how they would improve the technologies, and answer a short usability questionnaire. Users were then asked to authenticate again with their preferred technology and fill out the usability questionnaire once more. Authors found a strong correlation between preference and perceived convenience, but only a weak correlation between perceived security and perceived convenience/usability.

In a follow-up study, Weir et al. [271] compared the usability of password authentication and two 2FA technologies, specifically OTPs generated by a token or received via SMS. They performed a lab study involving 141 participants and a 30-question survey. The researchers concluded that familiarity with a technology – rather than perceived usability – affected user willingness to use that technology. Users perceived the single factor method as being the most secure and most convenient option.

Both studies by Weir et al. had the drawback that participants interacted with prototypes, performing artificial tasks. They did not log in to their own accounts and they were not interacting with the authentication interface under real-life constraints (e.g., time pressure).

De Cristofaro et al. [63] presented an exploratory quantitative analysis comparing the usability of three different 2FA technologies, based on a survey completed by 219 US users recruited through Amazon Mechanical Turk. The authors found that participants were relatively pleased with the usability of 2FA, regardless of the specific technology, and regardless of motivations or context of use. However, the perceived usability was correlated with demographics as female and older people felt they expended more cognitive effort to use 2FA. Also, in contrast to previous studies, users' perception of trustworthiness was not negatively correlated with ease of use and the cognitive effort required.

In summary, prior work on 2FA usability used expert assessments, survey-based studies, and studies on prototypes, incurring a number of shortcomings. Expert assessments did not involve users, yielding findings that only rely on researchers' judgement and often without the benefit of a structured usability assessment technique, such as GOMS, heuristic evaluation or cognitive walkthrough. Survey-based studies asked participants to make hypothetical choices or report behaviours based on what they can remember. Finally, studies with prototypes were performed in the absence of real-life

constraints: without reference to a primary task – such as paying a bill – or context of use – paying a bill from one's office during lunch break or in a hotel room while travelling. There has been a lack of studies focused on actual users of 2FA post-adoption, which is crucial for understanding how people use different 2FA technologies and how the effort required fits into their every-day activities.

2.5.3 CAPTCHA

CAPTCHAs ('Completely Automated Public Turing tests to tell Computers and Humans Apart') are tasks users are asked to solve before they are allowed to access resources or make purchases. By deploying them, service providers want to prevent automated access to their offerings that would result in for example spamming, mass ticket buying or a denial of service. The CAPTCHA tasks, such as deciphering distorted letters, are intended to be easy to solve by humans but difficult to solve by computers.

Yan and Ahmed [281] defined a framework for assessing the usability of both text- and audio-based CAPTCHAs. The usability properties they considered are distortion, content and presentation. An evaluation by the authors, essentially as subject-matter experts, systematically reviewed these dimensions, identifying a number of factors which could reduce the solvability of a CAPTCHA for humans (e.g., use of unfamiliar character strings, use of colour). The approach taken by the authors is rather technology-centric as they only consider user performance rather than their goals and needs. The main usability issue of CAPTCHAs is the fact that they disrupt the primary task and ask users to expend effort to compensate for the shortcoming of technologies. Pogue [186] says users are together wasting 17 person-years every day on solving CAPTCHAs, calling this immense effort *"a disgraceful waste of our lives"*. There are technical solutions that could replace CAPTCHAs that would remove the need for user effort, but they are difficult to implement and costly for the service operators [77].

Bursztein et al. [48] also exhibited a technology-centric approach by asking *"How good are humans at solving CAPTCHAs?"* (p. 399). The authors conducted a large-scale study where workers on Amazon Mechanical Turk and an underground CAPTCHA breaking service solved more than 318,000 CAPTCHAs (from 12 image-based and 8 audio-based schemes). The findings demonstrated that humans find CAPTCHAs difficult with audio-based ones being particularly troublesome. The study

is an interesting one in that real users were asked to solve CAPTCHAs. However participants were motivated to do it by economic factors and they did not have an accompanying primary task as users do in real life.

There have also been attempts to expose users to CAPTCHAs in gamified ways. Ho et al. [116] proposed metrics for quantifying the usability of CAPTCHAs, by way of a game which tracks completion of CAPTCHAs as a measure of progress. Participants are engaged through a game, rather than directly through surveys or similar means, specifically as a way to manage study resourcing. The work posits that the most appropriate way to assess CAPTCHA usability is to ask many people to solve CAPTCHAs repeatedly. Metrics include completion time, typing error, and number of abandoned CAPTCHA solving attempts.

Although studies often measure the time needed to complete a CAPTCHA, physical and cognitive effort is only rarely considered. Gossweiler et al. [95] evaluated a CAPTCHA alternative based on image orientation, where users rotate 2D images to an upright position. The authors stressed that *"[u]sing CAPTCHAs, services can distinguish legitimate users from computer bots while requiring minimal effort by the human user"* [p. 841]. The study presented in Section 5.3 will assess if the effort required of users is as minimal as the authors claim it is. The evaluation by Gossweiler et al. included a "Happiness Study", to determine participants' preference for either a traditional text-based CAPTCHA or the image-orientation variant. Participants were asked to solve both types of CAPTCHA, five times each. They were then asked to state directly which method they preferred (as a free-text response), where 11 of 16 participants preferred the image orientation variant. In terms of physical effort, the paper noted that *"many users [of the text-based CAPTCHA] referenced feeling like they were at an eye exam while deciphering the text"* [p. 849].

Belk et al. [21] investigated the link between users' cognitive styles (by way of a psychometric-based survey) and both performance with text- or picture-based CAPTCHAs challenges as well as preference. The motivating observation was that CAPTCHAs should minimise cognitive effort. Results are reported for 131 participants who chose to complete either a text- or picture-based CAPTCHA (where the choice was recorded as a preference). Users overwhelmingly preferred text-based CAPTCHAs, where this is attributed to users' familiarity with text-based CAPTCHAs.

Notably, the work implies that users can both exercise a choice and have individual qualities informing that choice.

Reynaga et al. [205] compared nine CAPTCHA schemes on smartphones with alternative input mechanisms that aimed to increase usability. It is very topical research as touchscreens consume more physical effort, as outlined earlier in this chapter. The researchers found that although participants considered traditional input mechanisms to be error-prone, they preferred them due to familiarity. The study is notable for capturing a variety of subjective user perceptions (e.g., ratings for memorability, preference) and free-text responses from participants. When discussing their findings, the authors stressed that both correctness of solving a CAPTCHA and user perceptions ought to be considered when evaluating verification mechanisms.

In summary, the overwhelming majority of existing research on the usability of CAPTCHAs has examined how different types of CAPTCHAs compare in terms of solvability. The usability of CAPTCHAs has been narrowly defined as users' capability to decipher "squiggly" characters. In some cases, qualities of user perception are identified secondary to the laudable (but restricted) pursuit of less cognitively and physically effortful alternatives to CAPTCHAs. Studies aiming to evaluate the usability of CAPTCHAs have often failed to assess them under realistic conditions. Participants have almost always been invited to studies to solve CAPTCHAs which might be priming and in dissonance with reality where users go online to accomplish some primary task (e.g., make a purchase, contribute to a forum).

2.5.4 Security and privacy indicators

This section discusses existing research on warnings and indicators for security and privacy.

Passive versus active security indicators

Currently, security indicators can be divided into two groups: passive and active. Passive indicators are adjacent to the users' task and do not get in their way, at the expense of being overlooked at times. For instance, a passive indicator for a secure connection is placed on the margins of a website highlighted through colour. On the other hand, active indicators, such as pop-up warnings, seek to attract the user's attention by interrupting them in their primary task. The user has to then actively interact with the

indicator, as they cannot continue with their primary task without acknowledging the indicator and taking an action.

An example of a classic passive indicator is the padlock icon indicating an SSL connection. Early studies showed that only few users notice it [68, 280]. Schechter et al. [227] confirmed and expanded these findings. In their study, they asked participants to perform some banking tasks and presented them with increasingly alarming cues that the connection was compromised. They gradually removed HTTPS indicators, the participant's authentication image and finally replaced the login page with a warning page. Passive indicators were not effective in preventing users from logging in to a phishing site – no participant refused to enter their password when the HTTPS indicators were absent and only 8% did so when their chosen authentication image was absent. However, when a security warning page replaced the login page 64% (8 out of 22) participants refused to enter their credentials.

In their study on phishing, Wu et al. [280] showed that participants could be tricked into submitting personal information 34% of the time. Instructing them to focus on the toolbars in the browser did not make much difference since participants reported assessing the legitimacy of the website by how it looked and felt. That is, the locus of attention and the users' strongest signal was the page itself, not the chrome surrounding it. Participants appeared to be immersed in what they were doing; 45% of them stressed they did not pay attention to the toolbar because they wanted to finish their task. The authors conducted a follow-up study and demonstrated that pop-up warnings that interrupted users were far more effective than passive warnings shown in the toolbar. These pop-ups were not only active, but they were also displayed over the webpage, that is in the area of the screen participants were already focusing on. One weakness of the study is that due to implementation difficulty, participants did not use their own details but these of a fictitious person called "John Smith". It questions whether the results of the study have ecological validity since if participants are not made to believe they are at risk, they might not behave in a natural way and their reactions might not be generalisable to the real world [69].

Whalen and Inkpen [275] studied users' visual attention to security details using eye-tracking. They instructed their participants to first "browse normally" and then browse paying particular attention to security. The authors admitted that they might

have failed to reproduce normal browsing behaviour since the data participants were asked to enter was not their own. During this phase, participants did not pay attention to security indicators which can either mean that they ignore them during normal browsing or that they ignored them in the study as it put them in an artificial situation. In the second phase, when participants were asked to pay attention to security, the authors found that the padlock icon was the indicator checked most often (11/16) and seven participants who also checked the padlock, also looked if the website was using HTTPS. Interestingly, two participants looked for the padlock in the wrong corner of the screen and the authors suggest the icon should not be located in the periphery and browsers should standardise its position. The study also found that participants stopped checking security indicators after having logged in as they felt secure after having authenticated. Also, no participant checked certificate information to establish if the connection was secure. It appears that participants either did not think some security indicators were interactive or did not believe they would be able to interpret the information they provide correctly. Research by Biddle et al. [29] confirmed this. In a study with 40 participants, the authors evaluated the standard certificate information provided by Internet Explorer 7 and their redesign of the way the information is displayed. They found that the traditional design was more difficult to interpret for the participants due to the use of technical language.

Egelman et al. [68] juxtaposed passive and active browser phishing warnings. Their results corroborated findings from previous studies. Of those confronted with the active warnings, 79% closed the phishing website but only 13% did so when the passive warnings were used. The authors found that passive indicators were not significantly different from not providing any warning at all.

In a large-scale field study of SSL warnings, Akhawe and Felt [8] used telemetry to investigate user warning behaviour (e.g., click-through rates). They found that users hardly every click on explanatory links and thus, it is important to provide users with sufficient information in the main text of the warning.

Privacy indicators

There is a number of privacy indicators available. However, they rarely come from the browser and are more often add-ons or plug-ins created by researchers or open

source developers. Privacy indicators tend to be passive, for example the Privacy Bird[5] changes the colour and the content of its speech bubble depending on whether the website's privacy policy matches the user's set privacy preferences.

There are also privacy indicators around cookies, the most popular of them is Ghostery[6]. Ghostery displays a list of companies that are tracking the user when they are visiting a particular website. This browser add-on provides the user with information on the type of data being collected and offers the option to block certain cookies. A purple bubble with a list of trackers in the top right corner of the screen, Ghostery is a passive privacy indicator. Leon et al. [155] tested nine tools that were designed to help users limit online behavioural advertising, one of which was Ghostery. The researchers found that participants struggled to use the tools effectively due to confusing interfaces, poor feedback and the use of technical jargon.

Strictly active privacy warnings have not been examined so far. There has been little research on privacy warnings altogether. LaRose and Rifon [153] used the term "privacy warning" but under closer examination this warning was not an active privacy indicator since it did not respond to the users' actions. For example, active security warnings are triggered if a user attempts to download a file. Thus, it should be rather considered a different way of displaying a privacy policy rather than a warning.

In summary, there is a paucity of empirical investigations into privacy indicators. In particular, it remains under-explored whether the passive vs. active dichotomy would also hold for privacy as it does for security. Although both warnings are aimed at the threats arising from cyberspace interactions, one fundamental difference between privacy and security warnings is the spectrum or duality of intended outcomes. Where there is a universal agreement that more security is better, privacy indicators must cater for diverse privacy preferences as individuals do not unanimously prefer revealing less about themselves. Another important lesson from existing research is that passive indicators tend to be overlooked; therefore, a way to calibrate the design of an indicator could be by testing it in an active form so that participants have to act on it. Finally, the summarised studies emphasised the importance of using clear language in the text of indicators. If technical jargon is used, users will not be able to understand them and take an informed decision.

[5] http://www.privacybird.org/
[6] https://www.ghostery.com/

2.6 Conclusions for this doctoral research

The literature review has emphasised that effort is a basic driver of human behaviour as humans aim to minimise how much time, cognitive and physical work they invest to accomplish a goal. The reviewed psychological theories discussed the factors that impact individuals' readiness to adopt a certain behaviour. While the early theories stressed that stronger intent would results in more willingness to expend effort, subsequent theories tend to be more nuanced. They emphasise individual appreciation of the threat and its impact (severity, susceptibility), and mention expected results and self-efficacy as important factors. Effort is described as a cost which might act as a barrier preventing an individual from taking an action.

Current research into usable security and privacy lacks a holistic approach to studying the user perspective. It often fails to consider the fit between security and privacy tasks and users' every-day online activities. Little research has been conducted that would examine the role effort plays in users' security and privacy behaviours online. This doctoral research aims to address this research gap. For security, it focuses on authentication as it is the most common security 'task' that users encounter. Studies so far have only taken a very reductive view of effort treating it as user performance asking: "Can the user authenticate or verify using the mechanism?" They have failed to take a holistic and realistic approach to how authentication/verification effort interacts with the primary task and how it changes the ways in which users go about their activities in the long run. The studies described in Chapter 4 will examine the user experience of authentication post-adoption and Chapter 5 will evaluate the effort needed to use proposed authentication and verification mechanisms considering context of use.

For privacy, this doctoral research focuses on information disclosure as an example of where users explicitly share information over the Internet. Existing literature has identified a range of factors that influence how much users disclose online but effort has only received limited attention. Privacy might have more tangible rewards than security (discount today vs. maybe nothing bad will happen) and it might be the case that monetary incentives and other motivations outweigh the effort the user needs to expend. These hypotheses are tested in Section 6.2. To help enable informed disclosure, active privacy indicators are evaluated in a follow-up experiment presented in Section 6.3.

Chapter 3

Methodology

This chapter outlines the methods chosen to conduct the research in this dissertation. First, it classifies and discusses the different ways in which security and privacy mechanisms have been evaluated. Second, the methodology chosen for the studies in this dissertation is described and justified. Third, the chapter discusses the concept of effort, the ways in which it can be assessed and concludes by elaborating on how effort will be examined in this dissertation.

3.1 Research methods in usable security and privacy

Research methods can be divided into qualitative and quantitative ones. Qualitative research methods focus on interpreting meaning, in-depth exploration and constructing a rich picture [47]. A typical qualitative research method is interviewing where a researcher asks a participant questions and engages with their responses to elicit their perspective. Quantitative research methods investigate observable phenomena with the use of statistical and mathematical methods [47]. A typical example of quantitative methods are surveys with closed-ended questions and actual logs of user behaviours.

3.1.1 Evaluating mechanisms

When creating a usable security or privacy mechanism, user involvement is necessary at each stage of the development cycle [96, 179]. This should range from requirements gathering in the beginning to evaluating the mechanisms after deployment in order to gather user feedback and learn about experiences of adoption. Evaluation methods of designs, prototypes and functional mechanisms can be divided into two groups: analytical and empirical.

Analytical methods

In analytical assessment methods, an evaluator inspects a design and assesses usability based on a set of rules and techniques. The classic methods are the heuristic evaluation and the cognitive walkthrough [159]. An advantage of analytical methods is that they can be applied quickly and on a low budget. A disadvantage is that no real users are involved and methods such as GOMS explained below only approximate the use of a system as a best case scenario.

GOMS (Goals, Operators, Methods and Selection rules) and KLM (Keystroke-level Modelling) [49] decompose the interaction between a human and a computer to basic physical and cognitive actions – such as recalling a password (GOMS) or clicking a button (KLM). Once a given task has been decomposed into steps, standard times a practised user needs to complete each step are assigned. A GOMS-KLM analysis is traditionally done as a paper-and-pencil analysis, but it can also be carried out automatically using tools such as CogTool [1]. CogTool takes screenshots of interfaces and a specified path through them as input to produce an estimate for task time. The timings that are assigned to actions are based on KLM but some real timings are calculated based on for example, the distance to an object and size of the target according to Fitts' Law [80].

Empirical methods

When using empirical assessment methods, various measurements are taken while actual users interact with interfaces and complete tasks. Empirical user studies of authentication mechanisms can be divided into studies conducted in a laboratory and studies done in the wild.

Lab-based evaluations are currently the most common way of evaluating proposed mechanisms. User evaluations are often conducted by the developers of technologies themselves. For example, Fujita et al. [84] developed and evaluated a verification mechanism called Chimera CAPTCHA which asks users to recognise merged objects. The study consisted of participants completing CAPTCHAs and then answering questions about their experience. The study seemingly assesses user experience but is in fact just focused on *user performance* aiming to answer to question: "Can the user successfully prove they are human?" As many other studies in the field, it

[1] http://cogtool.com/

fails to ask questions around the acceptability of effort, the mechanism's suitability to various primary tasks and user preparedness to adopt it. The study suffers from further methodological shortcomings as it is based on biased questions. Participants were confronted with leading questions such as *"Is it easy solving the CAPTCHA?"* and *"If you choose 1 or 2 in Question 1, please write why you think that it is not."* (p. 54). This is an example of biasing framing where the participant might choose a score of at least 3 in order not to have to explain why they did not find the CAPTCHA easy. The question also suggests the researchers are not interested in what made the CAPTCHA easy for the participant. The effort considered here is perceived in terms of "being easy" rather than eliciting what makes it easy or difficult. Also, no primary task is given to the participants making them know exactly that their performance solving CAPTCHAs is evaluated. The CAPTCHA is assessed in isolation as there is no elicitation around the acceptability of the mechanism or how the effort required would impact a primary task. Similarly, in an evaluation of video-based CAPTCHAs, Kluever and Zanibbi [136] asked their participants: *"Which task do you enjoy completing more?"* (p. 6) suggesting CAPTCHAs are there to be enjoyed. These examples illustrate a wider phenomenon that asking users about their experience has failed to be balanced. For example, asking open-ended questions enables learning about the user perspective rather than just obtaining a confirmation of the results that the researchers are hoping for.

Studies in the wild (or field studies) take place outside a research lab in the real world and are thus considered to be a more realistic way of assessing mechanisms. Studies in the wild often require the development of sophisticated methods of data capture and/or the cooperation from service providers. Studies in the field can both assess user reactions to currently deployed mechanisms as done by Felt et al. [8] for browser warnings and they can also be used to assess the impact different proposed designs might have on users [74]. In terms of capturing effort, studies in the wild can offer data on timing, reactions and for authentication, statistics on successes and failures to log in. There have been attempts to *monitor and log authentication* events directly without user involvement [81], to avoid a bias introduced in for example authentication diaries. Log data can be taken from the authentication server or data can be collected in the user's browser directly. An example of an analysis of data from an

authentication server is Bonneau's analysis of Yahoo passwords [33] or more recently an analysis of password resets at a university by Parkin et al. [181]. To assess the overall authentication effort of individual users, their authentication events were logged in their browser in a study by Florêncio and Herley [81]. Based on an idea by the author of this dissertation, Nikolai Vorkinn developed an ethical password monitoring browser extension for Google Chrome as part of his MEng project in 2014-15 [263]. Future research plans are to create similar extensions to be able to collect passwords and PINs on different devices and platforms users authenticate to (e.g., phones, service terminals).

However, a major disadvantage of studies in the wild is that they rarely capture the user perspective, for example participants' explanation of what they thought had happened. They also raise ethical issues as it is harder to obtain user consent and the researcher is not there to provide reassurance in case of an emotional reaction. Thus, studies analysing logs can be complemented by in-depth interviews with a smaller subset of users to shed light on the their individual perspective and learn more about the perceived effort [181].

An interesting alternative to self-reports and log data are *diaries* which aim to address the shortcomings of both methods. In diary studies, participants are asked to record events and experiences as they happen throughout their normal activities. Diaries are employed in research scenarios where it would be difficult for a researcher to be present to make observations. They have the advantage that the time of capture of event details is close to the event happening so there is less dependence on a participant's recall. However, this also brings a number of disadvantages in that diaries might disrupt the natural flow of activities for the participant and they require time to complete. In order to ensure reliability, participants might need training so that they record events consistently [31]. As in the case of any self-reported data, the information collected through diaries depends on participants' accurate and truthful reporting. Diaries also do not capture when a user had the intention to perform an action but gave up on it [50].

Usable security research has used authentication diaries to understand authentication habits and usability issues. Inglesant and Sasse [122] introduced password diaries to capture the details of authentication events happening 'in the wild' and found that

frequent password changes were perceived as troublesome, that users did not change passwords unless forced to, and that it was difficult for them to create memorable passwords adhering to the policy. Hayashi and Hong [107] also analysed two-week diaries to derive the average number of passwords, frequency of use, recall strategies across 20 participants. Their participants recorded 1,500 events, with 8.6 online accounts per person.

3.1.2 Self-reported data

Self-reports, for example in surveys and interviews, rely on participants' recall and their willingness to answer questions truthfully. The use of self-reported data is particularly difficult for exploring sensitive topics [32]. Security and privacy can be regarded as sensitive topics and users might avoid giving truthful answers because they want to have control over the image of themselves they convey. The reason for this might be the social desirability bias, demand characteristics but also the fact that not behaving securely can be a breach of contract in a work environment.

The author's own study [141] also showed that self-reports are hardly any indicator of users' actual behaviour. The study investigated users' reaction to two security warnings – one phrased vaguely and one more specifically mentioning what the consequences of downloading a file could be. It was a between-subjects design with 120 participants randomly assigned to two groups: 60 participants were confronted with a generic warning and the other 60 with a specific one. The study consisted of an experiment, where participants' actual reactions were recorded and an interview. During the interview, each participant was shown a set of four warnings: one of them was the one they were confronted with in the experiment, another one the other experimental group was confronted with and the remaining two were distractors. For the warning the other group was confronted with, participants were asked what their reaction would have been if they had seen that warning in the experiment. For the specific warning, 47 (77%) participants who were not exposed to it in the experiment stated they would not have downloaded the file. In the experiment, however, only 14 (23%) participants in the specific warning condition refused to download the file (Table 3.1).

Interviews are an example of a research method for collecting self-reports. In the field of usable security, interviews have been widely used to study employee security behaviours in organisations (e.g., [220, 201, 30]). In this context, interviews can be

	Download	Refusal	Total
Observed reaction	46	14	60
Hypothetical reaction	13	47	60

Table 3.1: Hypothetical and observed reactions to a specific security warning demonstrated in a study by Krol et al. [141].

used as a diagnostic tool and the identified themes can help form a basis for more quantitative studies such as a company-wide survey [19]. Interviews, often combined with a debrief, can also be used to shed more light on participants' reactions in an experiment [141].

Surveys have many merits, they enable the researcher to reach a high sample size and the wording of questions is kept consistent [154]. However, unless there is open-ended questions (which actually produces qualitative data), the researcher can only find as much as they asked for and respondents can only fit their answers into pre-defined categories.

Despite all this, surveys and survey-like studies continue to be used for research on users' security and privacy choices. For that reason, it does not come as a surprise that many studies appear to have provided exaggerated and sensationalist results. In a UPI survey, 93% of users reported having concerns about the privacy of their health records [257]. In a different study, 70% of users were found to be worried about online tracking [254]. The bias is obvious when compared to real data. In a study by Metzger [166], 33% of participants reported having seen the privacy policy but actually only 18.6% clicked on the link to see it. Those who lied about having seen the privacy policy had significantly more Internet experience. Similarly, a study by Microsoft [167] asked "technologies elites" if they read privacy policies. Although more than three quarters indicated that they read privacy policies before accepting them, the actual logs showed that the number of users who read them was significantly smaller.

For at least a decade, research has thrived where privacy choices were only simulated and participants were presented with scenarios asking of them to make hypothetical privacy choices. Ward et al. [266] asked their participants how happy they would be about a membership in a bookstore where customers would receive discounts for providing personal information. Wathieu and Friedman [269] asked participants to

imagine an alumni association shared their personal information with a car insurance company for a 30% discount. Castañeda and Montoro [51] made an attempt at recreating a real-world experience by showing participants a mock-up of a website asking about their willingness to make a purchase from that site. A different take on this are studies where participants are presented with a number of scenarios and asked to rank them in order of preference [105, 104]. There appear to be very few studies that explore users' privacy choices through a real exchange [252, 25]. The experiment by Tsai et al. [252]; however, was relatively coercive and did not leave participants the option not to purchase anything and participants were obliged to buy a number of products. Participants were also allowed to provide a fake shipping address which limits how real the experience must have felt to them.

Surveys and studies examining users' hypothetical choices have also been erroneously described as experiments. For example, Modic and Anderson [170] set to study the effectiveness of a range of different security warnings. However, they asked participants how they would behave when seeing a warning rather than observing an actual reaction.

Another important issue is priming which refers to exposing participants to some information that might influence how they subsequently behave in the study [148]. If participants suspect that the real purpose of the study is to learn about their security and privacy behaviour, they might act unnaturally due to for example the social desirability bias. Although this is common knowledge and the use of deception is widespread in psychology and usable security and privacy [69], there are still studies where participants are told exactly what the researchers are interested in finding out [280]. This can seriously affect findings as participants might not behave as they would in real life.

3.2 Chosen approaches

3.2.1 Data collection

In face of the shortcomings of research methods relying on self-reports, the author of this dissertation chose a mixed research methods approach, recording both actual behaviours and the ways in which users make sense of them. The best approach is always to triangulate research methods and all studies in this dissertation collect both quantitative and qualitative data [64]. Another important goal of this research is to take

a holistic view of effort not only assessing if a security or privacy task is doable but if users find the effort appropriate and how it interacts with actual primary tasks.

The studies in this dissertation use *experiments* to capture participants' reactions when confronted with a stimulus. There are three experiments conducted here, the first one captures participants' interactions with verification mechanisms (Section 5.3) and the other two focus on participants' disclosure behaviours (Chapter 6). *Interviews* are used in four studies in this dissertation to shed more light on the quantitative data that was collected but never as a stand-alone method. In disclosure studies that were conducted remotely on Amazon Mechanical Turk, free-text responses in the feedback questionnaire after the experiment serve a similar purpose as interviews.

3.2.2 Analysis of qualitative data

While this work applies standard statistical tests to analyse quantitative data, there is a number of ways to analyse qualitative data. Grounded Theory [244] is an established method for analysing interview data in the social sciences. It has been applied to research in usable security and privacy (e.g., [3, 243, 183]). Grounded Theory is appropriate for studies where the researcher explores an uncharted territory and constructs a theory. The research in this dissertation, however, aims to gather evidence based on lessons from existing research and the themes it identified. It is confirmatory in nature and driven by hypotheses. Therefore, *thematic analysis* was chosen as a more appropriate method to analyse the qualitative data in this dissertation. As described by Braun and Clark [41], thematic analysis identifies themes in the data that relate to the research question at hand. The researcher analyses the data by applying codes to the transcripts and then aggregates these codes to create themes.

3.3 Assessing effort

For the purpose of this research, the author distinguishes between three ways of considering effort. (1) **Actual effort** is the type of effort that can ideally be objectively measured. Since cognitive and physical effort are difficult to measure, actual effort is easiest to capture in terms of time elapsed. (2) Users can also be asked about their perceptions of effort. **Perceived effort** can be prospective and retrospective, depending on if users are asked about their expectations of the amount of effort before a task or their perception of effort after the task. In both cases, it is a subjective assessment of how

much performing a task is going to or had cost the user. Like actual effort, perceived effort is difficult to measure reliably. The assessment of effort is often an unconscious process and asking participants for their assessment might distort the picture due to cognitive mediation. (3) **Effort can be modelled**. Modelling using HCI techniques, such as GOMS-KLM described in Section 3.1.1, can be used to establish a baseline for how long a task would take.

3.3.1 Subjective measurement of effort

Subjective effort assessment focuses on asking the participant to rate the effort as they perceive it. Proposed measurements include Eilers et al.'s [70] SEA-Skala and Reid and Nygren's SWAT which assesses mental workload [199] by treating it as a multidimensional construct that is composed of three factors: time load, mental effort load and psychological stress load. Vidulich [262] used judgment matrices for workload assessment developing the SWORD technique. However, the most widely used technique has been NASA Task Load Index (TLX), created by Hart and Staveland [106]. When using NASA TLX, the user is asked to indicate the workload involved in completing a task on six sub-scales: Mental Demand, Physical Demand, Temporal Demand, Performance, Effort and Frustration. Participants are asked to indicate the level of workload for each dimension on a graphical scale with 100 points. Then they are asked to make 15 pairwise comparisons of the six dimensions indicating which one of the two aspects was more important to their perception of workload. NASA TLX has been previously used for assessing the workload involved in completing security-related tasks (e.g., [202, 188, 181]).

3.3.2 Objective measurement of effort

In order to avoid cognitive mediation, *physiological measurements* have been employed to capture stress as a proxy for effort. Common indicators of stress are galvanic skin response, heart rate and blood volume pulse [277]. Eye-tracking has been shown to be a good indicator of people's attention shifts [259] and pupil dilation can be treated as an indicator of cognitive workload [187]. Physiological measurements have the disadvantage that the researchers cannot be sure what caused stress as it can be induced by other factors than the ones studied. This problem is even more salient for security and privacy which are secondary tasks. Taking physiological measurements of stress

also requires the participant to wear specialist measuring equipment which might make the interactions feel unnatural yielding less ecologically valid results. Thus, the studies in this dissertation focus on the time elapsed as it is both an indicator of effort and of the interruption to the primary task. NASA TLX is used to complement it and capture the different aspects of workload including cognitive and physical demand as well as frustration.

Chapter 4

Authentication diary studies

4.1 Overview

This dissertation investigates the role of effort in users' security and privacy behaviours online. The two studies described in this chapter examine user behaviours in relation to managing their authentication effort. As the first studies of the dissertation, they aim to diagnose the struggles users are facing and learn how to better design processes and mechanisms to reduce the effort required.

The first study aims to understand how authentication fits into employees' day at a governmental organisation in the US. Employees of that organisation were asked to keep an authentication diary of all their logins over the period of 24 hours and were later interviewed about their experience. There was a discrepancy between actual and perceived effort with some logins being habitual while others disruptive and poorly integrated into work processes. Interviews painted a picture of authentication fatigue caused by significant authentication effort that affected productivity and employee satisfaction. This section identifies participants' coping strategies and proposes a taxonomy of the ways in which employees seek to reduce the workload and disruption created by authentication.

The second study examines the user experience of two-factor authentication (2FA) for online banking in the UK. A three-stage study with two interviews and a period of diary-keeping showed that, on the one hand, participants appreciated the convenience of online banking but on the other hand, they found the use of 2FA tokens to be effortful and frustrating. Also here participants reported on a range of coping strategies that helped them manage their authentication effort and the taxonomy of coping strategies developed for the first study was applied to categorise them. The high authentication effort led to disengagement as participants started using a particular account less often

or switched to a different bank entirely. In both studies, participants expressed doubts if the level of security and the ensuing authentication effort was justified for the level of risk they were facing.

4.2 Authentication in an organisation

4.2.1 Study aims

The main aim of the study was to examine employees' overall authentication effort in an organisational context. Security mechanisms have often been studied in an isolated laboratory setting and they received a stamp of approval deeming them usable and the effort required manageable. The study presented in this section examined the authentication effort from the point of view of users rather than from the point of view of technology. The study also had the goal of assessing mechanisms and processes post-adoption, to see how users had appropriated mechanisms and devices in real-life conditions. Further, the studies thematised the relationship between the effort required by authentication and the primary task.

4.2.2 Method

Design

The study consisted of two stages, participants were asked to complete (1) diaries logging all their authentication events over the period of 24 hours and (2) were later asked to share their experiences of authentication and security in an hour-long interview.

Participants

Participants were recruited through a call for volunteers throughout the organisation using email. Of the 25 employees who responded, 23 (14 male and 9 female) provided their authentication diary data at the time requested and it was possible to interview 22 of them. Age of participants was: 20s (3), 30s (5), 40s (3) and 50+ (11). All participants were knowledge workers, 13 were researchers with a computer science related education or job. Three were researchers without a background in computer science (but in physics, cognitive science etc.). Two were IT systems administrators. The remaining two participants were administrative support staff.

Materials

Quantitative data in the study was collected through diary forms. A page from the authentication diary is shown in Appendix A. The layout and form fields were based on a previous diary study by Inglesant and Sasse [122] and also NIST employees provided their input to adjust it.

Procedure

In the first stage of the study, participants were asked to keep an authentication diary for the duration of 24 hours in July 2011, all of them on the same day. Participants received their form sheets in a briefing session before the diary day. In the second stage of the study, each participant was interviewed within a month of the diary day. The diary data served as a point of departure for the interview which explored the participant's individual perspective on authentication and security in general.

4.2.3 Diary results

In total, 23 participants recorded 528 authentication events during the 24-hour collection period. On average, they authenticated 23 times during that day (range: 4–40).

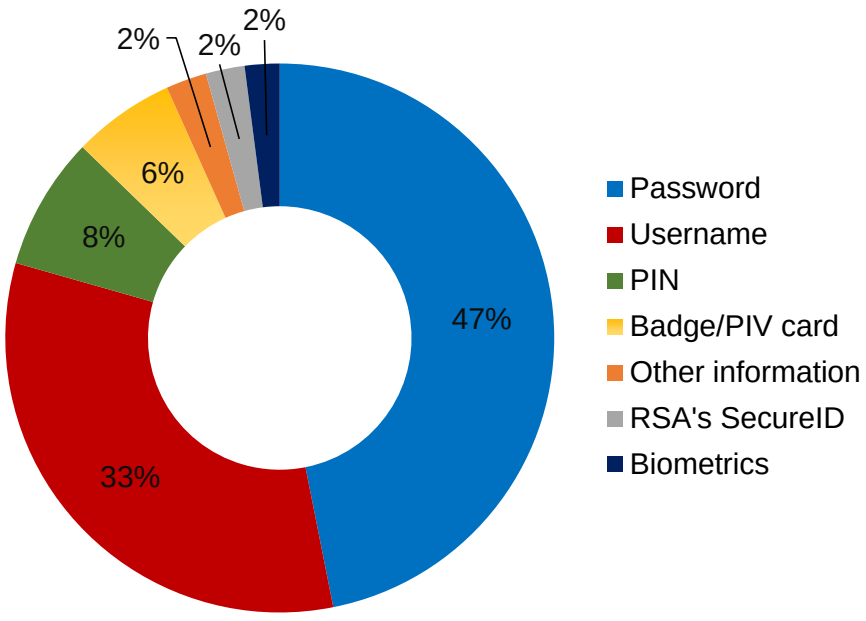

Figure 4.1: Frequencies of types of credentials required for authentication.

Figure 4.1 shows the types of credentials required for authentication. The most common credentials used to log in were passwords (47%) and usernames (33%). Passwords and usernames are almost always required together; however, the percentages

for both are not equal since in some cases usernames were already entered for the user (e.g., on lock-screens because of previous login). PINs (8%) and badges/PIV cards (6%) were the next common credentials used across the authentication events. PIV (Personal Identity Verification) cards are normally inserted to the computer to authenticate using a PIN. At the time of the study, all employees at the organisation had a PIV card with an associated 6-8 digit PIN. They can be used to log in in the same way as a username and password.

Out of the total of 528 events, participants experienced authentication problems in the case of 48 (9.1%). The most frequent sources of problems were mistyped (24 occurrences, 50%) and wrong passwords (7, 14.6%).

Of the 528 events reported in the study, 168 (31.8%) involved the use of some memory aid. From the point of effort (cognitive, physical or both), the memory aids can be divided into three groups:

1. Passwords are stored for the user.
   (a) passwords are entered for the user (no physical entry required and no mental effort of password recall): when stored in the client (e.g., in the browser, 13 mentions) or in a password manager (2 mentions);
   (b) passwords have to be entered by the user but are stored either on paper (11) or in a file (4) (physical entry required but no password recall).
2. Users create passwords to be memorable (physical effort required but mental effort reduced), when
   (a) a portion of a password stays the same as the previous one (1) or
   (b) passwords are based on a memorable phrase (e.g., song lyrics) (1).
3. Password entry is replaced with a biometric fingerprint scan (4) (mental effort removed and physical effort significantly reduced).

4.2.4 Interview results

The interview transcripts were analysed using thematic analysis [41]. The following sections present the themes that emerged.

Areas of high effort

The analysis of the interviews showed that there were eight aspects of authentication that required increased effort from participants which often led to frustration.

Frequent re-authentication. Participants reported struggling with re-authentication after a time-out. The computers in the organisation are set to lock after 15 minutes of inactivity to limit the window of opportunity for an attacker to exploit an unlocked machine while the employee is away. P21 emphasised this was unnecessarily causing stress: *"You end up having to almost set a timer in your head to go back to the computer and type something within every 10 minutes or so. And some minor studies of productivity I've been involved with indicate that it's better to be focused on a task as opposed to have lots of interruptions throughout the day."*

Infrequently used passwords. Participants reported having problems recalling passwords they used rarely: *"You have to think if you are going to be the responsible person, then you have to think of different passwords for everything and it is very frustrating. So when it gets to those infrequently used passwords, then I get really irritated."* (P11). P17 stressed: *"Once again security has gotten in my way and it takes me extra time because now I got to look that one up because I don't use that one often enough."*

Multiple passwords, their length and complexity. Participants reported struggling with password creation due to different requirements for each system. P19 explained: *"It can be a pain. And when I'm here, everything I log into, the password policy is completely different. So even if I wanted to, the password can't be the same. Some of them have to be eight characters, some of them have to be 12. So it can't be anything more than 12. So some passwords I have are 20 characters, some are 6. And it's just hard to remember everything, so I actually load a file on my computer that just has every password listed so I can just copy and paste it."*

Different password expiry intervals. Apart from different password change intervals, another obstacle to keeping one's passwords synchronised across applications is that some passwords expire at different intervals. To address both issues, many participants who synchronised passwords to some degree also used the complementary strategy of updating all their passwords at once.

Interruption of workflow. A number of participants emphasised the disruptive effect

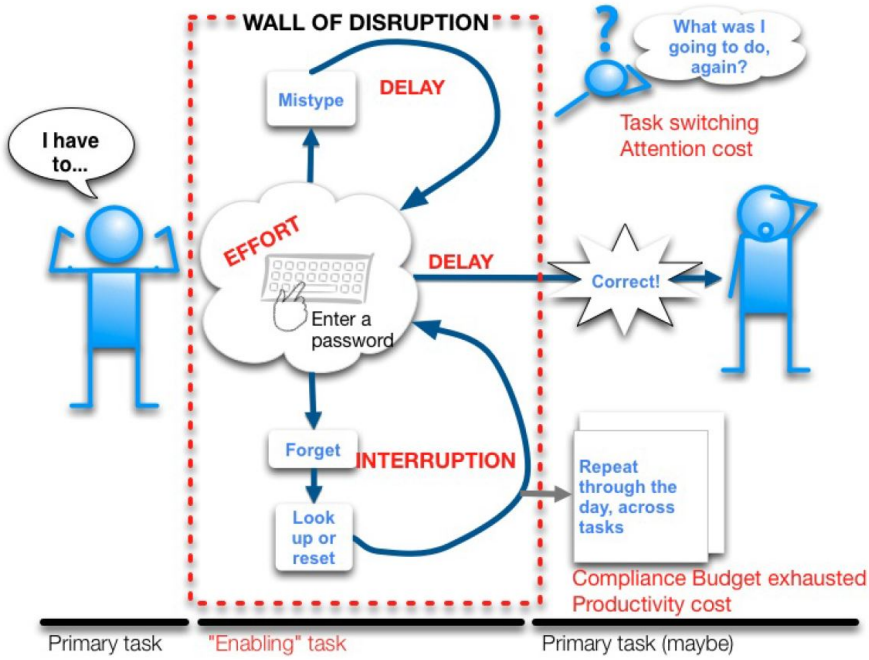

Figure 4.2: The wall of disruption authentication created (drawing by Dana Chisnell [242]).

authentication had on their work, they felt it had a negative effect on their productivity: *"And it gets in the way. It definitely takes way more time out of my day, both just time having to deal with this and then the break in the flow of work."* (P19) Any authentication task that requires time and effort creates a "wall of disruption" (Fig. 4.2) that affects productivity.

Similarly, it is evident from the analysis of diaries that most authentication events occurred around 9am, that is at the start of the working day, with another peak in authentication when participants came back from lunch after 2pm. P17 emphasised how waiting at the gate to be checked by the guard was annoying because they could not wait to start their working day: *"It was annoying [...] You've got to wind your way into the building, because, in some way shape or form, we are in a secure building, although there's nothing secure about it, other than the fact that you've got this card reader sitting out there."* Authentication and in general security was again a hurdle and participants used the time when they were highly productive to comply with security protocol that they believed to be unjustifiably strict.

Problems with tokens. Participants reported struggling with the usability of tokens. They found the two-factor authentication device – the RSA token – hard to use because (i) the display showing the six-digit code was too small and (ii) the change of number every 60 seconds was making the reading and entering of digits a cumbersome task.

At the organisation, RSA tokens are necessary for accessing the VPN (Virtual Private Network) from outside the campus. To authenticate, the user has to enter their username, PIN and a six-digit code given by the token, then their username and password. There is a problem of the system needing a few seconds to process the credentials at both stages and if the code expires before this is completed, the user has to start the authentication process again. P21 pointed several drawbacks of the RSA token when it comes to its physical usability: *"These are very low-resolution. You have to have bright light. It's frequently misread. [...] If you make a simple mistake, which it is easy to do. Sevens can easily be ones. Sometimes twos look like eights. If you make a mistake, then of course you've got to re-enter [...] There aren't clear signs on it to indicate which way it ought to be held and viewed. It looks like it's intended more to be held with your left hand than right."* P12 also found their token was out-of-sync and they sometimes needed up to three attempts to log in.

PIV cards posed some challenges to the participants. Only a relatively small number of participants used them on a regular basis, in part because they do not have the card reader and middleware or operating system. Apart from authenticating to computers, the card is also used to open doors, and P19 reported often leaving the card in the computer, exiting the room and then being locked out: *"So I would forget this every single time. And I've been locked out of the building. I've been locked out of the floor. I've been reprimanded for leaving this in the computer. And there's nothing I could have done other than somehow setting reminders every two minutes to don't forget this."*

Lack of integration. Four employees reported using a biometric fingerprint reader for some logins. One of the participants had physical usability problems with it since they needed to use a moisturiser to make their prints readable: *"This is worse, because it works by a biometric, and my finger doesn't read very well. I have to smear it with moisturizer, and then I forget what the actual password is. When it really just won't read my finger, I can't unlock it and I can't remember."* (P2). And in this organisation, the underlying infrastructure and policies had not been adjusted – the underlying password – which users avoid by using the fingerprint reader – expires every 90 days. When this happens, the user has to find their current password (often written on a piece of paper kept in the drawer), change the old password, write the new one down, and

enrol their fingerprint. This illustrates how the usability of authentication cannot be assured by simply replacing the 'front-end' but keeping the underlying policies and infrastructure the same.

Inadequate level of security. Throughout the interviews, there was a recurrent theme of employees thinking the level of security in the organisation was too high for the type of work that they were doing, P19 emphasised: *"If you're working for the CIA or a hospital with patient records, maybe I could understand that. [...] Nothing we do is sensitive. Everything is public, so they can get it anyways. It doesn't make any sense."* P21 felt they were being punished because of high-profile security breaches: *"Some of the security requirements on laptops especially, especially when traveling – resulted more from, I think, higher level screw-ups, people in upper level management having their systems being broken into."*

P3 emphasised that the organisation's approach to security has failed to consider the user perspective: *"I think that authentication effectiveness, the burden is placed on the user. And that I think is an issue, because providers just say, well, we need a secure wall, and they set it up. But I don't think they're doing usability studies before they do it."*

Perceived effort

Participants frequently addressed the effort required to authenticate. Some anticipated beforehand that the authentication effort would be greater than it actually was while others stressed that the experience of keeping a diary made them aware of the effort they had backgrounded due to habit and automation of the authentication process.

Foregrounding authentication. In the briefing meetings before data collection, some participants requested more diary forms as they were expecting to authenticate more often than the 40 events a standard set of forms could hold. After the diary day in the interviews, participants were asked what the biggest surprise about their experience was and seven out of 22 stated that they actually authenticated fewer times than they had expected beforehand.

In the interviews, P3 complained about how the use of RSA tokens was making authentication noticeable and disruptive: *"So actually taking the time away from the computer, I walk to get my bag. Someone else stops me in the hallway. I have a conversation with them. Then soon I go back and I remember to login, but I'm like,*

"Oh." The strong idea I had in my head of a message that I wanted to send might have become a little more fuzzy, the idea of what I was doing. "Why did I open that other new tab and not go to the…?" So there's the little things. I feel like those little things really, when you have an idea of what you want to do next and then you have to deviate from that, I think at least for my brain it throws you off a little bit." Similarly, P3 stated: *"It's that deliberate effortful, conscious…Ĩ really have to stop what I'm doing and think about it. Whereas if you're just doing something from muscle memory, you don't really even have to think about that."*

Backgrounding authentication. There were seven participants (32%) who emphasised how some logins were so second nature to them that they hardly noticed authenticating. This is an indication of habituation. Participants stressed that the passwords for some systems are in their muscle memory and they do not have to focus to recall and enter them.

Six participants stated that having to record their authentication events actually made them aware of how frequently they log in without realising, P11 stated: *"I was actually quite surprised by the number of times that I authenticated. I did not realize until I was asked to explicitly record and pay attention to these things that I was doing it so frequently. And if it's a password or whatever that I am very familiar with and very comfortable with, I just don't even think about it. I always get it right."*

Coping strategies

In the interviews, participants reported on numerous strategies for coping with unmanageable authentication effort. Figure 4.3 illustrates the stages leading to the emergence of coping strategies. Starting from the top of the graph, the entry of credentials itself causes disruption which is even greater if credentials are mistyped or forgotten. This affects the primary task leading employees to create coping strategies which are workarounds that aim to keep the authentication effort at a manageable level. The coping strategies can be divided into two categories: participants either tried to manage and be in control of their authentication effort, or to avoid authenticating, temporarily or permanently.

Management: Trying to stay efficient

Centralising. Most participants attempted to reduce their authentication effort by employing password managers or synchronising passwords across multiple systems:

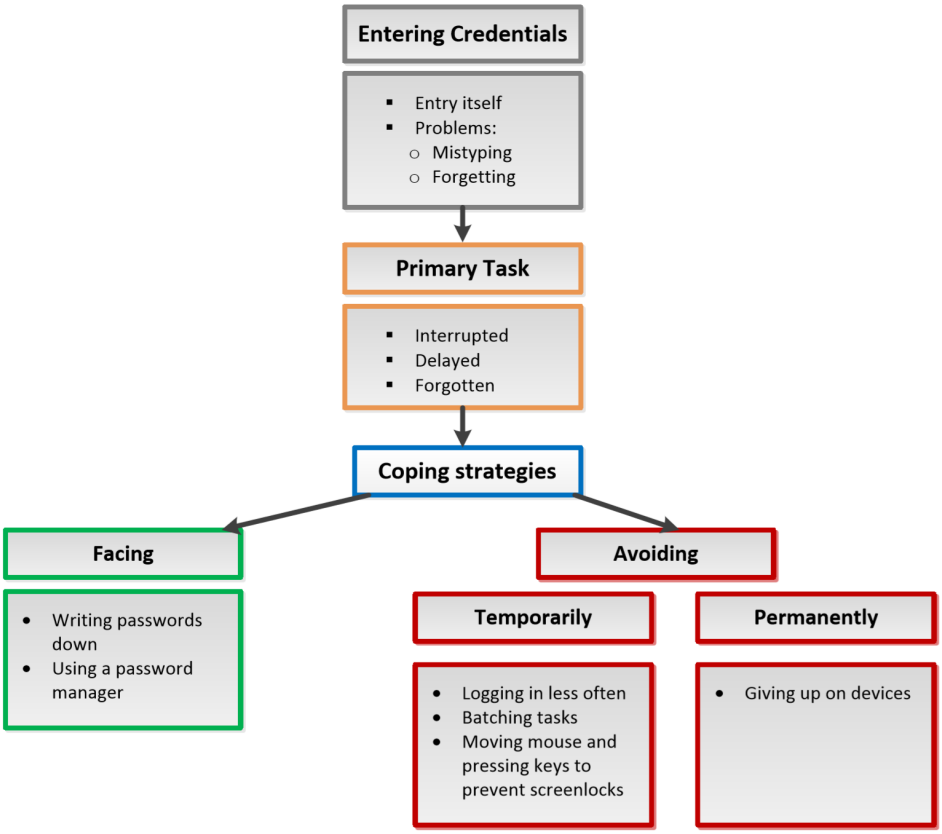

Figure 4.3: The emergence of coping strategies.

"centralizing it and creating the strongest password for centrally possible has been my tactic. Try to get it all in one bucket." (P3) Most had no awareness of the potential risks of doing so, others were aware but felt they had no choice – it was this, or constantly floundering when a password was needed. Another strategy was to employ a password manager (such as LastPass or KeyPass).

Proactive authentication. Participants also reported planning ahead and authenticating in advance if they expected a message or file they needed for their work, so they could access it without being delayed by authentication.

Avoidance: Deferring or removing authentication effort

Two types of avoidance behaviours were identified: employees either temporarily avoided authentication deferring it to when it was more convenient or they removed the need to authenticate more permanently.

Temporary avoidance

Preventing log-out. Participants reported reducing the frequency of authentication by preventing a logout. Often, they kept the screen active by moving their mouse.

Logging in less frequently. Participants also mentioned that they would turn around their email faster if it were not for the authentication burden. They reported that they particularly reduced the frequency with which they logged in remotely (e.g., from home), because this required additional security steps and the use of an RSA token.

Permanent avoidance

Giving up on devices. A number of participants reported that they coped with authentication by limiting the number of devices they authenticate to; P19 stated: *"If I had a laptop I would have to log in twice, once when you turn it on because the hard drive's encrypted, and then again to actually get into Windows or the operating system. [...] So I never wanted an agency laptop for that reason. I don't want to have to log in more times than I need to. That goes to the whole password security policy that we have here, is everything that leaves the agency has to be encrypted."* Those participants who refused to have an agency laptop were then incommunicado during business trips, meaning colleagues had to wait several days for responses required to complete their own work.

4.2.5 Discussion

One of the main lessons from this study is that not all logins are the same. With frequent logins, participants often reported having their passwords in their muscle memory. For more infrequent logins, they stated that having to recall or look up up their password was disruptive and they often developed coping strategies to reduce this disruption to their work. There were mentions of authentication affecting productivity as cumbersome login procedures prevented participants from logging in off-site and after hours. Participants also expressed frustration with logins they perceived as unnecessary or unjustifiably cumbersome for the level of security that they felt the organisation needed. High authentication effort led participants to develop coping strategies which helped them keep authentication at a manageable level and permit it at times when it was least disruptive. Participants shared their vision of a future workplace where authentication is implicit and thus well integrated into the work processes.

Perceptions of effort

The results of the study show a high level of frustration with authentication. The majority of participants estimated that they logged in more often than they actually

did. This might be an indication that they perceived authentication as a burden and a nuisance. Clifford Nass, from Stanford, an expert on multitasking and task switching commented on this finding saying: *"I suspect that unlike multitasking, in which the task switching is voluntary and hence people have an incentive to downplay the costs, authentication is not voluntary, thereby making people aware of the cognitive cost."* (Personal communication, October 2011). Other research confirms that task-switching has a 'ripple effect' on focus, concentration and efficacy – not only on the next task, but any task for the next 20 to 30 minutes (e.g., [171]).

The frustration over the disruptive effect of authentication could be attributed to the Zeigarnik effect [283] which states that humans are more likely to remember tasks that were interrupted and left unfinished. Research has shown that the lack of closure that stems from an interrupted and unfinished primary task can promote some continued task-related cognitive effort which might impact the secondary task of recalling a password [228].

Participants also thought that the amount of effort expected of them was unjustified for the level of risk they felt they were facing. They struggled with infrequently used passwords where similar problems have been documented in previous research. In a study conducted in the context of two organisations, Adams and Sasse [3] recommended that infrequently used passwords should be changed less frequently. Despite the study being over 16 years old, it appears little has changed in systems design to incorporate this suggestion.

Quantifying workload

Participants in the study developed a range of coping strategies to consolidate the effort associated with security, and minimise disruption of the primary tasks. To understand the workload they faced, the researcher conducted an analytical assessment of the workload certain authentication events would pose using the GOMS-KLM modelling technique. GOMS-KLM and CogTool were used to calculate how long authentication tasks would take (for examples, see Appendix A). Authentication with the username and policy-compliant password would take around 10 seconds. Using LastPass reduces this time to around 4 seconds. This number may not appear to be a huge time saving, but using LastPass posed significantly less of a disruption to the primary task – which means the cost associated with re-starting the primary task was significantly lower.

Participants said they preferred to keep a piece of paper with the password written on, even if it took several minutes to locate, rather than try to recall their password. This shows a fundamental truth from the field of HCI – users prefer extra physical to extra cognitive effort [174]. In the specific case of authentication, authentication is a hurdle dropped in the path of task completion. If the authentication task requires cognitive resources, it can interfere with the cognitive effort of the primary task. Several participants reported having authenticated, only to find they could not recall what they wanted to do – and needing significant time to backtrack in the task and resume work. A GOMS-KLM prediction shows the password manager minimises disruption caused by password recall and entry. A study by Trewin et al. [251] shows how participants were more likely to remember primary task-related information if the authentication task uses less cognitive resources.

The impact of authentication effort

Coping strategies. The study uncovered a range of coping strategies which can be divided into two categories: facing and avoiding authentication. These two main categories reflect the adaptive and maladaptive coping strategies in the protection motivation theory [208].

The transition when participants abandon an adaptive coping strategy and opt for a maladaptive one could also be framed in term of the compliance budget [20]. At a certain point, employee willingness to expend effort reaches a limit and they start avoiding tasks or technologies because of the associated authentication effort.

Productivity loss. Several participants said they had reduced the amount of additional work they did outside of regular office hours because of the authentication burden. Staff avoided logging in or did so less frequently from home or when travelling: *"Things get put off until when it's, 'OK, I have a block of time. It's worth it for me to get the token, to log in and to sit there and do like an hour's worth of work or half an hour or something like that." [...] But if it's for like fleeting little, 'Oh, I have this great idea' or 'I want to send this e-mail' or something, then I'm more likely to put it off until I have that sort of block of time where a log in is worth it. [...] especially if it's something that wasn't actually due. It's after hours. You've already put in your nine hours or however many hours you're doing and then you think of something at home, it definitely is less likely that you're going to get online to actually do that thing*

that you're thinking of. You're just going to wait until the next day." (P11) This can make colleagues and customers wait for information or assistance, which in turn is likely to hold up their work, and that great idea might have been forgotten by the next time the participant logged in.

Previous studies have similarly shown how authentication can affect productivity. Strouble et al. [245] who studied the Common Access Card (CaC) used at the US Department of Defense (DoD), found that employees stopped checking emails from home (due to the unavailability of card readers) and that many employees accidentally left their card in the reader. The authors estimated that the DoD lost more than $10M worth of time as a result of deploying the CAC. Similarly, a study by Inglesant and Sasse [122] found that staff re-organised their work to minimise authentication effort. The study presented here provides further examples of long-term consequences: some participants reported returning devices or avoiding logins when working off-site.

Embarrassment. Participants expressed worry that their organisation's security policy would cause them embarrassment, P23 emphasised it could even undermine the security of their accounts: *"everybody around me gets to watch me type my 12-character password in 15 times during the conference because it times out every 15 minutes. [. . .] somebody sitting next to me at the conference, by the end of the conference there's a good chance they can get in my computer easier than I can."* P17 described the fear of being embarrassed in front of customers and peers: *"I'd have 150 people waiting for a presentation. I'm waiting for 15 minutes for that sucker to boot up, so I can actually use it. Then, because I'm not admin and I can't change any setting on that box, when the screensaver kicks in, then I've got to log in all over again and reinitialize everything and start my presentation all over again and figure out where I was in the 120 slides that I use and make my way there. All the time the audience is right? That's a combination of security and other miscellaneous things. But, like I said, it gets in the way. You're trying to do something, which should be straightforward, but you can't."*

The fear of embarrassment caused by authentication has been mentioned in previous research. Beautement et al. [20] found that employees who were giving a presentation off-site took a copy of their presentation slides on an unencrypted USB device instead of an encrypted one because an encrypted one would require a password they might forget.

Limitations. The study was conducted in a US governmental organisation with employees who are knowledge workers and the findings might not be generalisable to organisation doing different types of work. Although authentication diaries were used to ground interviews in actual authentication events, the study relied on self-reports which are subject to a number of biases such as selective memory, exaggeration, demand characteristics etc.

4.3 Two-factor authentication in online banking

4.3.1 Background

This section summarises the 2FA landscape for online and mobile banking in the United Kingdom. All information provided is as of August 2014 when the research was conducted. The section contains an overview of authentication mechanisms for login purposes, setting up a new payee and using mobile banking functionality for 11 UK banks the participants in the study used: Barclays, Co-operative, First Direct, Halifax, HSBC Bank, Lloyds, Nationwide, NatWest, Royal Bank of Scotland (RBS), Santander and Ulster.

All 11 banks follow a two-step authentication process to log in to the online banking websites. In the first step, users provide an identifier and, only with Lloyds/Halifax, one authentication factor. Then, if the first step is successful, they are asked to authenticate by providing two authentication factors (one with Halifax and Lloyds). Table B.1 (Appendix B) shows the different credentials needed to authenticate to accounts for different UK banks.

During the identification step (i.e., entering username), banks usually provide users with a 'remember me' option: if selected, users do not have to re-enter their identifier when logging in from the same device and browser, unless they clear the cookies. In the second step, some banks display to the user a previously selected picture and/or text, aiming to provide some visual confirmation that the user is not on a phishing site.

Eight out of 11 banks rely on hardware tokens to generate the second authentication factor, giving a one-time password (OTPs). Some devices require the user to insert their debit card and enter the card's PIN (e.g., Barclay's PINsentry or NatWest's CardReader), while with other devices (HSBC and First Direct's Secure Key) users only

(a) Secure Key (HSBC) **(b)** Card Reader (NatWest) **(c)** PINsentry (Barclays) **(d)** Mobile PINsentry (Barclays)

Figure 4.4: Examples of two-factor authentication technologies.

OTP Source	Bank	Label
Secure Key	HSBC, First Direct	**SK**
Card reader	Barclays, Nationwide, NatWest, RBS, Ulster	**CR**
SMS	Santander	**SMS**
Mobile phone app	Barclays, HSBC, First Direct	**M**
Phone call	Lloyds, Halifax	**P**

Table 4.1: Technologies used to generate OTPs for different UK banks. The labels in the rightmost column will be used next to participant numbers to indicate which technologies they were using.

need to enter a previously selected PIN. Figure 4.4 (a, b, c) presents three examples of hardware tokens used by UK banks.

When setting up a new payment (e.g., to pay a bill or issue a money transfer to a new payee), all banks require an additional authentication step (besides those needed for login). Most banks require an OTP authenticating the specific transaction, amount, and/or payee (depending on the bank). Table 4.1 provides a summary of the different technologies used to generate OTPs across different banks.

All banks in the sample also support mobile banking, via smartphone apps (on iPhone and Android), allowing their customers to check their balance and issue payments. Table B.2 shows the authentication login enforced by each bank's application.

4.3.2 Study aims

The main aim of the study was to learn about the user experience of two-factor authentication (2FA) for UK online banking. This research examined it post-adoption to establish how users had learned to manage the increased effort caused by the use of 2FA.

4.3.3 Method

Design

The study consisted of three stages: preliminary open-ended interviews, authentication diaries and final follow-up interviews.

Participants

The research was advertised as a user study on online banking. Participants were invited to take the pre-screening questionnaire using flyposting in the local area, social media and the UCL Psychology Subject Pool[1]. The pre-screening asked for basic demographic information (age, gender, level of education) and whether or not they had previously used online banking and with which banks. Out of 66 individuals who responded, 21 participants were selected as described below.

Sampling and participant demographics. In order to maximise the diversity of the sample, potential participants were divided into five groups, one for each category of OTP generator, that is: Secure Key, card reader, OTP delivered via SMS, OTP generated by a smartphone app, and phone-based authentication. Within each category, participants were selected to achieve an even split between genders and a variance of ages.

The final sample consisted of 11 female and 10 male participants. Age of participants was: under 20 (1), 20s (9), 30s (6), 40s (4), 50 (0) and 60s (1). Mean age was 32.4 (range: 19–69, $SD = 10.87$). Out of the 21 participants, 10 had a postgraduate degree (a Masters or a PhD), seven an undergraduate degree, two had A-Levels, one had some undergraduate education with no completed degree, and one participant had completed vocational training. A total of five participants had an account with one bank, 14 with two banks, and two with three banks. Also, eight were customers of HSBC, six of Barclays, six of NatWest, five of Santander, four of Lloyds, three of Hal-

[1] https://uclpsychology.sona-systems.com/

ifax, and two of RBS. First Direct, Nationwide, the Co-operative Bank, Ulster Bank, and Citigroup all had one customer each. Finally, 43% of participants used Secure Key, 76% card readers, 24% OTP via SMS, 14% OTP via smartphone app, and 21% authentication through a phone call.

Materials

During the preliminary interviews, each participant was provided with an authentication diary which was a 20-page booklet where each page was designed to capture information for one authentication event. The design of the diary was informed by previous studies [122, 242]. A page from the diary is shown in Appendix B.1. On each page, participants were asked to write down the date and time for each authentication event, name of the bank, their location, the device from which they authenticated, what authentication factor(s) they used, and their reason to authenticate. They were also asked for how satisfied they were with the authentication process. They could also report on any problems they encountered (and if/how they were solved), they were asked to give a rating of their satisfaction with the authentication process and finally, there was space for any other comments.

Procedure

Preliminary interviews. In the first stage of the study, semi-structured interviews were conducted in a lab setting with one researcher and one participant at a time. Interviews lasted 30 minutes on average, and involved a total of 21 participants. The goal of these interviews was twofold: (i) eliciting a better understanding of banking preferences (e.g., online vs. offline), attitudes towards online banking, frequency of use of online and mobile banking services, and (ii) exploring attitudes towards authentication technologies deployed by the banks, aiming to identify sources of common errors, misconceptions, and frustration. Interviews were structured around a basic set of open-ended questions allowing users to talk about their authentication habits and experience. At the end, participants were briefed on the next stage – the authentication diary and received £5 for their participation in this part of the study.

Authentication diaries. Participants were asked to maintain a banking authentication diary for a period of approximately 11 days starting immediately after the first interview. Participation in the final interview, as well as remuneration, was independent of

whether or not participants completed the diary. Participants were asked to keep their ordinary Internet banking habits and not try to log in more or less than normally.

Final interviews. The last stage of the study involved 17 participants (80% of the initial sample), as four participants were unavailable to come for an interview within the set time frame. The second round of semi-structured interviews also lasted 30 minutes on average, again they were conducted in a lab setting with one researcher and one participant at a time. In the beginning of the debrief interviews, participants were asked about the experience of keeping an authentication diary, their authentication routine and some of their diary entries. Next, participants were asked to talk about other authentication systems they were familiar with, besides the context of online banking, placing more emphasis on systems that they enjoy interacting with, allowing participants to convey what elements they consider vital for a successful authentication process. Finally, the interview explored participants' perceptions regarding the notion of implicit authentication by explaining the concept and giving examples of how such a system would work. At the end of the session, participants received £10 for their participation in this final part of the study.

Ethics

The study design and protocol, along with all materials used, was reviewed and approved by UCL's Research Ethics Committee (approval number: 6010/001).

4.3.4 Diary results

The 17 participants who returned for the final interviews kept an authentication diary for an average of 11 days ($SD = 4.06$), starting from the day they had their first interview. A total of 90 entries were obtained from the diaries, with an average of 5.29 entries per person (range: 1–15, $SD = 3.9$). Out of 90 events, participants reported having had problems on 12 occasions (13.3%). The most frequent problems were mistyped credentials with five occurrences and misplaced tokens with two. The following problems were mentioned once: wrong memorable answer entered, wrong sequence of steps when using the token, a forgotten username, a frozen browser session and a slow Internet connection.

OTP generation

Participants could indicate their satisfaction with each login event on a five-point Likert scale (1 – very dissatisfied, 5 – very satisfied). From the 90 recorded events, 11 were missing a satisfaction rating (two participants provided no rating for one of their events and one participant did not provide any ratings at all). Satisfaction ratings for authentication events were compared for events with and without OTP generation (30 and 60 events respectively). Participant satisfaction was significantly higher when they did not generate an OTP to authenticate ($U = 1119$, $p < 0.001$, two-tailed Mann-Whitney test).

Number of credentials required

On average, each authentication instance required 2.44 pieces of information, ranging from 1 to 5 (e.g., 5 pieces of information were: surname, account number, last 4 digits of card, OTP, PIN for token). This quantity and satisfaction were negatively correlated ($\rho = -0.48, p < 0.001$): the more pieces of information were needed, the lower participant satisfaction was.

Time of day

For analysis, the events were divided in four times of the day: morning (6am-12pm), afternoon (12pm-6pm), evening (6pm-12am) and night (12am-6am). Out of 90 events, 24 took place in the morning, 32 in the afternoon, 30 in the evening, and 4 during the night. Morning events had a higher satisfaction score than events at other times of the day (4.51 vs. 4.07), this difference is statistically significant ($U = 978$, $p < 0.01$, two-tailed Mann-Whitney test). Participants experienced 8 problems (out of 30 authentication events occurring in the evening) versus 4 problems (out of 60 events during the rest of the day). This difference is statistically significant ($p = 0.0175$, two-tailed Fisher's exact test).

Type of hardware token

Satisfaction levels for banks using two different types of hardware tokens were also compared: events with banks using a card reader had significantly higher satisfaction than events with banks using Secure Key ($U = 203$, $p = 0.001$, two-tailed Mann-Whitney test).

4.3.5 Interview results

This section presents the results of a thematic analysis conducted on both preliminary and final interviews.

Struggling with hardware tokens

The study protocol for the study was such that the researcher avoided priming participants about authentication, this was reflected both in the recruitment process and also in the interview structure and the order of questions. The first questions participants were asked were around banking preferences and the user experience of online banking. Nevertheless, participants mentioned authentication and 2FA tokens long before there were asked any explicit questions about them. This can be seen an indicator of how strongly authentication impacts the user experience of online banking.

When talking about their experience with hardware tokens, participants discussed both what they perceived as advantages and disadvantages (e.g., compared to other 2FA forms), with negative feedback being predominant. Only four out of 20 participants found that the hardware token was "easy to use", "portable" or "easy to incorporate into everyday life." Seven participants complained that they had to remember to bring it with them, while four found the use of a hardware token to be frustrating. Some found it superfluous (2 mentions), unconvincing (2) and not thought-through (2).

Effort seen as unjustified. Five participants emphasised the effort required to use the token and questioned the necessity for it. P19 (SK, CR)[2] summarised their attitude towards the token as follows: *"I don't think it was really thought-through. I kind of wonder who they consulted in the beginning. [...] It's impractical to be constantly having that on you because if you lost your bag/keys/purse it would be very difficult to then log on or ask for another one. [...] It feels a bit long-winded at times, especially if I am in a rush or I forgot that I had to pay someone. So I find myself in the situation where I often have to order new card readers because I always tend to lose them and then find them. At the moment, I am trying to keep one in my bag and one at home. I've probably have about 10 or something because I keep losing them and finding them."*

Other participants felt that their account was secure enough before the introduction of the hardware token, as they had never experienced any fraud and/or felt that

[2]With each participant quote comes an indication in brackets what type of OTP generation technology they use, as per Table 4.1.

their password was hard to guess anyway. P19 (SK, CR) found the online banking authentication procedures to be excessive compared to offline transactions where only a four-digit PIN is needed to authorise a card payment and even no PIN is required for contactless payments under a certain threshold. Participants often disliked the extra step that the use of the hardware token introduced to their authentication task and found it irritating (3) or inconvenient (5). This seemed to be tied in with the fact that only one out of these eight participants reported using the hardware token to issue bank transfers, which probably contributed to the feeling of superfluousness.

Five participants stressed they thought a particular bank was more secure because it required using hardware tokens more often (e.g., not only for payment setup but also for login). Four participants (including two from the previous category) said they preferred a bank that used the token less, or did not use it at all. In fact, three participants reported changing banks or logging into their account less frequently because of the hardware token. P19 (SK, CR) said: *"Initially, I had an account with HSBC [. . .] and that was before they brought in the horrible key ring thing that you have to carry with you at all times to log in [. . .] With NatWest you have to remember passwords and stuff but you don't have to bring anything physically with you unless you're changing like payments and transfers [. . .] That's why I'm with NatWest, just because I find their online banking system [. . .] compared to HSBC, it's the most simple."*

P19 (SK, CR) was also disillusioned as to the reasons why the token was introduced in the first place: *"I've been using Internet 15 years now, and I've used it for a long time. I think people older than me should be using it as well and kind of be more afraid of it. . . and been told by their banks online banking is more convenient but then found the whole process not really secure. So I think it was about trust issues. It was the idea of having something more tangible similar to the machines in branch or a cash machine."*

The introduction of hardware tokens also affected participants' usual ways of working and brought an embarrassment factor. P11 (CR, P) reported that they stopped checking their balance and recent transaction while at work, since the login procedure makes it obvious to co-workers they are not working: *"It's OK when I am at home, but when you are at work and you are pretending you are actually doing work when you are actually checking on your account, then you have to bring out this calculator thing*

and it's kind of obvious you are not doing work. I'd rather have something where I am just on the screen and it's lot quicker."

Struggling with credentials

Authentication terminology. Another important issue related to confusing terminology used inconsistently across different banks. For instance, P8 (CR, SMS, M) reported: *"Is it a passphrase or passcode or key phrase what they need? [chuckling] I think it is slightly confusing. Although I'm experienced [...], it's frustrating."*

In fact, there is an astonishing number of different terms used to denote a few different forms of credentials: membership number, customer number, username, user ID; password, passcode, passphrase; memorable word, memorable answer, memorable information etc. (see Table B.1). As a result, participants complained they not only had to remember the actual credentials but also what they were called. P18 (SK, CR, SMS) explained: *"One of them was called a 'passcode' and the other one was 'ID', and I never knew which one is which [...] For the cases I forgot them, it was very confusing for me because when I was to recover my credentials it was difficult to know which I forgot!"*

Providing selected characters. Participants reported a few coping strategies when asked to enter selected characters from their credentials (e.g., the third and the sixth character), including repeating the password (or memorable word) in their head, whispering it, counting letters on their fingers, or writing it down and counting letters on paper.

Several banks even ask for the characters to be entered out of order, which makes the process significantly harder. For P17 (CR, M, P) providing characters out of order meant more effort and time: *"If I don't even bother to remember the password, I will first have to look for it, then do the math... I mean it's not that difficult but "Oh! This one is the fifth letter", I repeat it multiple times in my head."* Also, the methods of entering the characters seemed to matter to the participants. P11 (CR, P) stated: *"Actually when you are accessing it from a PC it is a box, when you access it from a mobile phone it's a drop down list. I prefer the box, it's just easier. I don't have to scroll up and down and try to find it."*

The researchers examined more closely the use of drop-down lists, such as those used by Barclays (both on the desktop and mobile devices) to let users enter selected

characters of their "memorable word." It could be assumed that their purpose was to reduce the risks of key-loggers installed on the user's device. However, with the exception of some versions of Internet Explorer, one can still type the corresponding characters from the keyboard (i.e., upon pressing a key, the correct character from the drop-down list is selected), thus users are actually still vulnerable. Also the characters are not hidden once they are selected from the menu, which makes them more vulnerable to shoulder-surfing.

Memorable answer. Beside usernames and passwords, customers of UK banks are asked to provide a "memorable answer" (e.g., to a question like "What was the name of your first pet?"). In general, participants did not report problems providing their memorable answers but three of them said they were often unsure how they had spelled them. P15 (SK, P) also complained that one of their banks was asking for their memorable answer but did not say what the question was: *"HSBC gives you prompts for the memorable information. The Lloyds one just asks you to enter the memorable information but it doesn't give you prompts about what it was."*

Three participants were concerned with the fact that their answers were easily guessable by their relatives and friends. They questioned how memorable answers actually added security to their account. P17 (CR, M, P) reported: *"This information is everywhere [. . .] It's just a waste of authentication, and a waste of time."*

Assigned usernames. The majority of UK banks assign usernames (that cannot be changed) – which often makes it harder for users to remember them. Participants in the study reported two main coping strategies: eight of them said they were writing their usernames down, while seven had their browsers remember them. Five participants reported that not being able to remember their usernames prevented them from logging in from other computers than their own. P2 (SK, CR) stated: *"For NatWest and HSBC, if I go somewhere else and I use another computer [. . .] then I cannot remember it and I cannot even get through the first step."*

Participants also stressed they found their username to be an unrelatable string of numbers, unlike their usernames for other services which they were able to pick themselves.

Passwords. When asked to create a password, three participants stated they re-used an existing password and nine that they partially re-used it. P7 (CR, P) said: *"I use*

one password for everything, and I vary it a bit I add a number an exclamation mark, a hashtag, but the core password is the same for all websites." Seven participants said they created participants to be highly memorable (to facilitate their recall) and other seven reported writing their passwords down. P21 (SK, SMS) struggled with creating a valid password for their bank since the instructions (describing length and type of characters) did not match what the system actually accepted, driving P21 to use a trial-and-error strategy in order to generate an acceptable password.

Credential recovery. Five participants reported that they had to reset their credentials at some point, describing the process as cumbersome. P12 (CR) reported: *"You have a number of times you can repeat the process before they lock you out. And then you have to phone them up and go through the whole thing and get them to reset it and it takes 24 hours. So you need to be very very careful [...] And I always get a bit nervous – the last thing you want to do is be locked out from your account. It's really frustrating as well because you could mistype m instead of n."*

Participants that went through a reset process reported they started to choose easier-to-remember credentials and to focus more during logins to prevent mistakes. P11 (CR, P) said they "learned" from their experience: *"It's quite a big deal afterwards. To reset the whole thing, you have to call them and they have to verify things... I chose easier information so that I am sure I will remember it and not have to go through the process again."*

Mental models of authentication tokens

During both interviews, three participants were particularly intrigued by their hardware token and they made speculations on how it was operating.

Where do the codes come from? Three participants were curious about how a card reader generated codes that were accepted by the website and wondered if it was connected to the Internet. A participant thought that the bank and the token have a list of codes and as soon as the customer uses one of them, it gets "crossed out" from the list. This belief could have been influenced by the fact that banking authentication in some European countries used to rely on this system. In fact, P10 (CR, M, P) mentioned having used this system back when they lived in Germany: *"I used to live in Germany, I really liked the way they do it there, my bank there was Deutsche Bank. It gives slips of paper where you have OTPs which are easy to replicate... For me, it's convenient*

to have both in the office and at home. It feels safe even if someone sees it because even if they see the code, they wouldn't know which one they would be using."

Do I need to protect the token? Two participants emphasised the token was valuable and some stated they were protective of it hiding it out of plain sight. P11 (CR, P) was worried about what kind of information related to their account could be collected from the card reader if stolen: *"With regards to this device, I mean, it's meant to be secure I would imagine, because you are supposed to be the only person that should have possession of it. But if you lose it what happens? What kind of information can be taken off it? ... You know relating to your account, I am not quite sure how secure that is to be honest."*

In some of the interviews, it became apparent that participants did not realise the card readers were interchangeable between banks. When asked if it would be possible to use a card reader from one bank to log in to another (which is the case for several banks), P7 (CR, P) expressed doubts: *"Well, Nationwide sent it to me so I assume it's unique for Nationwide, I don't think I can use it for other banks. [...] I think the number is unique for the debit card and the PIN, so if someone stole the reader they can't access the reader, they need to slot in the credit card. It's to make sure no one else has access to my account. To get the number, it has to be my debit card and they have to know my PIN – so it is very unlikely... unless they have the credit card, the PIN and the reader."*

The study also uncovered some misconceptions regarding authentication, as P12 (CR) commented on how authentication in the morning is faster than in the evening: *"I think the process itself can be quite slow. I think this is because the server is quite busy and because they don't have enough staff to check this kind of stuff and that can be a problem. Especially in the evenings it seems to take longer. One day I was at home, I wanted to pay council tax 10 in the morning and it went through so fast. It was incredible. Whereas in the evening there's more people on and maybe they don't have a lot of staff at the other end checking this stuff and so this takes longer and if you are in a rush this can be frustrating."* After a couple of clarifying questions from the interviewer, it became apparent that the participant thought verification of credentials happens manually with members of staff checking the usernames, passwords, memorable answers and OTPs customers enter online are correct.

Rituals that make users feel secure

Throughout the interviews, participants reported on their strategies for staying secure while banking. Although one might argue that some of them do not necessarily add real security, participants felt more secure by employing them, thus, they can be called 'security rituals'. P7 (CR, P) elaborated what made them secure: *"There is always a small chance, of a virus or people can steal your bank account details. That's always a small concern and I do as much I can to prevent that. On Windows, you can have a firewall, I have Avast Antivirus, they do regular virus definition checks. I have a password, it's not an easy password and it's also memorable information. I change them every six months and if I log in from a friend's computer or a library computer, I don't ask the computer to remember my username. I don't do that."*

Motivation for online banking

During the first interviews, participants were asked to discuss their preferred ways of banking. Out of the 21 participants, 17 reported they preferred online banking, with 11 participants appreciating its convenience, six its ease of use, and seven the fact that online banking was less time-consuming than other forms of banking. Obviously, this strong preference might result from the fact that the recruitment materials focused on online banking. Participants also highlighted several reasons why they prefer not to go to bank branches, including long queues and/or distance to the nearest branch, whereas, four of them reported preferring in-person banking due to faster/better resolution of queries and having face-to-face contact with members of staff.

Ideal authentication

As part of the second round of interviews, participants were asked what an ideal authentication procedure for online banking would look like. They provided the following suggestions.

Biometrics. Authentication using biometrics was the most frequently proposed idea. Some participants mentioned it already in the first interview, before the researcher even asked this question. For example, P6 (CR, SMS, P), after discussing their struggle with username, password, and token based authentication, said: *"If you could be in an ideal world where you wouldn't have to use your card reader, type in your password... you know I think, in a few 100 years from now you'll just put your finger on a machine*

and it reads your fingerprint. Today, it's slow – you know fast is good! The faster the better." P17 (CR, M, P) also said they would like a camera-based biometric system, as long as the process would be fast: *"Maybe through the camera, webcam looks at me and verifies it's me entering the five-digit code[. . .] I would be willing to change my position and stay still until the scan finishes – I still don't need to remember anything. [. . .] I think 5 seconds or 10 maximum, not beyond 10 seconds – I'd freak out – I'd prefer to remember."*

Reducing cognitive effort. Six participants emphasised they would like an authentication system that minimises the need to remember anything. Seven participants also emphasised they needed to concentrate to log in. They stressed that especially when tired or in a rush, they needed to stop what they were doing and focus. P14 (SMS) stressed they needed to make a conscious attention switch: *"If I am in a rush, I maybe misspell my surname or I do not enter the card number correctly. If there's any delays, it usually is one of those two. The device itself, I don't think I ever had problems with that stage, it's always been the first one. I'll have to get myself together mentally and let's say 'Focus! Whatever is in your mind, forget it.' "*

Reducing physical effort. Participants expressed their desire for an authentication mechanism that would reduce the physical effort they need to make – by avoiding the need to enter multiple pieces of information (5 mentions), or to carry additional devices with them (4).

Fast and simple. Reducing the cognitive and physical burden would lead to faster and simpler authentication procedures. In general, participants preferred fewer steps, as some of them suggested they should be logged in based on their actions, passively, rather than actively needing to enter their credentials. For instance, P6 (CR, SMS, P), while suggesting the use of biometrics, said: *"We have our unique fingerprint no one can replicate at the moment, and you just put your finger on the screen and that's it, a one-step process."* Interestingly, three participants said they preferred one bank to another because the login process was faster.

Portable. Two participants also highlighted the need for "authentication portability". They thought logins should be for more than just one system – for example, when using their fingerprint to access their computer, this should already log them in to their bank and also to other accounts.

Implicit authentication

At the end of the second round of interviews, participants were also asked about their views on implicit authentication since this suggestion was made in the previous study. The experimenter explained the basic idea of how an implicit authentication mechanism works, that it authenticates users by constantly monitoring some aspects of their online activity and their behaviour (e.g., their typing rhythm). Participants voiced two kinds of reservations about this type of authentication. First, five participants said the system would need to be highly accurate as they did not want to find themselves locked out of their machines. They raised concerns if the system would authenticate them even if they were tired or did not follow their usual patterns. They said they would feel more in control entering a password since if an implicit authentication system decided they were not a legitimate user, there would be nothing they would be able to do. Second, six participants voiced privacy concerns saying they might find it intrusive and could feel uncomfortable knowing their behaviour was being tracked. They also stressed they would want to know who was storing their data and how it was being used. P10 (CR, M, P) stressed: *"I could see implicit working but you'll probably run to privacy issues about that: Who's doing the software? How's the monitoring done? Who gets the information from the monitoring? blah blah blah. That would be the real issue."*

4.3.6 Discussion

This section presented an in-depth user study of the usability of two-factor authentication (2FA) in the context of online banking. It consisted of two series of semi-structured interviews, a period of keeping an authentication diary providing both qualitatively and quantitatively data.

Participants reported using 2FA without too many mistakes or lockouts, which is not surprising as the participants were banking customers accustomed to 2FA and online banking. However, the study uncovered a number of issues with credentials in general and, in particular, with hardware tokens. The study showed that the effort expected of banking customers is high with some participants considering it unjustified. The demands placed on users, such as the need to produce and remember a wide range of different credentials (often with confusing nomenclatures), or having to carry around and operate extra devices, have affected participants' banking routines.

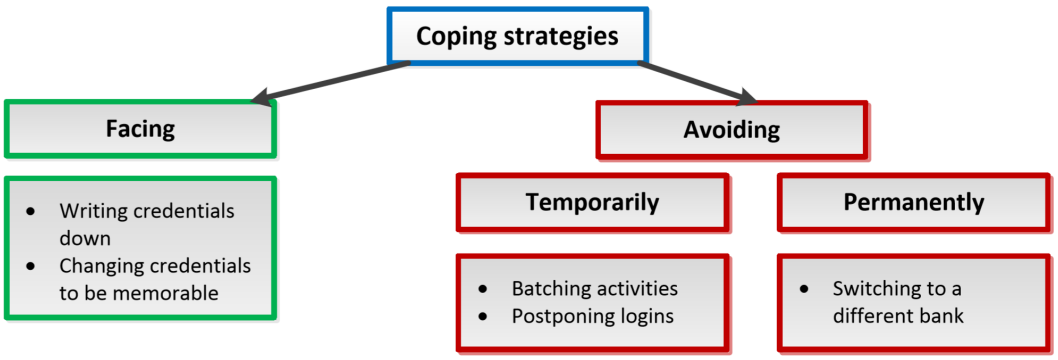

Figure 4.5: Examples of coping strategies adopted by the participants categorised using the taxonomy developed in Section 4.2.4.

They have changed how often, from where, and how, they access their accounts. The findings from the study reflect the results obtained by Gunson et al. [102] who found that two-factor authentication was perceived as less usable than single-factor authentication. Like in the study described in this chapter, Gunson et al.'s participants were confused by the terminology used for credentials. They were asked to enter a *secret number* and an *access code* after one another and both credentials were six digits long. In order to avoid confusion, the authors recommended to avoid credentials with such similarity in real-life deployments of two-factor authentication systems.

Interviews, as well as the analysis of authentication diaries, showed that participants' satisfaction dropped when extra effort was required, that is when they had to provide multiple pieces of information and use hardware tokens. Again, habit mattered and several users were not happy with these as they could not automate the authentication process and their primary task was disrupted by having to stop and shift their focus onto authentication. One participant actually switched banks to avoid having to use hardware tokens, others used accounts requiring effortful login procedures less often.

Throughout the interviews, participants reported on a number of coping strategies they were using to manage their authentication effort. Figure 4.5 shows a categorisation of coping strategies participants reported employing based on the taxonomy established in Section 4.2.4.

Several participants proposed biometric-based authentication as an ideal solution that would require less effort (nothing to carry or to remember). Implicit authentication was also a well-received idea, but participants stressed it would need to be highly

accurate and respectful of their privacy. Participants asserted, on several occasions, they knew what kind of level of security was required for them and took measures to obtain it. If participants found that their memorable answer was easy to guess, they used a password instead. They also found authentication procedures to be excessive compared to offline transactions where only a four-digit PIN is needed to authorise a card payment or nothing at all for contactless payments under a certain value. Anti-viruses and firewalls gave them a sense of "doing the right thing." Some also reported being careful with keeping the hardware token secure, even if it could not be operated without their debit card (as with Barclays' PINsentry).

Prompted by the findings from the US governmental organisation, participants were asked what they thought about logging in implicitly using behavioural biometrics. While participants very much liked the idea of biometrics and frequently suggested it to us as the ultimate solution where all authentication was headed, they raised privacy and reliability concerns with regard to behavioural biometrics. These participants worried what this data would reveal about them and who would be able to access it.

The diary data showed that the more effort was required, the lower participant satisfaction was. Also time of day mattered which could be used to optimise what is asked of participants and when, in a contextual way.

Recommendations

Reducing the number of steps required for online banking authentication was crucial for the participants. For most of them, an ideal authentication process would involve fewer steps and would not require them to carry tokens. This highlights the need to reduce the number of steps throughout the authentication process, and to minimise the use of the hardware token. As participants have different capabilities, needs and preferences, banks should consider providing customers with a choice of which 2FA method they can use. In the study, some participants exercised this choice themselves and moved their banking to a different account to be able to use their preferred authentication method.

Since this study, banks have been increasingly encouraging their customers to switch to smartphone banking applications. For the users, this reduces the burden of carrying a dedicated device while for the banks, it removes the need to provide OTP generating devices. The question arises if a smartphone app offers an equivalent

level of security to that of a dedicated token. For example, security concerns were raised after it was announced that RBS and NatWest would allow their customers to access their banking apps using a fingerprint scan (Apple's Touch ID) instead of a PIN [196, 157].

In a short-term perspective, the study also yields a few immediately actionable suggestions. In order to minimise confusion, UK banks should unify the wording for credentials (which now includes: password, passcode, passphrase, memorable word, memorable information). The use of different authentication terminology by each bank was a particular burden for participants who had accounts with more than one bank (16/21 in the sample). This highlights the need for banks not to consider authentication in isolation, but the actual tasks and contexts of use that their customers face, and how design of the mechanism impacts the lived experience of their customers. It could be argued that banks should remove features that do not add any real security but actually negatively affect the user experience by increasing disruption or time needed to authenticate. For instance, some banks, aiming to thwart key-logging attacks, force users to enter selected password characters by choosing from drop-down menus. However, this is cumbersome for many users and not really effective as one can still type the corresponding characters from the keyboard. Also, the characters are not hidden once they are selected from the menu, which makes them vulnerable to shoulder-surfing.

Limitations and future directions

Firstly, the participant sample in the study was small and consisted of well-educated, highly computer literate and relatively young individuals. One might argue that this is representative of those who currently adopt online banking but given that this group of participants struggled, it is likely that older and less technology-savvy user groups are discouraged from online banking by the lack of usability which is an accessibility problem. For banks, it is a missed commercial opportunity as they are looking to close down branches and lower transaction costs. Secondly, the study mostly relied on self-reported data, which might be subject to exaggeration, selective memory and the social desirability bias. The findings from the study can be tested in a future, larger online study. Objective empirical data about success and failure, password resets, and customers leaving can only be obtained from the banks themselves who are reluctant to share such information based on grounds of commercial sensitivity.

4.4 Conclusions

The two studies described in this chapter examined the role of effort in user authentication behaviours in different context – in an organisation and for personal online banking. Despite the different contexts, both participant groups raised similar issues. They struggled with hardware tokens finding them inconvenient and they thought the level of security was too high for the level of risk they were facing.

In both studies, participants emphasised the disruptive effect of authentication which led them to develop coping strategies to reduce the effort required to an acceptable level at a given point in time. Participants demonstrated a desire to control their effort moving it to when it was convenient and least affected their productivity. The type of coping strategies show what kind of effort is acceptable and when. For example, physical effort was preferred when participants were immersed in intellectual work. Often, they did not access the accounts that authentication guarded access to when their primary task required it but when they could handle the authentication effort. This had an effect on their productivity and satisfaction as numerous examples showed.

Both participant groups appeared to have some tolerance towards the increased effort that authentication required of them. In an organisation, they felt the obligation to keep the organisation secure and they might have considered it to be part of their contract. For online banking, 2FA is mandatory for some banks and participants rationalised the effort by saying they did not have to physically go to a branch and could do banking from their computer. Having said that, there was a small proportion of participants who gave up accounts or devices if they felt the effort required was unreasonable, especially if they considered the effort to be too high for the level of security they felt they needed. In both cases, there is evidence that systems were not designed with the users in mind and users are asked to expend effort to compensate for the shortcomings of the systems and procedures which fail to support them in their tasks.

While the studies in this chapter investigated how participants cope with authentication effort as part of their every-day lives, the focus of the next chapter is on two mechanisms that have been proposed but not yet introduced. The studies are laboratory-based user evaluations of the effort required to use them and their acceptability in different contexts.

Chapter 5

Authentication and verification mechanisms

5.1 Overview

This chapter focuses on different authentication and verification mechanisms to examine the actual time taken, the perceived effort and participants' acceptance of them. In the first study, participants performed logins through the Vernitski Authentication Grid (VAG). It is a 6×6 grid reshuffled on each login making the entry more difficult but potentially allowing the user to have a simple memorable password. In the second study, participants verified using reCAPTCHA, PlayThru and NoBot as part of a ticket-purchasing task. NoBot holds the promise of being effortless since the user does not need to decipher characters or drag and drop elements as is the case for reCAPTCHA and PlayThru. Both studies consisted of two parts – participants first tried the mechanisms and were then asked to reflect on their experience. Both studies show that participants considered security and privacy very contextually, elaborating on in what situations using the given mechanism would be appropriate. In both studies, there is a comparison of the usability of the mechanisms on a physical and on a virtual keyboard showing that authentication using the VAG was faster on a virtual keyboard but verification using NoBot took longer on a tablet. The results show that participants are willing to use a mechanism that requires effort if they believe it would make them more secure, especially for high-value and infrequent logins and transactions.

5.2 Vernitski Authentication Grid

5.2.1 Background

The Vernitski Authentication Grid (VAG) was named after its inventor the mathematician Alexei Vernitski. Users choose a password consisting of an even number of letters and/or digits. The entry of this password is through a 6×6 grid (Fig. 5.1) with the position of characters being random and reshuffled on each authentication. The user needs to enter the characters of their password in pairs. They need to find the row that contains the first character and the column that contains the second character and enter the letter/digit that is in their intersection. Then this is repeated for each next pair of characters until the end of the password.

The VAG is not a graphical password since there is no graphical element, such as a picture or a pattern, that the user would need to remember. Instead, they are asked to remember a password and then enter it using a grid. The mechanism is a form of challenge-response (C-R) authentication. The aim of the grid is to guard against shoulder-surfing and key-logging. It can protect the password at three stages: (1) at entry in a case when the interface cannot be trusted, for example, when using an unknown computer, (2) at entry when the physical environment is hostile, for example, in the presence of a potential shoulder-surfer and (3) in transit when communication can be eavesdropped on. Capture of the secret password shared between user and system is made by the characters in the grid being reshuffled each time. If the attacker has a key-logger installed on the users' computer, all they receive will be a set of random-looking characters. A statistical security analysis of the VAG was conducted by Papanicolaou [180].

Another potential use case for the grid would be as back-up or very infrequent authentication – such as annual tax returns. Long periods of not recalling a password or backup credential result in high failure rates. To increase memorability of the shared secret, users could choose a password with high personal entropy [71] – a word or phrase that is hard to guess, but meaningful to the user, and strongly embedded in biographical memory (using episodic memory).

In terms of effort, using the grid with a simple highly memorable password could potentially be easier than memorising and recalling a complex password. The ability to use a grid is stored in procedural memory which is a type of implicit memory and

after repeated use, it could become habitual and unconscious. In contrast, a password would be stored in semantic memory and it requires uncued recall.

5.2.2 Study aims and hypotheses

The goal of the study was to assess the usability and user acceptance of the VAG. As users struggle to remember passwords, for the grid they could have a simple password but the entry would be through a 6×6 grid with characters and digits reshuffled on each login. The purpose of the study was to evaluate the user experience of the VAG both quantitatively and qualitatively. The study was meant to be a preliminary evaluation, assessing the effort required to use the grid, the learning curve and participants' perceptions of the grid. This often led to a more generally discussion of participants' experiences of authentication, their expectations and preferences.

The following hypotheses were devised to answer the research question:

H1: There will be a difference in the time that users take to enter a password on a PC using a physical keyboard and a tablet using a virtual keyboard.

H2: The time of entry will decrease with practice.

H3: Authentication speed will depend on users' individual characteristics such as age, computer literacy and experience with touchscreens.

5.2.3 Method

Design

The study was conducted in a laboratory with one experimenter and one participant at a time. There were two groups of participants: first one with 31 and the second one with 5 participants. The first stage was the same for both. After being introduced to the scheme, participants were asked to perform six logins. While the first group authenticated six times during one lab session, the second group was asked to return for another session a week later where they were asked to perform another six logins.

Participants

Participants were recruited through the UCL Psychology Subject Pool. There was no pre-screening for the study and there were no restrictions on who could participate, apart from the standard requirement that they had to be at least 18 years of age. Participants were paid £6 for their participation which took about 30 minutes. Overall, there were 36 participants in the study, 19 male and 17 female. Mean age was 26.9

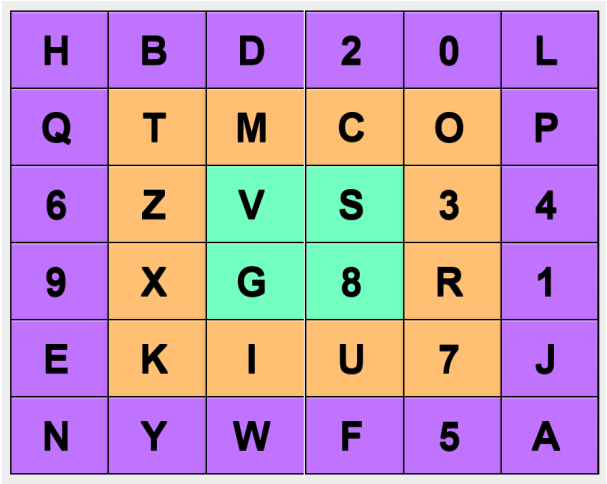

Figure 5.1: A screenshot of the implementation of the VAG used in the study.

years (range: 20–49, $SD = 7.2$). In terms of education, 28 participants had completed a university degree and seven had A-Levels.

Apparatus

The implementation of the idea was done by Constantinos Papanicolaou as part of his MSc project in the academic year 2012-13. The prototype of the VAG used in the study (Fig. 5.1) was programmed in Java (using JSP/Servlet technology) and linked to a MySQL database. The JQuery framework was also used to enhance the interactivity of the prototype and provide a better user experience. The study was performed on a PC and a tablet. The PC was running Windows XP and had a $22''$ monitor with a resolution of 1920×1080 pixels. Participants used a standard QWERTY keyboard with a UK layout. The tablet was a $9.7''$ iPad 2 with a resolution of 1024×768 with an on-screen virtual keyboard with a UK layout.

Procedure

Upon arrival, the participant received an explanation of what the study involved and the experimenter answered any questions they had. They were asked to read an information sheet and sign a consent form. The workings of the grid were then explained to them by the experimenter who used a laminated sheet of paper with a grid on it and a marker pen. The participant was then asked to enrol by setting up a username and a password. Once they completed the enrolment, they were asked to perform six logins: three on a PC and three on a tablet (order counterbalanced). After having performed these logins, the participant was interviewed about their experience and then asked

to complete a brief questionnaire asking for demographic information, their computer literacy and cyberthreat exposure.

Ethics

The research was approved by the Head of Department in the Department of Computer Science and exempted from a full review by UCL's Research Ethics Committee on the grounds of minimal risk.

5.2.4 Quantitative results

Overall, across all participants, devices and trials, a login attempt (regardless of if successful or not) took 63.7 s. However, in real life failed attempts to authenticate add to the time needed to access a system, therefore the following analyses consider the cumulative time needed for a successful login, which encompasses the times of failed attempts too. The cumulative average time for a login using the VAG across all participants, devices and trials was therefore 88.6 s (median=32.5). This average is skewed by some trials requiring several attempts. Therefore, to illustrate it better, 45% of logins took under half a minute, 70% under 1 minute, 79% under 1.5 minutes and 85% under 2 minutes. Figure 5.2 shows login times across different trials and treatments. Each participant had to successfully log in six times and the number of attempts needed ranged from the required 6 to 14 ($M = 8.16$, $SD = 2.11$). On average, 1.4 attempts were needed for a successful login ($SD = 0.8$). Passwords chosen by participants were on average 6.7 characters long ($SD = 1.24$, range: 4–10). Out of 36 participants, 22 chose a 6-digit password, presumably influenced by the example password (*zebra1*) given by the experimenter which had 6 characters.

A 2 (Device: PC, Tablet) $\times$ 6 (Trials) $\times$ 2 (Order: PC first, Tablet first) repeated measures ANOVA was conducted on the time needed to log in. There were significant main effects of Device ($F(1,34) = 12.6$, $p = 0.001$), Trial ($F(2,33) = 9.74$, $p = 0.001$) and Order ($F(1,34) = 9.8$, $p = 0.004$). There was a significant Device $\times$ Order interaction ($F(1,34) = 5.24$, $p = 0.028$) and Device $\times$ Trial interaction ($F(2,33) = 3.35$, $p = 0.041$). There Trial $\times$ Order interaction was not significant ($F(1,34) = 2.49$, $p = 0.091$). The three-way Device $\times$ Trial $\times$ Order interaction was non-significant ($F < 1$).

H1. Device type

It was hypothesised that there would be a difference between how long participants would take on a PC using a physical keyboard and on a tablet using a virtual keyboard. On average, participants needed 118.8 s to log in on a PC and 58.4 s on a tablet (Fig. 5.2). The ANOVA test described earlier showed a significant main effect of Device indicating that participants were faster to log in on a tablet than on a PC. Post-hoc effects indicated this difference is statistically significant ($U = 319.5$, $p = 0.02$). H1 is therefore supported.

H2. Time

It was hypothesised that login times would decrease with practice. The login times reflected a learning curve as participants were authenticating faster with each trial. They took longer to log in at the first trial but became faster by the third trial. Upon switching to the other device, the login time was longer at first trial but decreased again with practice (Fig. 5.2). Participants who started on a PC were slower in their first trial than those who started on a tablet ($U = 39$, $p = 0.008$). After switching to the other device, participants who used a tablet first and switched to a PC were marginally faster than those who switched from a PC to a tablet but this difference was not statistically significant. The group that started with a tablet was on average authenticating faster than the group that started on a PC ($U = 3931$, $p < 0.001$). Assignment to these groups was random and there were no significant differences between the two groups.

For the smaller sample of five participants who authenticated on two occasions, it was hypothesised that they would be faster in their second authentication session. For the first session, the average authentication time was 54.4 s and for the second 39.9 s. Each participant authenticated faster in the second session by an average of 14.6 s. Despite this trend, there was no statistically significant difference there ($p = 0.4$). A larger sample of participants would be needed to be able to show a possible effect.

H3. Individual characteristics

It was hypothesised that age, computer literacy and experience with touchscreens would influence the speed with which participants authenticated. The analysis showed a strong positive correlation between age and authentication time ($r = 0.42$, $p = 0.01$): the older the participants were, the slower they authenticated. There was a moderate

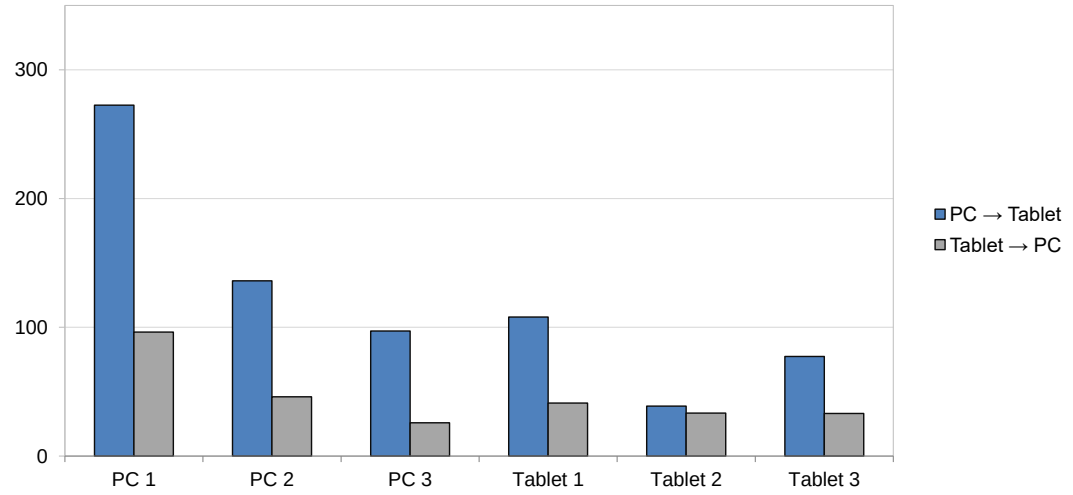

Figure 5.2: Average time, measured in seconds, for participants to log in using the Vernitski Authentication Grid on a PC and a tablet.

negative correlation between computer literacy and authentication time ($r = -0.36$, $p = 0.03$): the more computer literate a participant, the faster they authenticated. Finally, there was no statistically significant correlation between authentication speed on a tablet and experience with touchscreen devices ($p = 0.095$).

While the system recorded the time taken and numbers of attempts, the experimenter in the lab conducted structured observations of participant behaviour while they authenticated. Out of 36 participants, 31 pointed at the grid with their fingers or mouse while they were trying to find the row and the column where the characters of their passwords were. Three participants wanted to write their password down to facilitate breaking it down into pairs of characters.

5.2.5 Qualitative results

After having performed six successful logins, participants were asked to share their experiences. The interviews were audio-recorded and later transcribed. The following results emerged from a thematic analysis.

Effort

Out of 31 participants, 16 emphasised the authentication scheme was complicated. Eight participants said it required mental effort with P2 saying it was *"too taxing on the mind"*. P31 explained: *"You do have to give full attention to it, so you can't be doing any other stuff. You can't be on the phone and be like: 'Wait a minute I will just check my email. Oh, I have to login, hold for 5 minutes.' "* As this quote suggests, par-

ticipants also found the scheme time-consuming to use with 13 interviewees stressing it required more time than traditional passwords. P7 emphasised that authentication should be fast especially if it guards access to a critical task: *"If someone is going into cardiac arrest and those seconds matter, you need a procedure that's going to be as quick as possible, not something that's going to complicate things and, potentially, lose a life. I mean I realise logging on a system isn't life and death, but sometimes it is that crucial that you get in as quickly as possible."* Both effort and time needed contributed to the feeling of frustration in some participants. Six of them stressed they found the grid frustrating, P7 explained: *"It just seems like just the password itself is just far easier to remember and quicker to type in than having to find all these different letters and matching up and stuff. Like, if this was a real login for a site that I was on, and especially, you know, a business site or a professional site, it would drive me absolutely up the wall. I mean it would waste a lot of valuable time. So, like I don't understand the purpose of it actually. It's a big time waster."*

Out of 31 participants, 10 emphasised the learning curve for the practice of how to use the grid. They mentioned using the grid became easier with time which is confirmed by the quantitative data described in Section 5.2.4. Nevertheless, P31 stressed that using the grid would never be as easy as entering a password since it could not be automated: *"Because you have to look at, I mean it's not something you can just memorise. If you just want to check email you memorise a password, or you check in some password, like I memorise the key strokes on my keyboard, that's fine. I hate logging in using tablets and stuff but you know like, it's usually like if I was checking email it's usually just yeah, the pass code and a mobile and everything is there. But there is only your computer that can memorise your keyboard strokes, so you are like whatever, it's muscle memory, you know your password. So that is like, oh no, you actually have to be awake and you have to do like puzzle type thing and then it's annoying to do on an iPad, like say if you've just woken up or are really tired, then you can't log in."* The difficulty P31 described is that a login using the grid requires thinking and focus whereas standard password entry can be done out of muscle memory and there is no attention switch from the primary task.

High-value accounts

When asked if they would use this authentication scheme to log in in real life, participants generally gave varied responses. P15 responded: *"If I had to, yeah."* This might indicate that the participant had been forced to use mechanisms which required effort in the past and they had to get used to it because of the lack of alternatives. Out of 31 participants, 15 stressed their willingness to use the grid would depend on the context of its deployment. In 10 cases, participants emphasised it depended on what type of account they were trying to protect. Participants mentioned that better security would be needed for systems that hold sensitive or confidential information or that could lead to money loss as with access to online banking. P18 explained they would be more likely to use the VAG *"for websites for banks and stock exchanges and all, where the real money is involved and your interest is at stake"*. However, participants were not unanimous on this, four of them stressed their banking was secure enough already, P31 emphasised: *"I'm fine with my bank account I think because I have a key and I have to press one number and then it gives you a number and I do that, so it's not like I have additional mental stress. So, I wouldn't use it for my bank. And anyway my bank is really good because I have had fraud twice and they return it, they block it when I have had fraud, so I am fine."* Nevertheless, P24 stressed they would prefer to use the grid since it did not require carrying an additional device with them: *"Definitely this one is easier for my bank account than my token because I have to have my token everywhere I go. This way I can do it anywhere in the world, any platform, yeah. So in terms of easier sign-in, it's easier for bank account but in terms of security, I don't know."*

Frequency of use

Overall, nine participants stated their willingness to adopt the VAG depended on the frequency of use of the account it would protect. In four cases, participants made a link between frequency of use and account importance. They said they logged in to important accounts, like banking, less often than to their email, thus they would be more willing to put the effort into logging in for something that is high-value and infrequent. P23 explained: *"Logging in to like my email, because I do it so frequently, I probably wouldn't want to go through the hassle of that. But if it's something more secure, I probably would."* P25 expressed a similar view saying: *"I think that would*

take too long given how often I log into my email account." Perceptions of effort can change with repeated use as P07 explained: *"I found it quite enjoyable because I knew it was a different thing but I know if I had to do that every time I might enjoy it less."*

Security

Overall, ten participants stressed the authentication grid was more secure than passwords. Five of them stressed the complexity made it more secure, P26 explained it is difficult to use for the user, so it must be for the attacker too: *"It sounds like it's safer yes, because it's so complex, even for the user themselves."*

Experience of fraud

Related to security was prior experience of fraud which also influenced participants' stated willingness to adopt the VAG. P23 stressed: *"If you could choose, so I would probably want something like this for like my bank account or actually just recently my iTunes password was hacked, someone in Canada got my iTunes password so I had to close my account and cancel my credit card."* Conversely, P24 stressed they did not feel the need to strengthen their authentication through the use of a grid since they had not experienced fraud: *"I might feel more receptive to taking this up if I had ever been a victim of password fraud before, but I haven't. So, let's say if I'd had my account hacked into before, I would probably be much more receptive to using this one."*

5.2.6 Discussion

The study described in this section was a preliminary laboratory evaluation of the Vernitski Authentication Grid. A login took 88.6 s on average. The more practice the participants had, the faster they authenticated. Younger participants and those with higher computer literacy authenticated faster as well. There was no statistically significant relationship between experience with touchscreen devices and authentication speed on a tablet. Interestingly, participants authenticated faster on a tablet and subsequently faster on a PC. Previous research has shown that authentication on touchscreens poses many usability challenges [16, 99] and it is surprising to see that a login to the VAG was faster on a tablet than a PC.

Interestingly, in theory this authentication scheme was meant to guard against shoulder-surfing but actual user behaviour undermined this as participants pointed at the screen which might reveal to an attacker what the characters of their password are.

In terms of qualitative feedback, half of the participants thought the authentication grid was complex and over one third described it as time-consuming. Interestingly, when asked about their willingness to use it in the future, 19 participants made their decision dependent on the context where it would be deployed. They thought the grid could add extra security for systems holding important and sensitive information like banking and systems that they do not access as frequently. They emphasised the notion that for frequent accounts the password is in their muscle memory and they can enter it fast without much thinking. In such a situation, the use of the grid would not be suitable as it requires focus, effort and time. Especially since activities like email or Facebook are quick, users log in for a few minutes just to check if there has been anything new they need to attend to. This is in line with findings from previous studies. Brostoff and Sasse [46] found that if a login procedure was elaborate and time-consuming, participants logged in less often and once they logged in, they worked on the system for longer than those who logged in just using passwords.

The differentiation that participants made in terms of account importance and frequency of use is very interesting. Passwords were invented for a certain purpose (administering a shared computer), then introduced to various other systems regardless of context as a one-size-fits-all solution and nowadays virtually any website offering some service requires users to register with a username and password. Context of use is a fundamental HCI concept [246]; however, it is often forgotten by security researchers who do not account for differences between individuals and contextual factors such as account type or frequency of use [221].

Participants' stated willingness to use the VAG was associated with their risk perception. A participant who had experienced fraud stated they would be more likely to use it than a participant who had not. This is in line with previous studies where participants' exposure to cyberthreats made them more cautious in subsequent online interactions [141]. Interestingly, a proportion of participants inferred that the grid must be more secure than passwords because it is more difficult to use. This might be related to the "effort heuristic" as discussed in Chapter 2.1.2. In a study by Kruger et al. [150], participants assessed the quality of artwork as being higher when the effort put into it was greater. Participants relied on effort more if it was difficult to ascertain the quality of the artwork being evaluated. Similarly here, participants could have applied the

effort heuristic when they used the VAG. It cost them more effort, so they concluded it must offer better security.

Moreover, participants frequently compared using the VAG with entering traditional passwords. With an authentication time of 88.6 s, the grid performed poorly in this comparison. To put this number into perspective, Roth et al. [211] tested several types of cognitive trapdoor games, that is PIN-entry methods offering resilience to shoulder-surfing. The longest average entry time for these was around 25 s. To give another example, a standard login where the user has to enter a username and a password (both eight characters long) was analytically predicted to take approximately 14.8 s [242]. Additionally, participants emphasised that using the grid required their undivided attention and they could not enter their password from muscle memory. It is a general problem with one-time credentials that their entry cannot be automated. As noted by one participant logging in with the grid could be enjoyable as it breaks the routine but this might be a short-lived effect.

Limitations

The study was a preliminary laboratory evaluation of a new authentication mechanism. Such evaluations are multi-stage processes starting with a lab study, through a real-life deployment to an assessment post-adoption. Being the first stage in a long process of evaluation, the study had a range of limitations. Participants knew what was being studied what might have made them behave in an unnatural way (due to demand characteristics). Also, in real life a login is a gateway to some primary task and users' focus is not on security but on that primary task. This is something that could not have been recreated in this study since the explanation of the workings of the VAG needed to be quite elaborate and hiding the fact that the study was evaluating the grid was not possible. Finally, due to recruitment through a university participant pool, the study suffers from a volunteer bias and the sample consists of relatively young and well-educated individuals.

5.3 Human verification mechanisms

5.3.1 Background

The second study presented in this chapter evaluated three human verification mechanisms: reCAPTCHA, PlayThru and NoBot. reCAPTCHA (v 1.0)[1] is a text-based CAPTCHA from Google that requires the user to decipher distorted characters and enter them into a textfield. PlayThru[2] asks the user to drag and drop elements based on instructions (e.g., putting fish in the ocean as in Fig. 5.3). NoBot[3] verifies the user is human by making their screen flash different coloured lights and capturing images of the user's face with their device's camera. Although initially designed as an authentication mechanism, here NoBot is evaluated solely as a human verification mechanism. Verifying using NoBot does not explicitly require cognitive (solving a puzzle) and physical effort (entering characters, dragging and dropping elements), but requires that the user position their face in view of the camera. Visually, NoBot starts off as a widget like reCAPTCHA and PlayThru but then fills the whole screen by displaying instructions and then flashing lights. The idea for the study was conceived when the author's research team was approached by NoBot's creators. They asserted NoBot was effortless to verify and authenticate with and they wanted the team to conduct a usability study to test this empirically. Note that the study was designed and partially conducted before the introduction of reCAPTCHA v 2.0 that only requires a tick from the user [94].

5.3.2 Study aims and hypotheses

The study aimed to assess the time and perceived effort needed to verify using three verification mechanisms: reCAPTCHA, PlayThru and NoBot. Since in the previous study, participants voiced in what contexts they would find it appropriate to use the authentication grid, this study formalised it by asking participants to indicate in which every-day situations they would find it suitable to use a given verification mechanism.

Since in real life security is not a user's primary task, in the study participants were tasked with buying three event tickets and completing three distinct verifications as part of that. There were three experimental conditions in the study. In the first condition,

[1] https://developers.google.com/recaptcha/old/intro
[2] http://areyouahuman.com/demo-playthru
[3] The developer did not commercialise this product and wishes not to be named.

participants were asked to complete three ticket purchases with a NoBot verification on each occasion, using a laptop (condition denoted as NB_{lap}). In the second condition, participants completed the same tasks but on a tablet (NB_{tab}). In the third condition, participants also made three purchases, but used a different verification mechanism each time (denoted as $C3_{mix}$): reCAPTCHA (reC_{mix}), PlayThru (PT_{mix}) and NoBot (NB_{mix}) with the order of the mechanisms being randomised each time.

The following hypotheses were devised to assess the time and workload required to verify using the three mechanisms.

Time. It is hypothesised that (H1) there will be a difference in how fast users will verify between the three mechanisms used: reCAPTCHA, PlayThru and NoBot. (H2) There will be a difference in the time that users take to verify using NoBot on a laptop and a tablet. (H3) The time needed to verify using NoBot will decrease the more practice users have.

Workload. It is hypothesised that (H4) there will be a difference in perceived workload between different human verification mechanisms. (H5) There will be a difference in perceived workload between verifying using NoBot on a laptop and a tablet.

5.3.3 Method

Participants

Participants were recruited through the UCL Psychology Subject Pool. Mean age was 25.5 years (range: 18–53, $SD = 6.8$). Out of 87 participants, 57 were female and 30 were male. Four participants had A-Levels, 29 some undergraduate education (no completed degree), 27 an undergraduate degree, 26 a postgraduate degree (Masters or PhD) and one person had vocational training. There were 27 participants in condition $C3_{mix}$, 31 in NB_{lap} and 29 in NB_{tab}.

Apparatus and materials

The ticket purchasing website. A ticket purchasing site called "SimTikats" was created for the purpose of the study. It offered different tickets for a variety of events ranging from concerts to football matches.

The devices. Depending on condition, participants either completed the study on a Dell laptop E5540 with a screen size of $15''$ or a Nexus 7 (2013 model) tablet with a screen size of $7''$.

Recordings. While participants were purchasing each ticket, the time it took them to complete the verification process was recorded. A voice recorder was used to capture participants' reactions to the mechanisms and their responses in the interviews. Participants' interactions with the devices were video-taped over their shoulder. For participants in the mixed condition where there were questionnaires between the different mechanisms, the camera was switched off for these and switched back on for the next purchase.

Questionnaires. Participants were asked to fill in the following post-task questionnaires. (1) NASA TLX tool [106] was used to assess participants' perceptions of the workload involved in the human verification process. A pen and paper version of NASA TLX was chosen since Noyes and Bruneau [175] found that it required less cognitive effort than processing the information on a screen. Participants were asked to complete the full NASA TLX with cards for pairwise comparisons of the different aspects of workload, to capture perceived importance of the workload factors. (2) Participants were asked to pick three adjectives from a list of 24 different adjectives that best described their interaction with the mechanism. The population of the adjective list was informed in part by the work of Benedek and Miner [22], and partly by the adjectives suggested in the pilot study. As shown in Appendix C, the adjectives were displayed in two columns – positive on the left and negative on the right – to shorten the reading time that was needed. (3) Since effort and acceptance are contextual, participants were asked to indicate in which context they would be willing to use the mechanism. There was a list of seven different contexts (Table 5.5) which were taken from real-life uses of CAPTCHAs. For each, participants could choose from three options: "Sure, no problem.", "No, no way." or "I don't do this." if they never engaged in an activity (e.g., had never contributed to an online forum). (4) Finally, an instantiated Technology Acceptance Model (TAM) questionnaire [61] was given to participants to gauge how acceptant they were of the three mechanisms.

Procedure

Upon participant's arrival to the laboratory, the study was briefly explained to them by the experimenter. The experimenter stressed that none of the technologies tested in the study were created by the researchers themselves but by external companies and the researchers were acting as independent assessors. The participant was asked if they

were sensitive to flashing lights as the NoBot verification process involves shining light on the user's face. If no sensitivity was reported, they were then asked to read through the information sheet, encouraged to raise any questions they might have, after which they would sign a consent form. Once this was completed, the experimenter switched on the voice recorder and the video-camera and the participant was asked to make three ticket purchases.

When on the mock-up site, there were three steps: (i) selecting a ticket for an event of the participant's choice, (ii) completing one of the human verification processes, and then (iii) entering the details (e.g., name, address) of a fictitious person named Adam/Anne Johnson at the checkout to complete the transaction.

The experimenter stayed in the room throughout the session, taking notes and reacting to any comments raised by the participant, where responses were kept to an absolute minimum. After the purchase of the third ticket, the experimenter switched the camera off. The participant was then asked about their experience with the mechanism they had just used, and encouraged to voice any speculations about its purpose. The real purpose – if not deduced already – would be revealed by saying it was to replace CAPTCHAs and a print-out with various CAPTCHAs was shown to make sure the participant knew what was meant.

A brief interview followed where the participant was asked to elaborate on their experience with NoBot and text-based CAPTCHAs. After that, they were asked to complete a NASA Task Load Index (TLX) questionnaire, select three adjectives that described their experience with NoBot, indicate in what situations they would use it and fill in an instantiated Technology Acceptance Model (TAM) questionnaire. Afterwards, they were encouraged to voice their final impressions of NoBot, both the advantages and disadvantages of using the mechanism. Then, they were asked if they would like the company to keep or delete the images of their face taken in the study. Finally, they were thanked for participating in the study and received £10 in cash.

In the third condition where participants tried all three mechanisms ($C3_{mix}$), they were briefly interviewed and asked to fill in the questionnaires after each ticket purchase (i.e., after having experienced each mechanism). In addition to the questionnaires used in the other two conditions, they were also asked to rank the mechanisms in order of preference.

Ethics

The study design and protocol were reviewed and approved by UCL's Research Ethics Committee (approval number: 3615/004). The Committee requested that participants are told in the information sheet that one of the mechanisms will take images of their face and that they are asked if they are sensitive to flashing lights. Both changes were subsequently implemented. Several measures were taken to protect the participants in the study. There was a written agreement put in place with NoBot's developers that if requested by the participant, they will delete the images of the participant's face taken in the study. During the study, participants were asked to enter the details of a persona rather than their own. Participants' interactions with the interfaces were video-taped over their shoulder in order not to record their faces.

5.3.4 Quantitative results

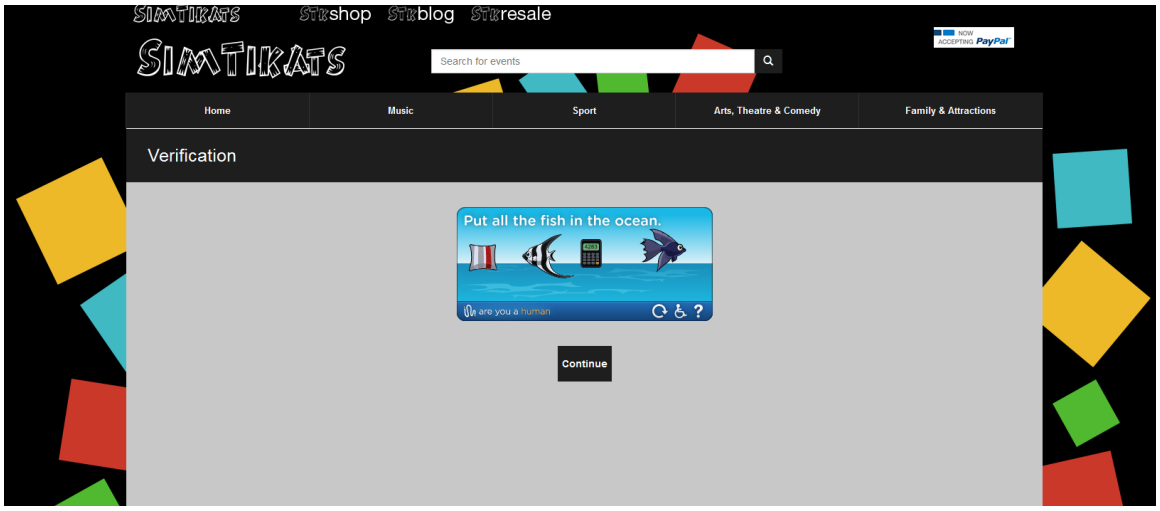

Figure 5.3: A page from the study website showing PlayThru.

Time

The set-up of the website used for the experiment was such that the verification was on a separate page, participants had to verify using reCAPTCHA, PlayThru or NoBot and then click the 'Continue' button underneath as can be seen in Figure 5.3. This set-up mimics the one encountered on real ticket purchasing sites where the CAPTCHA is either on a separate page or as an overlay (e.g., Ticketmaster UK). When establishing how long participants took to verify, the entire time how long they spent on the verification webpage was considered.

Three mechanisms. In the mixed condition where participants tried all three mechanisms once, it took them on average 20.2 seconds ($SD = 13.4$) to solve a reCAPTCHA, 28.7 seconds ($SD = 13.9$) to complete a PlayThru game and 70.1 seconds ($SD = 50.7$) to verify using NoBot (Fig. 5.4). A repeated measures ANOVA determined that the time taken to use each mechanism differed statistically significantly ($F(1.23, 35.681) = 27.076$, $p < 0.0001$). There were no significant order effects. H1 is therefore supported.

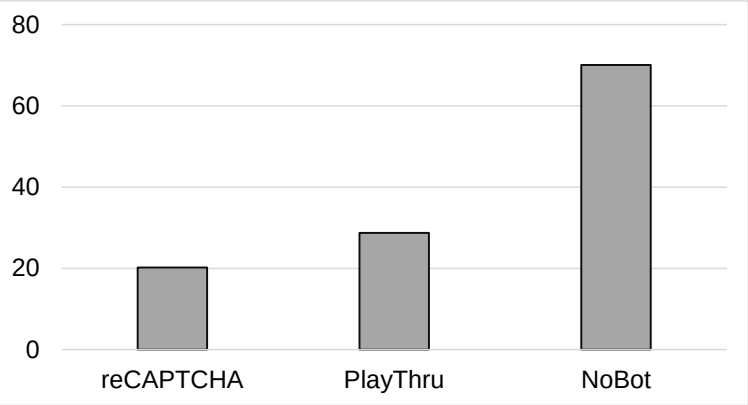

Figure 5.4: Average time, measured in seconds, for participants to verify using reCAPTCHA, PlayThru and NoBot on a laptop.

NoBot: laptop vs. tablet. Figure 5.5 shows the time needed to verify using NoBot across devices and verifications. On average, across the three verifications participants took 42.2 seconds on a laptop and 54.4 seconds on a tablet, this difference is statistically significant ($p = 0.022$, two-tailed t-test). H2 is therefore supported.

Practice. A repeated measures ANOVA determined that the time taken to verify on a laptop (NB_{lap}) differed significantly between verifications using NoBot ($F(1.964, 54.986) = 5.673, p = 0.006$). Post-hoc tests using the Bonferroni correction revealed that participants' verification time dropped with practice from the first to second verification (50.3 vs. 41.2), which was not statistically significant ($p = 0.131$). Again, time to verify using NoBot dropped between the second and the third verification (41.2 vs. 35) but this difference was not statistically significant either ($p = 0.549$). However, there was a statistically significant difference between the first and the third verification ($p = 0.011$). A repeated measures ANOVA determined that the time taken to verify on a tablet (NB_{tab}) differed significantly between different attempts to verify using NoBot ($F(1.983, 53.528) = 4.686, p = 0.014$). Post-hoc tests using the Bon-

ferroni correction revealed that participants' verification time dropped with practice from the first to second verification (78.7 vs. 44.7), which was statistically significant ($p = 0.003$). Again, time to verify using NoBot dropped between the second and the third verification (44.7 vs. 40.9) but this difference was not statistically significant. However, there was a statistically significant difference between the first and the third verification ($p = 0.009$). It is worth noting that the time for the verification process is impacted by the speed of the machine's Internet connection, the reliability of that connection, and any need to repeat the process (e.g., failures due to not being positioned in view of the camera). H3 is therefore supported.

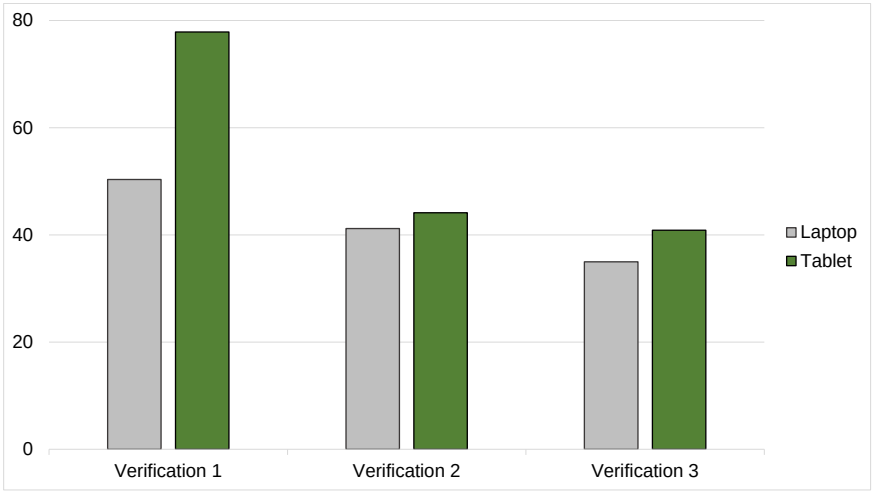

Figure 5.5: Average time, measured in seconds, for participants to verify using NoBot across devices and verifications.

Workload (NASA TLX)

Table 5.1 shows NASA TLX scores for using reCAPTCHA, PlayThru and NoBot. Frustration was the only aspect of workload where there was a significant difference between the three mechanisms $F(1.93,\ 54.036) = 7.132$, $p = 0.002$). Post-hoc tests using the Bonferroni correction revealed that participants reported a frustration of 8.73 for reCAPTCHA and 6.31 for PlayThru, which was not statistically significant ($p = 0.167$). The difference between reCAPTCHA and NoBot (8.73 vs. 11.7) was not significant either ($p = 0.27$). However, there was a statistically significant difference between PlayThru and NoBot (6.31 vs. 11.7) ($p = 0.002$). H4 is therefore partially supported.

Table 5.2 presents the NASA TLX scores for using NoBot on a laptop and a tablet. Although the scores tended to be higher for the tablet, the difference was only

TLX aspect	reCAP	PlayThru	NoBot	P-value
Mental Demand	10.75	9.24	8.89	0.62, n.s.
Physical Demand	4.33	8.45	8.34	0.64, n.s.
Temporal Demand	4.62	3.08	6.47	0.59, n.s.
Performance	3.62	5.46	5.51	0.38, n.s.
Effort	8.12	8.21	6.46	0.71, n.s.
Frustration	8.73	6.31	11.7	0.002**
Overall workload	48.4	37	46.7	0.19, n.s.

Table 5.1: Comparison of NASA TLX scores for reCAPTCHA, PlayThru and NoBot.

significant for *Physical Demand*. While the physical demand of using NoBot on a laptop had an average score of 2.9, using the mechanism on a tablet had an average score of 7.2 ($p = 0.004$). H5 is therefore partially supported.

TLX aspect	Laptop	Tablet	P-value
Mental Demand	4.2	4.8	0.13, n.s.
Physical Demand	2.9	7.2	0.004**
Temporal Demand	5.6	5.1	0.85, n.s.
Performance	7.2	7.9	0.5, n.s.
Effort	5.4	7	0.18, n.s.
Frustration	5.4	5.4	0.4, n.s.
Overall workload	29.9	37.4	0.13, n.s.

Table 5.2: Comparison of NASA TLX scores for verifying using NoBot on a laptop and tablet.

Adjectives

Tables 5.3 and 5.4 show the top six adjectives participants selected to describe the mechanism(s) they used.

Laptop		Tablet	
effortless	13	effortless	11
fast	9	intuitive	10
intuitive	8	slow	9
weird	7	easy to use	8
easy to use	6	acceptable	8
exciting	3	fast	7

Table 5.3: The top six adjectives participants chose to describe verifying using NoBot on a laptop and tablet. The numbers refer to the number of participants who chose these adjectives. Each participant was asked to choose three adjectives to describe NoBot.

On both devices, participants found NoBot “effortless”, “intuitive” and “easy to use”. However, the perceptions of speed differed. For ‘fast’, nine participants chose this adjective to describe verifying using NoBot on a laptop and seven participants on the tablet. For ‘slow’, one participant picked the adjective to describe verifying using

NoBot on the laptop and nine participants on the tablet. This difference is statistically significant ($p = 0.037$, Fisher's exact test.)

reCAPTCHA		**PlayThru**		**NoBot**	
normal	14	acceptable	14	unpredictable	9
acceptable	13	exciting	9	weird	9
effortful	9	effortless	7	creepy	8
easy to use	8	intuitive	7	fast	8
predictable	7	great	8	effortful	6
fast	6	easy to use	5	untrustworthy	6

Table 5.4: The top six adjectives participants chose to describe reCAPTCHA, PlayThru and NoBot. The numbers refer to the number of participants who chose these adjectives. Each participant was asked to choose three adjectives for each mechanism they tried.

However, participants were far less positive about NoBot in the mixed condition. While they found reCAPTCHA and PlayThru to be "normal" and "acceptable", NoBot was "unpredictable", "weird" and "creepy". The only positive adjective about NoBot that made it to the top six was "fast".

Different contexts

Table 5.5 shows the percentages of participants willing to use the three mechanisms in different contexts. Conditions NB_{lap} and NB_{tab} were combined into NB_{l+t} since there was no statistically significant difference between them. Overall, participants indicated they were most willing to use the mechanisms for ticket purchasing which is no surprise since this was the scenario used in the study.

Context	**NB**$_{l+t}$	**reC**$_{mix}$	**PT**$_{mix}$	**NB**$_{mix}$
Contributing to an online forum	16	59	15	24
Buying tickets online	76	93	79	55
Browsing for plane tickets	50	76	55	45
Checking in for flights online	62	86	54	52
Topping up your Oyster online	52	66	69	31
Bidding on items on eBay	38	66	69	31
Logging in to Facebook from a different computer	47	66	66	34

Table 5.5: Percentages of participants willing to use reCAPTCHA, PlayThru and NoBot in different contexts.

Participants were least willing to verify when contributing to a forum using PlayThru and NoBot. For PlayThru, some of them explained the game was not serious enough in the context of a serious activity. For NoBot, six participants stressed that the nature of forums is that one wants to stay anonymous. This reveals a com-

mon perception that the service provider or other users would see the pictures taken by NoBot which is not the intention of the company.

Frequency of use mattered too, PM03[4] stated: *"Topping up my Oyster[5] is a small task so why do you have to get my picture? Because I top up like every month or every week so I don't want to do it, it gets in the way."* Similarly PM05 stated frequency of use and value of the transaction mattered in the different contexts, *"For Oyster, I felt it's such a basic thing where you're not going to be spending that much money, it doesn't make that much sense. For bidding on eBay, it would slow everything down, it wouldn't work. I spend a lot of time on eBay so that just wouldn't be a thing."*

TAM

To gauge participants' acceptance of the three mechanisms, they were asked to complete an instantiated TAM questionnaire.

Statement	reCAP	PlayThru	NoBot
1. Using this product increases the security of my online activities.	5.3	5	4.48
2. Using this product gives my greater protection of my online activities.	5	5	4
3. This product enables me to accomplish tasks more quickly.	3	3.6	3.3
4. This product supports critical aspects.	4.2	4.2	3.6
5. This product increases my productivity.	3.2	3.4	2.8
6. This product encourages me to conduct more activities online.	3.4	3.6	2.5
7. This product allows me to finish tasks quicker and more securely.	3.7	4.4	3.2
8. This product enhances my effectiveness in protecting my online activities.	4.8	4.3	3.8
9. This product makes it more secure to do my activities online.	5.1	4.8	3.8
10. Overall, I find this product useful for my online activities.	4.9	4.5	3.3

Table 5.6: Statements for an instantiated Technology Acceptance Model and an average score of in how far participants agreed with them on a seven-point Likert scale.

Table 5.6 presents the statements used and average scores of in how far participants agreed with them on a seven-point Likert scale (1 – strongly disagree, 7 – strongly agree). An ANOVA and then post-hoc tests were conducted on the mean scores for each mechanism and any significant differences are highlighted. Where table cells are shaded there was a statistically significant difference between mechanisms, with the darker cell being the lower value in each pair. For conditions NB_{lap} and NB_{tab}, there were no statistically significant differences between the scores that participants gave. In the comparison $C3_{mix}$, the differences in average scores between reCAPTCHA and PlayThru were small and not statistically significant. NoBot re-

[4]The first letter of the participant number indicates which condition they were in: "M"–$C3_{mix}$, "L"–NB_{lap} and "T"–NB_{tab}.

[5]An Oyster card is a payment card for London's public transport.

ceived the lowest ratings of all mechanisms for all statements that were statistically significant. Participants rated NoBot as offering them not as great protection of their online activities, less encouraging them to conduct more activities online and overall, being less useful for their online activities.

The lower scores for protection and security are interesting here since in the interviews ten participants considered NoBot to be more secure than other mechanisms. An explanation for this apparent discrepancy could be that our participants perceived security more broadly, considering not just the protection of the website from bots but also the security of their own user account and associated data. Looking at participants' comments, 17 of them imagined myriad attacks that could be performed on NoBot and 32 worried about the implications of a compromise of the NoBot image database.

Ranking

Towards the end of the study session, participants were asked to rank the mechanisms in order of their preference. There were five choices: reCAPTCHA, PlayThru, NoBot, no CAPTCHA-like mechanism and a different mechanism. Figure 5.6 shows how many participants ranked the mechanisms on each position.

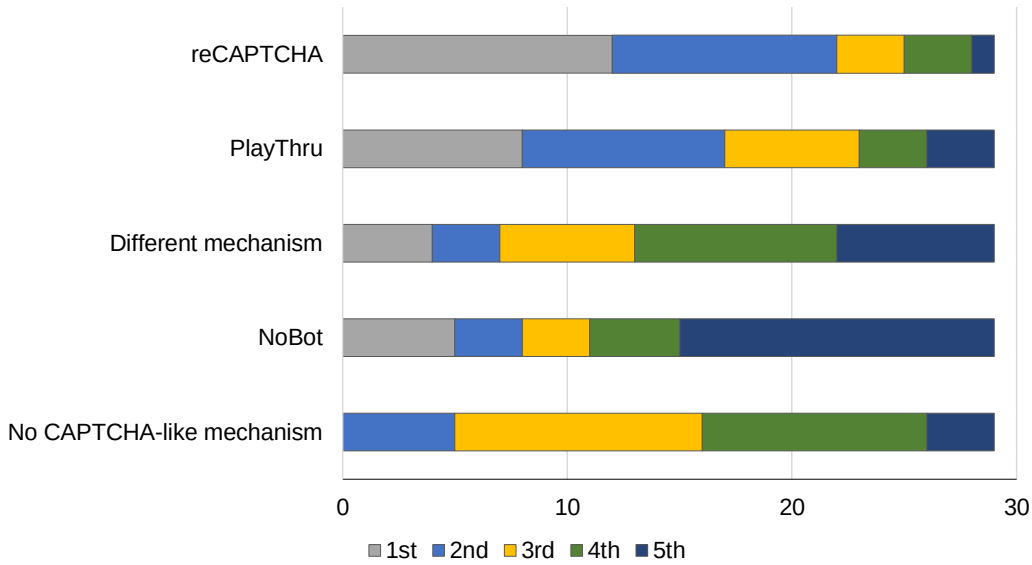

Figure 5.6: Numbers of participants in C3$_{mix}$ providing rankings for each mechanism and hypothetical alternatives. Mechanisms are ordered from highest to lowest average rank.

reCAPTCHA had the highest average rank followed by PlayThru, a different mechanism and NoBot. "No CAPTCHA-like mechanism" seemed to be the least desirable choice, it had the lowest average rank and was never ranked as the top choice.

PM06 motivated their choice of ranking it low by saying: *"I expect there is probably some benefit to all of us to having some kind of a security mechanism on some websites and some purchasing processes. I definitely wouldn't want to get rid of it altogether, not that I know in much detail what benefits they have."*

PM00 explained their ranking: *"I have always being using that, and it's a habit although it's sometimes quite difficult to read the letters but I feel cool with it. The PlayThru, it's like a game, it's not very serious so I don't... but I still quite like it though. If we don't have the reCAPTCHA, I'm OK with the PlayThru. [Why?] It's easy, it's fun and then you don't really put effort to doing that."* Similarly, PM03 stressed the importance of being used to reCAPTCHA: *"It [PlayThru] disturbed my usual routine, it required some kind of effort which actually I didn't exhort that much, I just guessed through it and it was correct what I did. If not I imagine, it would have been very frustrating. So this one [reCAPTCHA] is fine and I'm more comfortable with doing it.*"

5.3.5 Qualitative results

Any comments participants made about the mechanisms tested were transcribed and coded by two researchers. Each researcher coded half of the transcripts and both coded a subset of six transcripts (two from each condition) and they discussed any differences. The following sections present the results of the thematic analysis of the data.

Experiences with text-based CAPTCHAs

Throughout the study sessions, participants discussed their experiences with text-based CAPTCHAs, which were overwhelmingly negative. "Annoying" (44 mentions), "frustrating" (20) and "hate" (10) were the most common words participants used to describe their experiences with CAPTCHAs. Out of 87 participants, 64 stressed they found CAPTCHAs to be hard to read. 28 participants stated their coping strategy for reCAPTCHAs was just to press the 'refresh' button and try another one. PL27 called for an alternative: *"[CAPTCHAs] are just annoying, they should find a better system."* These negative experiences were a confirmation that users to get used to expending effort on security mechanisms if they have no choice but this does not mean they will start liking it but would rather consider them a necessary evil.

Perceptions of time and effort

In the interviews, participants often made comments about the speed of the mechanisms, sometimes saying which one was faster or slower. There was an observable discrepancy between actual and perceived length of the verification processes. To investigate this, the actual timings of mechanism use were compared with comments about mechanism speed as made by participants in the comparative condition. To illustrate, PM10 stated *"for me I'd want to do things much quicker, so I'd prefer NoBot as it's faster so I can get the tickets I want"*. This participant's actual timings were 43 seconds for reCAPTCHA, 19 for PlayThru and 53 for NoBot meaning NoBot was actually the slowest. Overall, six such discrepancies were identified. In four cases, NoBot was perceived to be faster than it actually was. One can speculate why this is the case, PM14 explained: *"[PlayThru] was quick but it didn't really matter because it was a game, it was nice, you could get sidetracked a little bit but it's nice."* It seems one does not notice the passage of time when the activity is entertaining.

What might explain this difference in perceptions could be the fact that the effort expended was coloured by the type of effort required, for NoBot there was less physical and cognitive effort and mostly time. PM01 stated: *"It's the first time the screen just recognises my face and don't need to do anything, just stay here, effortless."* The participants further explained what made this experience so fast, saying: *"[For PlayThru] I still need to read the sentence to know what they want me to do. I need to think and for NoBot, I just stay here and it's done."*

Participants had the tendency to rationalise why the verification effort required of them was justified. Some stressed by completing CAPTCHAs, they contributed to the services remaining available online, otherwise they would need to physically go to a box office to purchase their tickets. The inconvenience of CAPTCHAs was perceived as small relative to the convenience of being able to buy tickets over the Internet. Additionally, two participants stressed CAPTCHAs were a good thing for society since they were used to digitise books.

Participants also had a good feeling about contributing to overall Internet security. When asked why they ranked "no CAPTCHA-like mechanism" the lowest, PM05 explained: *"I guess there was a reason why they had these in place. That is I may not appreciate but it probably saves a lot time compared to when a lot of machines would*

be entering stuff. If there was no verification process at all, it would probably take longer to filter through what's simulated and what's real."

Some participants also expressed the view that CAPTCHAs had to be hard to give them a sense of security, PM02 explained: *"The chess one [PlayThru] was a bit simple, it feels like anyone can do it, it feels actually less secure, although I guess it's just checking if you are human. [...] And PlayThru might actually be the best out of the three because reCAPTCHA still requires a bit more effort than PlayThru, I think [...] I think it's quite easy to use so I feel like everyone and anyone could use it whereas with reCAPTCHA you have to go through some effort and NoBot, you have to fit your face. But PlayThru, because it's so easy and accessible, I feel like it's too easy in a way."* The experimenter then asked why CAPTCHAs had to be hard, and the participant explained that a hard CAPTCHA gave them time to think through whether they wanted to complete a transaction: *"Let's say if you're buying something and it's too easy, you'd just got through with it and if you want to change your mind or something, it's kind of too late."*

Apart from giving participants a sense of security or a moment longer to re-consider the transaction, some participants also over-attributed the security that CAPTCHAs gave them, assigning them "powers" they do not have. When asked why they rated "no CAPTCHA-like mechanism", PM04 stated *"If my information was stolen, my credit card and there was nothing to check that it wasn't me making the purchase.* [Do reCAPTCHAs help with this?] *But it helps fight spam bots too. If someone stole my password, my information was say Ticketmaster UK, if it was hacked and my information was stored, someone can buy a ticket. If someone can hack Sony, they can hack Ticketmaster."*

Security

When asked about their experience with the mechanisms, participants often expressed views on their security, rather than just usability or acceptability. Interestingly, they perceived the security of the mechanisms not based on how difficult they were for bots to solve and how well they protected the site, but they spoke about how well their transaction details were protected and how their images were stored. 32 participants described scenarios where the security of their images would be compromised, for example if the company was hacked. Also, rather than thinking of these mechanisms

as black boxes, many participants actively deliberated how the mechanisms operated, with 17 participants actively questioning their resilience against attacks and thinking of ways to break them, for example through the use of pictures or masks.

PT28 explained they were not sure if CAPTCHAs were providing any security at all: *"I really really hate them, they give you a hurdle to jump over and you've been onto that site 100 times, I find them really infuriating, I don't even know if they work, I can't see the point of them. I can't read them, they're too close or the letters are at an angle, or they have these dots which sometimes you can't see."* The CAPTCHA hurdle is even more frustrating if the user is under time pressure, PM07 shared their story: *"A couple of months ago, I was ready half an hour before they* [the tickets] *even went online, I had everything ready, my information. And then I like got through and I got the ticket and it had this thing that... CAPTCHA box, it's called. And the laptop that I was on had an ad blocker or something, every time that I typed the word that was in it, it wouldn't allow it and I was so panicking, I had to get that ticket... I didn't realise I had to turn it... that I had to change some settings to allow it or it was or I genuinely couldn't read it, so I tried to play the audio but that was confusing and I oh... then I ended up losing the ticket. Basically, I did get another one but it was pretty stressful."*

Five participants were disillusioned with human verification mechanisms in general saying they were protecting the businesses rather than the users. PL15 talked about NoBot: *"I'm even less inclined for people to use my photo, you know, just for the company's benefit. That is something that fully identifies you and it's actually for the company's benefit more than even my own. OK, it's security of buying my own ticket, so probably I would be even less inclined actually. Especially for what it's asking for to protect a company, no! [...] It's being sold to me that it's actually for my protection but it's not."*

Privacy

Often after their first encounter with NoBot, participants speculated that the mechanism was there for identification rather than verification that they were human. This is not surprising since the technology is using a biometric solution. Because of the scenario of a ticket-buying website, 18 participants thought that a picture of them was taken, to check if it was really them when they arrived at the event venue with the ticket they purchased online.

Throughout the study sessions participants raised privacy concerns with NoBot capturing and storing images of their faces. A total of 29 participants did not like the fact that someone could see their picture, and within this group, they either worried about the service provider such as the ticket purchasing site seeing their images, or other parties such as NoBot's developer or other users of the website (such as other forum members or bidders on eBay). Ten participants emphasised that they did not like the possibility that a company capturing personal images might store the images and not delete them.

When asked about what specifically invoked privacy concerns, eight participants stressed that a human face is special, by which they meant it was unique and identifying. Participants also elaborated on what kind of reassurance they would need to be able to use NoBot confidently. Ten participants would have liked a confirmation that images will be kept securely or deleted. Nine participants would have welcomed a privacy statement and two the display of security certificates. Interestingly, seven people stressed they would be more confident using NoBot if they saw that other people were using it too. Related to this, 20 participants emphasised that the technology was novel and in turn that they needed to develop a trust relationship with it.

There were also participants who said they were too ignorant to care, PM03 explained: *"Maybe some persons are not comfortable with showing their face.* [What about you?] *I am comfortable with doing that. When you are ignorant about computer systems, what they do with your data and all that, it's easier for you to give information. I'm kind of like that person so I don't really mind giving my face. I can trust the privacy statements that they give."*

PM05 noted that there might be a trade-off between privacy and convenience: *"I prefer the two to NoBot because they are less invasive but equally there was a little bit more effort than the reCAPTCHA."*

Context

Throughout the study sessions, participants specified situations where they believed it would be appropriate to use the different mechanisms. This was in part prompted by them being asked to indicate their willingness to use mechanism in different contexts for which the results are described in Section 5.3.5, but even before being prompted participants elaborated on when they would or would not use it and why. Six par-

ticipants stressed that using NoBot would be more suitable for high-value purchases where there is more money at stake, PM06 explained: *"I guess if someone is buying something more expensive, there is kind of more risk of fraud, using someone else's details. In that case, it might be a lot more beneficial but other than that I don't think it's necessary for small amounts of money or like logging in to things, it's definitely a bit extreme."* Because online contexts vary, nine participants emphasised they would like to be given choice which mechanism they would like to use to verify.

Some participants also explained that they would not like for the image of their face to be associated with what they were buying, PM05 explained: *"I liked it less because I don't like putting my face on the Internet; taking a picture of me and who knows what I'm buying."* This again had to do with participants' perceptions of NoBot as an identification mechanism that shares the images with the service provider.

26 participants stated that NoBot was too invasive or heavy-handed for simply checking if the user was human. PM04 explained: *"I'm trying to buy a ticket and not getting into a governmental building. It seems a bit much."* They further elaborated: *"I do support some sort of verification but I prefer getting less of myself, as in my face, facial recognition if I have to. Obviously, if that was part of my job or my livelihood, I wouldn't have a problem with NoBot, the visual confirmation, personal recognition but for purchasing tickets it's a bit too much."*. Other participants stressed that the gain in security was not worth the loss in privacy, PM06 explained: *"I can see why it would sometimes be useful in terms of fraudulent activity and purchasing things but I still don't think it's worth the breach of privacy to store an image of anyone ever who buys anything online or tried to log in to a website."*

Control

There was a recurring theme amongst participants of not feeling in control when using NoBot, because the outcome of the verification process depended on many factors that were outside of their influence. PM04 stressed: *"You're kind of held hostage to the quality of the camera"*. Participants emphasised that they felt more in control with reCAPTCHA because it was for them to read and enter the characters and they could always 'try harder' if the system did not accept their submission. With NoBot, they could only do as much as positioning themselves in front of the camera, and if the system decided they were not human, there was little more that they could do at that

time to change the outcome. PT26 explained: "[About reCAPTCHA] *I really don't mind having several attempts actually it's not something that is annoying and it's easy to do. This* [NoBot] *can have similar potential problems, my picture couldn't be taken properly, had to try several times, three times that was annoying and I'm not sure why, could be the flashes, really testing my patience, perhaps that's no fault of my own. With CAPTCHA, I know it was somehow my fault, because I didn't interpret the* [letters] *and numbers correctly, the* [letters] *are there, I might have misread it, it works better next time around, I've a certain amount of control* [over] *the CAPTCHA system compared to this NoBot system...*"

Similarly, PL11 said they would not rely on NoBot for critical activities because they considered verifying using a text-based CAPTCHA to be more reliable: *"like checking in for flights, I would not use NoBot because there might be some error and I could not check in for my flight. Then I would use the traditional because it's easier, I can refresh the images and try again, so I think it would be faster. So if there is something urgent, I would not use it* [NoBot]. *But for every-day uses, I would advocate this."*

Similarly, PL16 expressed being worried about system failure: *"I'd be concerned that just because you know computers and machines and that have been proven to be imperfect, I'd be a little bit concerned about it being maybe not able to accurately read my face. It's just the squiggly words, CAPTCHAs, it's just the way these words sometimes, the letters are undecipherable which is a kind of quality imperfection, if you will. I'd be concerned that someday, one of these websites wouldn't be able to read my face and I can't get in. And if it's an important one, let's say self-assessment [tax submission] or checking in online or something, then I'm really screwed because of a system failure."*

Worries over NoBot's reliability also revealed participants' mental models of security, PL17 elaborated: *"I wouldn't really use it though personally because I've had fraud on my debit card before so I don't trust this kind of thing anymore. I would rather do it myself using the CAPTCHA or get it from the tickets' case, I'd just get it from some office and do it. Then I know I properly get it myself and then there is no fraud in it. Whereas with the face thing, no."* PL18 expressed they had a low level of tolerance for the inaccuracy of security measures: *"If something is to guard me, protect*

me and if it's not accurate, it pisses me off, you know. It's like having a guard at a gate who is a drunk."

PL14 stressed that they would not like to be photographed when buying something sensitive but they would see the advantage of storing a picture in case a criminal successfully completed a transaction in their name: *"I'm actually not sure about the picture though, yeah. Because I wouldn't want them to know, if it's for verification purchases, then why do you need to store the picture? [...] Sometimes I think it's useful to store pictures though. Let's say some criminal wants to buy some... purchase something online and then the police can find out who bought that thing online and they can use that image to find the criminal. I don't know. But what are the chances of this happening?"*

5.3.6 Discussion

The study described in this section examined the time, perceived effort and suitability of three human verification mechanisms: reCAPTCHA, PlayThru and NoBot. Out of 87 participants, 29 used all three mechanisms once, 27 used NoBot three times in succession on a laptop, and 31 similarly used NoBot on a tablet. The results show that participants were on average the fastest verifying using reCAPTCHA, followed by PlayThru and NoBot. Although practice reduced verification times, participants took longer verifying using NoBot on a tablet than on a laptop finding it more physically demanding on a tablet. Further, they were the most frustrated by NoBot and the least by PlayThru.

Familiarity

Several participants emphasised that it was hard to compare reCAPTCHA with the other two mechanisms, due to their familiarity with text-based CAPTCHAs. This is supported by the quantitative data where "normal" and "acceptable" were the top adjectives chosen to describe reCAPTCHA. Participants were largely annoyed by traditional CAPTCHAs but emphasised they understood they were there to add security, although they were not always clear about how this security advantage was afforded. reCAPTCHA also had the advantage of having been around for longer which was a source of trust, PM26 explained: *"the fact that it already exists, is pre-existing, somebody approved it somewhere, it's been around a while... that seems fairly safe"*. The

fact that participants ranked no CAPTCHA-like mechanism as the least favourable option indicates that security was seen as important.

Participants were used to reCAPTCHAs so they did not question the effort that was demanded of them, they were in a habit of solving them and they had developed coping strategies. They were fine with the increased effort of trying to decipher them and re-trying was seen as a normal part of the process.

Perceptions of effort

In the mixed condition, there were discrepancies between actual and perceived time to verify. Perceptions of time could have been coloured by the entertainment factor of PlayThru and NoBot, as well as their novelty. In future work, one could examine if perceptions would change as a person becomes familiar with the technology. The NoBot face capture mechanism does not require explicit effort, but preparation disrupts the flow of the primary task (positioning the face in front of the device camera and remaining still, etc.). NoBot was the most time-consuming of the mechanisms, which might have led to increased user frustration.

Participants were more critical of NoBot in the mixed condition, calling it "unpredictable", "weird" and "creepy" rather than "effortless", "fast" and "intuitive" as when used in isolation. It is important to be careful not to draw any far-reaching conclusions from this since two factors varied between the conditions: repeated vs. single use, and having direct comparison with other mechanisms and having only one mechanism. One can only speculate that repeated use might have made participants more accepting of the technology. Coventry et al. [56] showed that participants were more acceptant of a technology after having tried it. It might also be the case that direct comparison may have encouraged participants to be more critical. This would confirm an important design rule, that when asked to evaluate a technology users should – where there are alternatives – be shown more than one design, to be able to compare and better articulate their needs and preferences through relative assessment.

NoBot, as a biometric mechanism, was perceived to be the most frustrating for users, where taking pictures of their faces was regarded as invasive, and its implicit nature left users feeling outside of control. Responses also suggested that if the actions of a technology are not adequately explained, this can lead to the development of incorrect mental models – in the case of NoBot, participants thought the mechanism

was used to singularly identify them or capture a photo that would appear on their ticket (for the purposes of security and convenience at the venue). Participants also worried about who would process and store the images. There was additional concern at potentially having photos of their face linked with purchases that they had made.

Context

Participants reported they would make the decision of whether to use NoBot based on the context of use. They found NoBot too heavy-handed for more frequent and low-value transactions yet more appropriate for high-value purchases. This is similar to the findings from the study on the VAG described in Section 5.2.5 where participants believed that a more time-consuming mechanism should be used for infrequent and high-value transactions. This can be related to CAPTCHAs, in that a number of participants believed that security should be difficult.

Participants disliked reCAPTCHAs but saw them as a necessary evil, trusting that they were there for the right reasons. Participants using PlayThru thought the entertainment factor made it unsuitable for serious online activities and "too easy" to provide enough protection. For NoBot, participants did not notice the passage of time when verifying using NoBot but they objected to the collection and storage of their images.

The study with its holistic approach stands out from previous research of CAPTCHAs and usability of CAPTCHAs in several ways. Prior studies examined how the difficulty of human verification is pushed to users, without fully capturing the user's perspective. Here, humans are not seen as CAPTCHA solving machines but as individuals who consider the effort of security in terms of their goals and expectations [113]. Another challenge is to afford effortless security that strikes the right balance between providing reassurance and demanding effort, as users recognise a need for security.

Limitations

The study was a laboratory-based evaluation and as such it suffers from a number of shortcomings. Although participants were given a primary task, it might not have been a realistic one since participants were in an unfamiliar environment using unfamiliar devices and mock buying tickets. Another limitation is that that the mechanisms were

tested on a sample of participants who were relatively young and well-educated, which might limit the generalisability of the results.

5.4 Conclusions

In both studies, proposed mechanisms were tested in a preliminary lab evaluation that assessed the time needed, the perceived effort and participants' acceptance. Participants used the mechanisms repeatedly, also between-subjects on devices with a physical and a virtual keyboard. The studies considered effort and usability broadly, they did not evaluate how fast participants were to examine 'user performance' but aimed to learn more about the context of how the authentication effort would fit in with their primary tasks.

The first study, a preliminary evaluation of the Vernitski Authentication Grid, showed that authenticating using the mechanism was perceived to be effortful and time-consuming. Participants stated they would only find it acceptable for infrequent and high-value transactions. The second study examined three human verification mechanisms: reCAPTCHA, PlayThru and NoBot in more detail. The findings show that users were the fastest verifying using reCAPTCHA and it was their preferred mechanism, with familiarity playing a role. NoBot was considered privacy-invasive due to the capture of users' pictures. Although especially in the first study, the context of use provided was limited, the results show that participants saw authentication and verification very contextually. They considered frequency of use, value of the account and level of security required. Participants translated high level of effort required to log in or verify into thinking that it would offer them better protection as completing effortful security tasks gave them reassurance. The case of NoBot shows that users want to be in control of the verification process and might not trust a technology that keeps the process hidden away from them.

Both studies looked at technologies deployed "in a brownfield", meaning the VAG, PlayThru and NoBot were designed to replace existing security mechanisms – passwords and CAPTCHAs. The results showed that participants had a preference for incumbent technologies due to familiarity and acceptability. To be adopted, a replacement technology not only needs to be better than its predecessor but users also need to be motivated enough to make the switch and abandon the use of the existing one.

Chapter 6

Disclosure in webforms

6.1 Overview

Webforms are the primary way in which data is explicitly collected from users online. Webforms stand in the users' way of achieving a goal and designers have developed ways to minimise the user effort required in completing them. The studies presented in this chapter investigated how users fill in webforms and what motivates them to disclose information about themselves.

The first study showed a high prevalence of deliberate and unpaid over-disclosure of personal information. Careful experimental design ensured that participants understood that additional data disclosure was voluntary. Participants regularly completed more form fields than required, even providing more details than requested. Making some fields mandatory jeopardised voluntary completion of the remaining optional fields. Conversely, monetary incentives for disclosing those same fields brought a positive spillover by increasing completion for other optional fields.

The second study was designed to explore what could help users better manage information disclosure. It aimed to assess the effectiveness of privacy-enhancing warning messages and to explore the design space in a search towards more usable interfaces. A range of eight warnings was trialled which alerted users that they have completed non-mandatory fields. The results show that the warnings successfully allowed participants to limit the submission of optional, personal information in the webform. Interestingly, participants preferred the warning that left them most in control over which information they submit by allowing manual reconfiguration. Against the principle of least effort, the warning that required minimal effort from the user by deleting all optional data for them was least preferred.

6.2 Quantifying disclosure

6.2.1 Background

Introduced into the HTML standard twenty years ago [26], webforms are now part of users' daily web browsing routine. The form is the primary mechanism for collecting personal information from users, who type in personal details such as their name, date of birth or address to complete online transactions.

Design to lower completion effort

Completing webforms is time-consuming and users are often confronted with them when they are about to perform a transaction meaning they are a hurdle to completing a primary task. For instance, even major service providers see conversion rates as low as 10% on their account registration pages [161]. Thus, web designers and researchers have been trying to find ways to ease the completion of a webform. This strand of research focuses on lowering the cognitive and physical effort of completing forms: label positioning (above the textfield, on the left, below) and formatting, the mechanism for indicating mandatory fields (e.g., with a red asterisk) and unified text field to reduce tabbing and mouse-keyboard switching [14]. Although some practitioners have strong opinions – for instance on label formatting – it does not seem to matter as long as the user experience is consistent [124].

Autocompletion

One way to reduce the effort associated with the completion of webforms is to automatically fill in fields for the user. The autocompletion of webforms was first introduced in IE4 in 1997 as 'Form AutoFill'. The browser uses a combination of cues, including commonly used field names and names of previously completed fields, to match them across websites. The browser can then suggest values for formfields it has seen before, reducing the need to type the same information again. Today, autocompletion is limited to values typed into text fields. For one-time or sensitive entries, such as payment authorisation codes, webform creators often prevent automatic form filling, thereby mandating manual completion of fields. Users can configure their browsers not to store and suggest form values and delete individual values from their autocomplete suggestions.

Autocompletion vs. security and privacy

In the early 2000s, despite built-in browser support for the autocomplete feature, there was also demand for third-party tools, such as FormWhiz [185] and Gator eWallet. Gator eWallet called itself "the smart online companion" [86] and was marketed as enabling the user to "fill in FORMS with no typing". In addition to reducing the effort to complete forms, it also transmitted the user's data to the GAIN Publishing advertising network to optimise the adverts displayed to the user until it was shut down in July 2006 [87].

The emergence of the autocomplete feature led consumers to question whether their privacy might be violated. For example, PC Magazine in 1999 asked "What's to stop a hacker from stealing your personal data [that] is popping up in the Auto Complete window"? [212, p. 108]. Despite previously entered form data being stored locally in an encrypted file, researchers were able to read out the autocomplete suggestions [101]. In summary, web developers were encouraged to use autocomplete because it can "collect demographic data more easily" and faster [83], the common assumption being that the form itself is a nuisance or a "pain" [279]. Yet, there does not seem to be any thorough study or analysis demonstrating the extent to which the use of autocomplete encourages data entry on webforms. Seemingly obvious assumptions regarding webform completion have shown surprising results before. In the economics of privacy, it is regularly assumed that monetary or other incentives would encourage users to ignore their privacy concerns when filling out webforms [248]. Yet, recent research has found that consumers show no preference for merchants with less privacy-invasive web order forms even when product and price are equal [25].

6.2.2 Study aims and hypotheses

Webforms are ubiquitous on the Internet, they are the primary way in which data is collected from users online. Filling in a webform is a classic secondary task that stands in the way of completing a primary task. Practitioners have looked into ways in which completing forms could be made easier but true empirical evidence is missing. The aim of this study was to establish how users disclose information through webforms, what motivates them to do so and what role effort and time play in the process.

H1: Effort. Users will fill in the minimum amount of information required to receive payment.

H2: Mandatoriness. Making some fields mandatory will increase completion rates for those fields.

H3: Bonus. Providing a bonus for filling in certain fields will increase completion rates for those fields.

6.2.3 Method

The study was conducted on the Amazon Mechanical Turk (mTurk) platform, it consisted of an experiment and a follow-up questionnaire. In the experiment, participants were given a 12-question webform (Fig. 6.1). In the follow-up questionnaire, participants were asked about their experience with the form, perceived effort and reasons for disclosing information.

Design

The study was a between-subjects design. Table 6.1 provides an overview of the five experimental conditions. A 2×2 experimental design varied the payment: either \$0.25 (denoted T_{25}) or \$0.50 (T_{50}), and the number of mandatory fields: either just check questions ($T_{\underline{\star}}$) or check questions and favourite colour and weather ($T_{\overline{\star}}$). In the fifth condition ($T_{B\underline{25}}$), only the two check questions were mandatory, but participants were awarded a bonus of \$0.25 if they answered the questions on favourite colour and weather.

data requirement		compensation	
minimum	extra	\$0.25	\$0.50
high	—	$T_{\overline{25}}$: 209	$T_{\overline{50}}$: 445
low	—	$T_{\underline{25}}$: 202	$T_{\underline{50}}$: 216
low	bonus for high	$T_{B\underline{25}}$: 181	

Table 6.1: Experimental conditions with numbers of valid entries.

Participants

Participants were workers on the mTurk platform, all recruited to be based in the United States. Neither the webform nor the follow-up questionnaire asked participants for demographics. All the same, age and gender may be inferred for those who provided their date of birth and first name. The median year of birth was 1981, corresponding to an age of about 30 years at the time of the experiment, after outlier

correction. The cohort of the 1980s accounted for 41% of the sample. To establish participants' gender, the first names they provided were matched against the list of common first names from the 1990 US census[1]. For ambiguous names, the most prevalent gender was chosen. According to this inference, 34% of participants were male and 33% female. For the remainder, the first name was not given or could not be found in the census name files.

Apparatus

Experiment. Figure 6.1 shows a screenshot of the form participants were asked to complete. The form was simply headed "About yourself". There was no cover story for the experiment and no explicit indication of the purpose for data collection. The form was designed to ensure that everything took place on a single page, making it clear to participants that submitting the form also finished the task. To avoid a trust bias, the form itself did not mention any university affiliation in words or pictures.

The form contained questions of varying levels of sensitivity ranging from date of birth (DOB) in Q11 to favourite colour in Q3. In the middle of the form, check questions were placed to ensure that participants have read and understood the instructions (Q5 & Q6). These questions did not stand out visually. They also had no visual indicator, such as an asterisk, to show they were mandatory.

All information was collected using text fields. Participants could enter data in any format they wished. For instance, the field asking for DOB did not require the participant to enter data in a specific way, making partial input of day, month and/or year possible. The questions were phrased to make them of comparable length so that none of them stands out. In order to obtain a homogeneous population, the study was restricted to participants based in the US. Thus, the questions used American English spelling (e.g., "favorite color") and US currency (e.g., $100).

Follow-up questionnaire. At least one day after completing the experimental webform, participants were invited to fill in a feedback questionnaire for an additional payment of $0.65. There was a response rate of 74%.

At the beginning of the questionnaire, participants were reminded of the original form they completed with a small screenshot, and then asked a series of twelve questions regarding their motives for participating, time spent, enjoyment, and willingness

[1] http://www.census.gov/genealogy/www/data/1990surnames/names_files.html

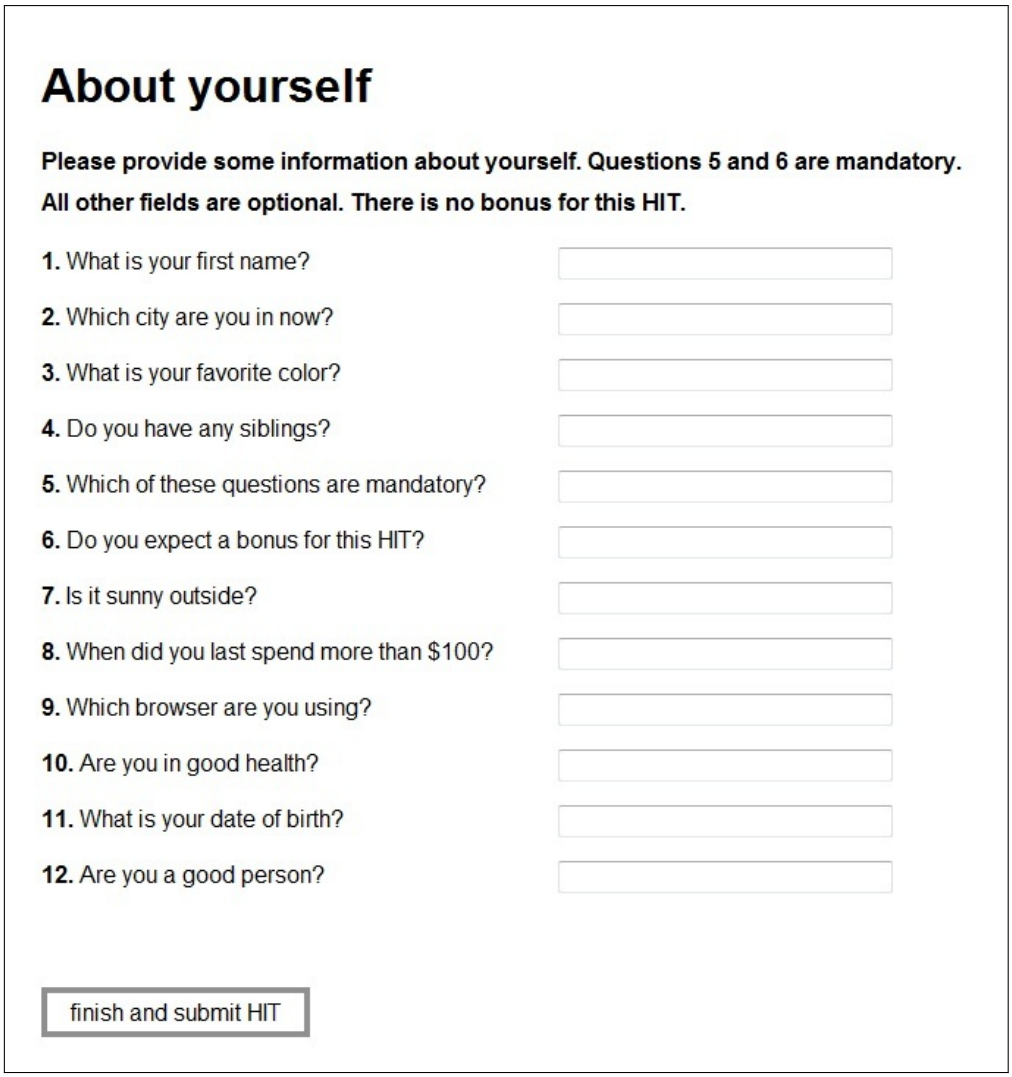

Figure 6.1: A screenshot of the webform used in the study as it appeared in conditions $T_{\underline{\star}}$.

to participate in a similar study. Finally, they were asked whether they had revealed any personal or sensitive information, and if so, which data items they considered to be sensitive. Depending on their disclosure on the original form, participants were also asked about their motives for (not) providing their DOB.

Ethics

Both studies described in this chapter were conducted after successfully having gone through an ethics review at the University of Cambridge where the author's collaborators were based as outlined in Section 1.4.

6.2.4 Experiment results

Figure 6.2 shows the disclosure rates for the ten items of information collected across different conditions. Overall, date of birth was the data item omitted most often. Of all

received submissions, 57% of participants included full details (i.e., day, month, year) and 68% provided parts of their DOB. Full DOB was submitted significantly less often than the second- and third-most often omitted data items, which were spending and first name ($p = 0.005$ and $p < 0.0001$, two-tailed paired t-tests). The DOB can thus be considered the most sensitive item on the form. Spending had a relatively low completion rate, one could speculate that this could be due to it being a sensitive finance-related information or participants not being able to recall their last major purchase. Answers concerning weather and favourite colour were included most frequently.

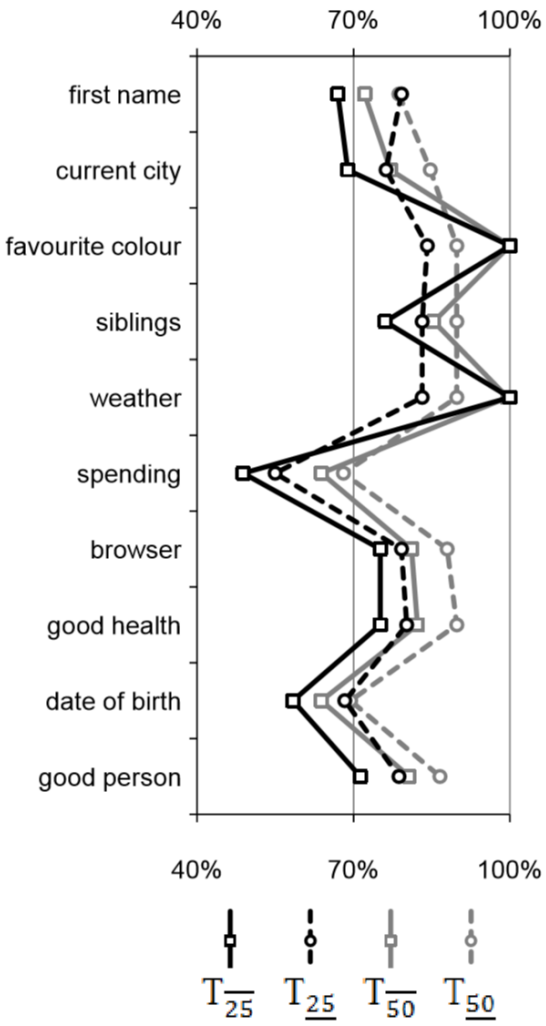

Figure 6.2: Disclosure rates for each of the ten items across four treatments.

As outlined in Section 6.2.3, age and gender of participants were inferred based on the information they provided. There was a significant ($p < 0001$, two-tailed t-test) trend that older participants answered fewer questions. However, age only explains 2% of the variance in completion rates and the effect size is minute. On average, the older

half provided 0.18 items less than the younger half of participants ($r^2 = 0.024$). There was no significant difference in completion rates between men and women ($p = 0.22$, two-tailed t-test).

Effort

Measured effort. It was hypothesised that in order to minimise effort, participants would fill in the minimum required information to receive payment. Completing optional fields requires time, cognitive effort of coming up with responses and physical effort of typing them in. Not surprisingly, those participants who completed all instead of none of the optional fields took significantly longer ($p < 0.0001$, t-test). A regression analysis reveals that participants spent around 57 seconds reading the form plus additional 3.5 seconds per field completed ($p < 0.0001$). Figure 6.3 shows the time participants took to complete the form fields.

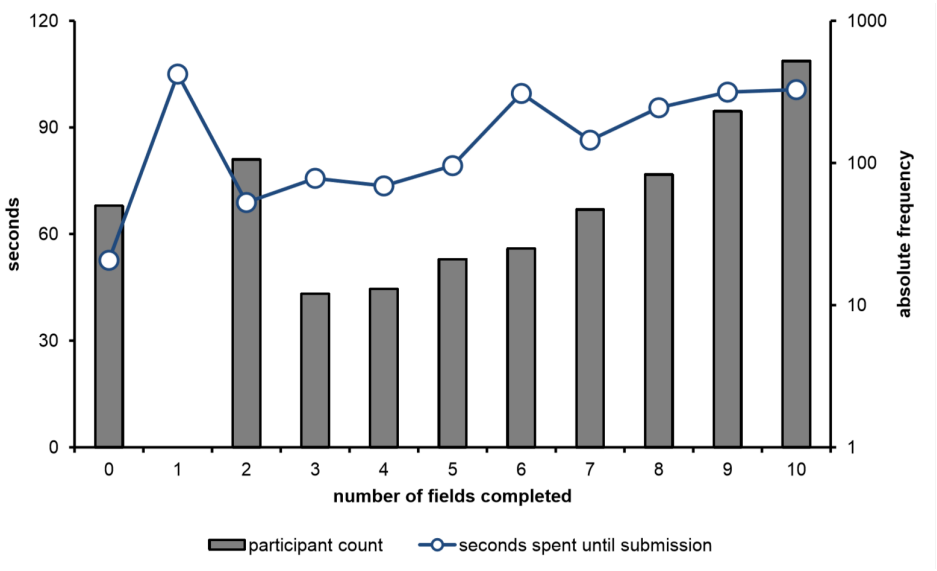

Figure 6.3: Average time spent on completing the form and the number of fields completed on top of the check questions.

Participants who over-disclose also have to type more. The minimum number of characters typed in total increases linearly in the number of fields completed at a rate of 4.0 characters per field ($p < 0.0001$, t-test on the regression coefficients; outlier detection based on inter-quartile range; 98% variance explained). However, the median number of characters typed in total increases quadratically in the number of fields completed (99% variance explained). A special case of over-disclosure are overly verbose answers to simple questions. Examples of overly verbose responses include:

"No. It's currently cloudy and rainy"[2] (P21) or *"no, its cloudy and snowing"* (P504) when asked about if the weather was sunny, and *"last week on textbooks"* (P526) or *"buying oakley sunglasses last summer"* (P1096) when asked when they last spent over $100. 6% of all participants answering the weather questions provided details that were not asked for; 14% of those indicating the time of their $100+ purchase also indicated the purpose of spending. In both cases, the prevalence of overly detailed answers is significant ($p < 0.0001$, Fisher's exact test). H1 is therefore rejected.

Perceived effort. The actual time it took each participant to complete the form was compared with the estimate they provided in the follow-up questionnaire. Most participants largely overestimated the time spent on the form – 86% of participants had an estimate of the time spent that exceeded the actual time. For 13% of the participants, their estimate was more than 10 times higher than the actual time.

Making fields mandatory

It was hypothesised that making some fields mandatory would increase completion rates for those fields. Data disclosure in conditions $T_{\underline{\star}}$ and $T_{\overline{\star}}$ was compared to establish if the mandatory revelation level makes a difference for the disclosure of the remaining, optional fields. In $T_{\overline{\star}}$, weather and favourite colour were mandatory; they are the two data items revealed voluntarily most often in $T_{\underline{\star}}$.

Making two low-sensitivity fields mandatory decreased the revelation ratio for the high sensitivity item date of birth ($p < 0.02$, G-test) as well as for the medium sensitivity item of being a good person ($p < 0.04$, G-test).

Only participants who provided answers to questions 3 and 7 are considered for this analysis, regardless of whether they were in a condition where these questions were mandatory or optional. Within this group, disclosure behaviour for the remaining fields depended on whether questions 3 and 7 were marked as mandatory or optional. The average number of fields completed when questions 3 and 7 were marked as mandatory is reduced by about 1.3 fields in $T_{\overline{\star}}$ compared to $T_{\underline{\star}}$ ($p < 0.0001$, two-tailed t-test). The data therefore suggests that as the number of mandatory fields in the form increased, the total number of completed fields decreased. Also negatively affected is the revelation ratio for date of birth: fewer participants were willing to disclose it in $T_{\overline{\star}}$ than in $T_{\underline{\star}}$ ($p < 0.0001$, G-test).

[2] All quotes in this chapter are reproduced as they were written by participants, no *[sic]* is used.

Providing a bonus

It was hypothesised that providing a bonus for filling in certain fields would increase completion rates for those fields. Providing a bonus for certain fields yields the same disclosure ratio as mandatoriness ($p = 0.55$, Fisher's exact test). A bonus improves disclosure for fields that are optional ($T_{B\underline{25}}$ vs. $T_{\underline{\star}}$: $p < 0.0001$, Fisher's exact test). A comparison between $T_{B\underline{25}}$ and $T_{\overline{\underline{50}}}$ shows that incentives for disclosing low sensitivity data also increase disclosure for the remaining, optional, medium, and high sensitivity fields on the same form (good person: $p = 0.002$, DOB: $p < 0.001$; Fisher's exact test). H3 is therefore supported.

6.2.5 Questionnaire results

Reasons for completing the form

Using an open-ended question, participants were asked in the follow-up questionnaire why they had completed the form. The free-text answers were coded into up to three reasons for each participant.

54% of participants stated they participated for the money. This comes as no surprise for study conducted on a crowd-sourcing platform. 30% said they completed the form as it looked easy. 15% indicated they had participated out of joy while 25% because it was interesting. Original responses include: *"I enjoy filling out surveys"* (P1022), *"I enjoy doing surveys as a way to destress"* (P48) or *"It looked interesting, fun and easy to do"* (P3). This is opposed to the received wisdom that form-filling is a nuisance. To have their opinion heard or to help research was named by 3% and 8% of the follow-up participants respectively. Exemplary answers include: *"I think it's really cool to be part of a statistic analysis, to contribute my thoughts and experiences to a collective body of information"* (P67), *"the opportunity to present an underrepresented demographic (conservatives, mothers) in surveys"* (P163) or *"I like taking surveys to get my opinions heard"* (P232). Those being motivated by helping research named both the researcher and research per se, such as: *"I enjoy helping researchers"* (P83), *"Any help I can be for research, I am glad to do"* (P520) or *"I appreciate helping (even if only a little) with research"* (P90).

These motives work towards completing specific fields in the form (e.g., for opinion shaping) and completing more fields on the form (e.g., when wanting to help re-

search). They can be understood as antecedents for over-disclosure. There is a high proportion of participants motivated by the base reward or the form itself, combined with the low proportion of those who used the form to express opinions (2%), or were motivated by trust in the university (4%). This is promising as it hints that the results might have more general validity. A G-test for each condition individually and for all conditions together indicates that over-disclosing date of birth, being a good person and favourite colour, does not depend on whether participants were motivated by a financial incentive. Participants who provided more data items (nine or ten versus eight and below completed fields) also enjoyed the form more (approaching significance: $p = 0.06$, G-test).

Reasons for disclosing DOB

In the follow-up questionnaire, participants were asked about the reasons why they provided or did not provide their DOB. The group that provided their DOB on the original form and answered this question in the follow-up questionnaire had 455 participants. A thematic analysis of the free-text responses identified four main reasons why participants provided their DOB: (i) they felt it was not a sensitive piece of information, (ii) they wanted to be helpful, (iii) they had the habit of submitting a complete form, and (iv) they hoped for future rewards.

First, 207 participants (45.5%) explained that they did not mind sharing their DOB as it was not a sensitive piece of information. P1021 stated: *"That doesn't bother me - it's not really personal."*, similarly P24 added: *"I didn't feel like it was personal or sensitive."* P712 explained this information was already available out there: *"It does not bother me to share my date of birth (I study astrology, so that information tends to be more public than for most people)."* Second, 108 (23.7%) participants stated they wanted to help research, P5 explained: *"I wanted to complete the survey with all the information that was asked of me, to further help any research that was necessary."* Third, 61 (13.4%) participants felt the urge to complete every field of the form: *"I feel a certain obligation to completely fill out surveys"* (P437), *"A completionist instinct I guess. I didn't feel like leaving any fields blank."* (P370), *"Completeness"* (P609), *"i felt that i should complete all aspets"* (P557), *"I like to fully comply with requests"* (P63). Participants were habituated to fill in forms. The following responses were given: *"Force of habit"* (P794) and *"I probably automatically put it down without*

thinking" (P856). Some participants reported they did not know, if they had provided their date of birth, which indicates the lack of a conscious decision which might be an indication that they filled in the form habitually. P341 stated their first instinct was just to fill in the blanks: *"I had filled out all the boxes at first, until I realized it wasn't required"*. Fourth, 36 (7.9%) participants anticipated future payoffs and intended to improve their social capital on the platform. P744 stated they participated *"to boost my mturk hit approval rate"* and P44 responded saying: *"Even though it was optional, I thought that if I did not disclose the information, you would be unable to classify me for future HITs and I would miss out on the opportunities."* Additionally to the four groups of reasons, there were also cases of extroversion, P292 stated *"I have a unique birthday, it being on christmas, so i just wanted to share."*

Reasons for not disclosing DOB

The group that did not provide their DOB on the original form and answered the question about their motivation in the follow-up questionnaire had 363 participants. They named different reasons for not having provided their DOB and these can be grouped into the following categories: (i) participants found the DOB to be too personal, (ii) they stressed it was optional, (iii) they worried about potential misuse of this information or (iv) could not remember.

First, 138 (38%) participants stated this information was either personal and/or identifying, P1016 stated: *"It's private and could potentially identify me."* Second, 111 (30.6%) participants responded saying that this information was optional. P851 explained that DOB was not among the required questions: *"You didn't ask for it in questions 5 and 6."* Third, 47 (13%) participants mentioned the possibility that the information could be misused, voicing both security (*"DoB is a key piece of information for identity thieves."* – P146) and privacy concerns (targeting for marketing). Fourth, 36 (9.9%) participants could not remember. While some could not remember being asked (*"I don't remember being asked for it. Apparently I missed a question."* – P622), others could not remember the reason why they did not provide it (*"I'm not sure why I didn't to be honest."* – P1139). Out of these 36 participants, three actually provided their date of birth as a response to the question why they did not provide their date of birth. For example, P410 wrote: *"I did not realize I did not give mt date of birth, its 11/27/75."*

6.2.6 Discussion

This section reports on a study conducted with around 1,500 on Amazon Mechanical Turk asking them to fill in a form with 10 items of personal information. There were five experimental conditions varying the payment and mandatoriness of fields. Participants largely over-disclosed providing information that was not required to receive payment, even filling in additional details that were not requested. Once certain fields were mandatory, the disclosure for the remaining optional fields dropped. Conversely, once there was a bonus for some fields, the disclosure for the remaining fields went up. In terms of effort, the analysis of participants' questionnaire responses gave four interrelated reasons. Participants wanted to be helpful, they filled in the form out of habit, they did not want to leave any fields blank and they took satisfaction from being compliant.

Each new field that a participant completed cost them time, cognitive and physical effort. Participant effort was captured in terms of the time they spent on filling in the form. Interestingly, 86% of the participants overestimated how long it took them to submit the form, which might be an indicator that they perceived the form to be a nuisance. In terms of physical and/or cognitive effort, 30% stated that the form was easy.

There was a strong impact of habit on participants' disclosure behaviour. Also what links to that was being conversational. If one is asked in a normal offline conversation if it is sunny outside, very few would just reply 'no'. Instead, they would politely elaborate on the weather conditions. Similarly in the study, P526 explained: *"a little, not very, more like slightly overcast"*. The influence of habit has not been studied extensively in relation to privacy behaviours. An exception is a study by Ko [137] who investigated self-disclosure through blogs amongst Taiwanese youths. The researcher found a strong influence of habit that accounted for the continuous use of such platforms.

Participants also wanted to help research, it appears they projected on the task what was expected of them, what it meant to be a good worker/participant/person and acted accordingly.

6.3 Informing disclosure

6.3.1 Study aims and hypotheses

The preceding study showed that users do not keep their effort in dealing with webforms to a minimum. The study showed that participants spent more effort and time on webforms than required as they knowingly provided optional details at 3.5 seconds per field. Prevalence of over-disclosure was quantified at 57 to 87% for data items such as date of birth or favourite colour respectively.

More prevalent than voluntary over-disclosure is accidental over-disclosure: the onus is on the users to tell mandatory and optional fields apart and this can be a difficult task. When surveying 140 websites, Preibusch and Bonneau [192] found that less than one third provided visual or textual indicators of which of the webform fields were mandatory. The preceding study, however, showed that users make clear distinctions between mandatory and optional fields and – if possible – selectively decide which fields to leave blank. On a form which features a mix of mandatory and optional fields, the latter see significantly lower disclosure rates compared to the former and compared to a form featuring no mandatory fields. This might be an indication that if users knew what was required, they would only fill in fields that were mandatory and in this way limit disclosure and protect their privacy.

In the absence of visual or other hints regarding the mandatoriness of input fields, identifying the minimum set of data items could be time- and effort-consuming. It would require a user to leave a field blank and attempt to submit the form to see if the form is accepted despite the field(s) being empty. Depending on the server logic behind the form this would need to be done with every combination of fields. Additionally, after an unsuccessful submission, the previously entered data might become removed and would require re-entry.

The study aimed to identify the characteristics of actionable privacy alerts, focusing in particular on the wording of the warning messages, the different options offered to the user, and the characteristics of the audience. Eight hypotheses guided the analysis of the results.

Warning text

Warnings currently deployed in browsers typically feature a short message that mentions the diagnosis, why the warning was displayed, and optionally explains the threat (Fig. 6.4).

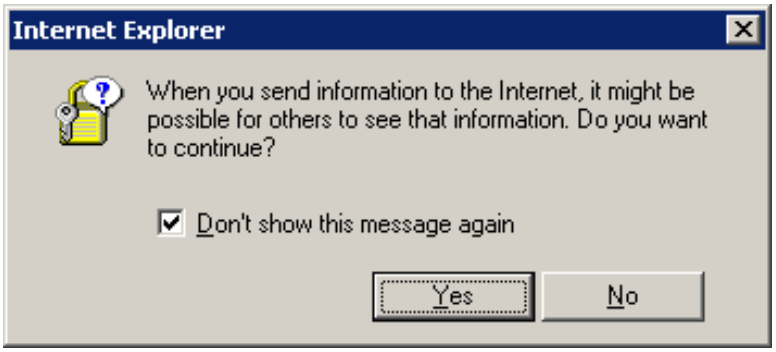

Figure 6.4: Screenshot of an existing privacy warning in Internet Explorer.

Since as discussed earlier, security events are associated with a more definite outcome than privacy events, it is hypothesised that users who will see a security warning will be more likely to delete information than those who will see a privacy warning. It is further hypothesised that the mention of security and privacy will have a stronger impact on the user than not providing such an explanation.

H1a: Users presented with warnings mentioning privacy and security threats will delete previously entered data.

H1b: Users presented with a warning mentioning security will be more likely to delete previously entered data than those presented with a warning mentioning privacy.

H1c: Users presented with a warning which has no accompanying explanation will be less likely to delete previously entered information than those presented with a warning that mentions a privacy or security threat.

Item sensitivity

It is hypothesised that the impact of the warning will depend on item sensitivity. Users who are asked to provide personal data on a webform can restrict its proliferation by either leaving the fields blank or by deleting previously entered data. Data items with low sensitivity *do not need* to be deleted; data items with high sensitivity *cannot* be deleted if users did not provide them in the first place. One can therefore predict a U-shaped relationship between data item sensitivity and its likelihood of deletion after a privacy warning.

H2: Users will be more likely to remove sensitive information while keeping the less sensitive information in place (U-shaped relationship).

Time and effort

Filling out a form requires an amount of work and humans tend to be protective of what they have produced. Therefore, it is hypothesised that the more time the participant spent on filling out the form, the less likely they will be to remove their entries and let all this work and time invested be for nothing. This reflects the sunk cost fallacy and the escalation of commitment as discussed in Section 2.1.2.

H3a: The more time users spent on filling out the form, the less likely they will be to remove the information entered.

Furthermore, as current security advice instructs the user to perform time- and effort-consuming checks, tool support should be better integrated with security mechanisms to take some of the effort off the user. It is therefore hypothesised that a warning that provides the options to remove non-mandatory information for the user will be preferred.

H3b: Users will prefer options that minimise effort and save time over other options.

Demographic characteristics

It is hypothesised that the decision to go back and delete some information after having seen a warning could be associated with users' computer literacy as the level of knowledge might help them more accurately assess the risk linked to disclosing personal information online. As previous research demonstrated an impact of cyberthreat exposure on user behaviour [141], it is hypothesised that prior experience of viruses, fraud and scam will be associated with users' willingness to provide the requested information.

H4a: Users will differ in their deletion/altering behaviour depending on their levels of computer literacy.

H4b: Users will differ in their deletion/altering behaviour depending on their cyberthreat exposure.

6.3.2 Method

Design

As previously, the study consisted of an experiment and a follow-up questionnaire.

Experiment

The webform and all instructions used were directly replicated from the preceding study (Section 6.2.3).

Upon attempting to submit the webform, a warning was shown if the participant had filled in any non-mandatory fields (Fig. 6.5). No warning was displayed unless at least one optional field was filled in. The warning text typically consisted of two sentences – the statement saying the user has completed some non-mandatory fields and the mention that this can be a threat to their security or privacy.

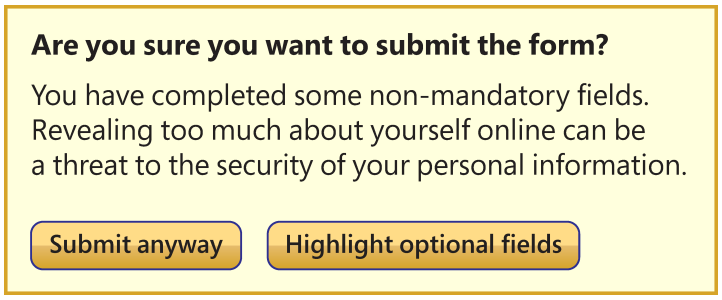

Figure 6.5: An example of a warning used in the study (condition WS1HO, *security warning – highlight optional fields*).

The study was a between-subjects design. The form presented as the task included two check questions that required reading and understanding the instructions. Payment of $0.50 was unconditional of participants' answers to the check questions. In the following, those who answered both check questions correctly are considered as having understood the instructions including the optionality of the form fields. The other participants who answered at least one check question incorrectly or neither are classified as not having read and/or understood the instructions.

Eight types of warnings were used in the study. They can be divided into two groups: (1) four manual warnings varying the message and (2) four one-click warnings varying the options on the buttons. Table 6.2 provides an overview of the warnings used in all eight conditions. The first warning (WS) mentioned security, while the second privacy (WP) as the explanation for why it was being shown. The warning in the WNX condition (no explanation) did not provide an explanation and only stated:

"You have completed some non-mandatory fields". The warning in the WO condition (o as in 'orthography') served as a control: it stated that there may be spelling mistakes in some fields and gave the explanation that it can be difficult for others to read a text that contains spelling mistakes.

A control condition was chosen that displayed a warning instead of removing the intervention altogether. The rationale behind it was to confront participants with the stimulus warning to see if it impacted their altering behaviour. The idea is to benchmark against the absence of a privacy/security warning rather than the absence of a warning at all. The baseline of undisturbed disclosure rates is available by observing participants' behaviour in the preceding study or in this study, before participants saw the warning.

A spelling warning is well-suited for the following reasons. First, spelling errors are a real and credible 'threat'. They are potentially embarrassing but are not related to security or privacy. Second, it is a plausible pop-up since spelling is something where tool support is already widely implemented. Third, manual investigation is needed from the user: similarly to the security and privacy warnings, the participants needed to go back and inspect each field. The warning in the control condition was structured in the same way as in the conditions WS and WP, they all provided a diagnosis and an explanation, totalling three lines of text.

For the one-click warnings, the same warning text was used as in the security warning (condition WS). Again, there were two buttons for each warning. The first one always said "Submit anyway" while the second one was subject to our manipulation. The four phrases used for the second button can be subdivided into two categories: highlighting and manipulating. For highlighting, "Highlight filled optional" was juxtaposed with "Highlight all optional". A screenshot of what the highlighting looked like in condition WS1HO can be seen in Appendix D. For the manipulating buttons, "Submit only mandatory" was juxtaposed with "Clear optional fields". When confronted with one of the warnings, participants had to make a choice, they could not proceed unless they clicked on either of these two buttons.

Why use warnings? Although active warnings are (at least initially) more effective in alerting users and making them change their course of action, there is also a cost associated with them. For instance, research has shown that most certificate warnings

condition	warning
WS	*Security warning:* You have completed some non-mandatory fields. Revealing too much about yourself online can be a threat to the security of your personal information.
WP	*Privacy warning:* You have completed some non-mandatory fields. Revealing too much about yourself online can be a threat to the privacy of your personal information.
WNX	*Warning (no explanation):* You have completed some non-mandatory fields.
WO	*Orthography warning (control condition):* There may be spelling mistakes in some fields. It can be difficult for other people to read a text with spelling mistakes.
WS1HO	*Security warning: highlights all optional.* highlight optional fields
WS1HFO	*Security warning: highlights filled optional.* highlight optional fields
WS1SM	*Security warning: submits only mandatory.* submit only mandatory
WS1CO	*Security warning: clears optional.* clear optional fields

Table 6.2: Overview of experimental conditions: over-disclosure warnings used in the study.

appear to be false positives [112]. Users heeding a certificate warning would not benefit from added security as there was no threat in the first place. But more importantly, heeding security advice often means users are prevented from completing their primary task. For instance, checking URLs is time-consuming, while attacks are rare. For that reason, Herley stresses that ignoring security advice is rational from an economic point of view [112] as human memory and attention are a finite resource [113].

The use of warnings can also be desensitising since studies have shown that users are habituated to warnings and ignore them. Krol et al. [141] showed that users dismiss warnings if they have the impression that they appear indiscriminately for all the items they attempt to download.

Despite this overwhelming evidence against warnings, it was decided to implement over-disclosure warnings as active pop-up windows that interrupt the primary task since otherwise they might not have been noticed by the participants. The evi-

dence from studies on security indicators shows that passive warnings are less likely to attract users' attention. This represents a 'fail early' approach – to know whether an over-disclosure indicator works, it is better to test it in its active form that would be noticed. If it is effective and receives user acceptance, then it should be explored if and in what form it could work as a passive indicator.

Furthermore, this study is meant to map out and explore the design space of different warning parameters. Importantly, the differential assessment of wording effectiveness and the acceptance rate of different privacy remedies suggested by the warning, are independent of whether a warning is passive or active.

Success metric. Security warnings may be dismissed because they are seen as a false alarm but an accurate security warning applies to all users. For privacy threats, the situation is rather different: even an accurate privacy warning may still be dismissible depending on the user's specific privacy preferences. In other words, security is a 'yes or no', privacy is a spectrum. While in security users cannot partially get infected with a virus, in privacy they can disclose some personal information but withhold other. The users' risk analysis must be more nuanced and the warning design has to have a built-in feature allowing them not to follow the warning's privacy-enhancing advice.

Our study therefore adopts a success metric that does not presuppose universal privacy preferences. Warnings are considered successful not only if participants deleted previously entered information. Instead, this research adopts two different criteria. (1) The warnings are considered effective if users return to the form and revisit the information entered. The warning's actionability is assessed by the click-through rates generated by the privacy-enhancing response option on the warning. (2) Also the post-experimental ratings for usability are analysed to establish if users considered the warnings helpful, easy to understand and relevant.

Follow-up questionnaire

At least one day after completing the experiment, participants were sent an invitation to fill out a feedback questionnaire for which they received an additional payment of $1.50. A reminder was sent to those who had not completed the questionnaire.

The questionnaire started by reminding the participant of the original webform with a screenshot. There was no option to enlarge it so participants could not read the original questions or instructions. The aim was to make participants recall the

original experiment rather than base their answers on what they saw. The questions in the questionnaire can be divided into four main groups. (1) The first nine questions related to the survey asking if the participant enjoyed the original task of completing the webform, how long they spent filling it out, if they had revealed any personal or sensitive data in it, and what they thought the intended purpose of the original form was. (2) Further questions related to computer and security experience. Within that, the first one aimed to gauge the participant's computer literacy and consisted of 14 sub-questions. The second one focused on cyberthreat exposure and contained eight sub-questions. (3) There were 10 questions relating to the warning. (4) Finally, four questions asked for demographic information.

The follow-up questionnaire was completed by 87% of the original participant population (response rate). For 3,203 participants, there were the complete records of form filling behaviour plus feedback data. For a random subset of the participants, the link could not be established with the first phase.

Participants

Participants were workers on the mTurk platform, all recruited to be based in the United States. According to the data provided in the follow-up questionnaire, the majority of the participants were aged 18 to 24 years (36%), 25 to 29 (23%) and 30 to 39 (21%). 16% of participants were aged 40 or older; less than 4% refused to indicate their age bracket. Male participants made up 53% of the sample. Regarding formal education, 28% had completed one or more years of college and an additional 7% had completed their college with an associate degree (for example, AA, AS). 30% had completed a bachelor's degree (for example, BA, AB, BS), and additional 7% a master's degree. For 16% of participants, English was not the language spoken at home.

6.3.3 Experiment results

General results

Findings from the preceding study as to the level of disclosure were corroborated. Amongst participants who were aware that none of these details were required, all personal details were disclosed by at least three quarters of participants, with the exception of date of birth, which was still revealed by more than two thirds of participants. Favourite colour was the data item volunteered most often (84%). First name occupied

rank five when ordering data items by disclosure rate; in the subsequent analysis date of birth, first name and favourite colour are considered as representative for data items of high, medium and low sensitivity, in line with extant literature that takes disclosure rates as a proxy for sensitivity [191]. Date of birth was also the field left blank most often (19%), well ahead of spending, which was the field left blank second-most often.

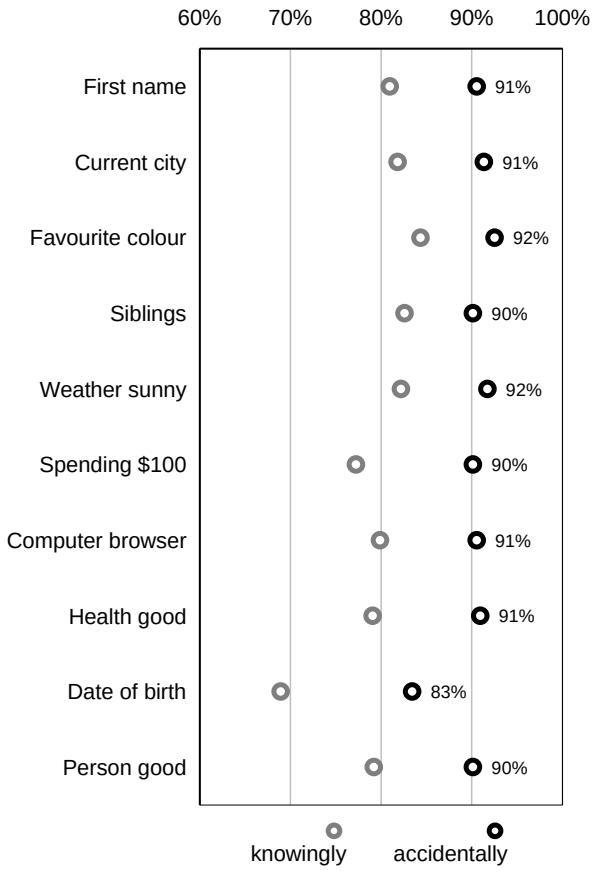

Figure 6.6: Initial disclosure rates for each of the ten items of personal information collected on the form. Knowing disclosure applies to those participants who had read the instructions and correctly answered the check questions; accidental disclosure applies to those who had not.

Interestingly, participants who had not read the instructions and who were thus unaware that disclosure was voluntary, shared more personal details with high statistical difference ($p < 0.0001$, two-tailed t-test). The average number of fields completed by participants who had not read the instructions was 9.1 and thus 1.0 items more on average. Amongst the unaware, 80% fully filled the form with all ten items of personal data compared to 64% amongst those who had read the instructions. Similarly, a blank form was submitted less often by the former than by the latter (6% versus 10%). Consequently, disclosure rates are much higher when participants ignore the optionality of

the data requirements, with all of them being above 90% except date of birth with 83% (Fig. 6.6).

Hypotheses

For simplification, the following analysis is mainly based on three data items to cover the whole sensitivity spectrum. The sensitivity of these items was established based on three sources. Exogenously based on the results from two different, previously conducted experiments which requested them and then asked participants to rate them in the post-experiment questionnaires [193, 162] as well as endogenously. Endogenously, the initial (i.e., pre-warning) completion rates are considered from this study, as outlined earlier in this section. Based on these, date of birth (DOB) was identified as the high-sensitivity data item as it had the highest proportion (19%) of participants who left it blank (69% completion rate, lowest), first name as the medium-sensitivity item (81% completion rate, ranked fifth out of 10) and favourite colour as the low-sensitivity one (84% completion rate, highest).

H1a: Mentioning privacy and security. It was hypothesised that warnings evoking privacy and security threats will make users delete previously entered data. When comparing the deletion ratios for the security and privacy warnings with the control condition (WO), for example for date of birth: the deletion ratio was 26% for WS, 25% for WP but only 4% for WO. The comparison for the three items of differing sensitivity showed that warnings evoking security and privacy threats made users delete previously entered data ($p < 0.0001$ at least for all items, G-test of independence). H1a is therefore supported.

H1b: Security vs. privacy. It was hypothesised that those confronted with a warning mentioning security will be more likely to delete previously entered data than those confronted with a warning mentioning privacy. For the security warning (WS), the deletion ratio was 27% for DOB, 26% for first name, and 14% for favourite colour. In the post-experiment questionnaire, participants indicated the security warning as being more severe than the privacy warning (rated 4.6 versus 4.4 out of 7, $p = 0.03$, two-tailed t-test). The distribution of severity ratings is shown in Figure 6.8. For the privacy warning (WS), deletion rates were very similar (Fig. 6.7). It was hypothesised that the security warning could have a stronger impact on the users than the privacy one. However, this difference was not statistically significant for any of the three levels

	First name	Current city	Favourite colour	Siblings	Weather sunny	Spending $100	Computer browser	Health good	Date of birth	Person good
ws	26%	31%	14%	16%	10%	14%	13%	11%	27%	10%
wp	25%	29%	14%	17%	10%	14%	12%	10%	27%	9%
wnx	12%	10%	6%	6%	6%	6%	7%	6%	12%	6%
wo	4%	5%	1%	2%	1%	1%	1%	0%	3%	1%

Figure 6.7: Deletion ratios per data item and condition after the warning was displayed.

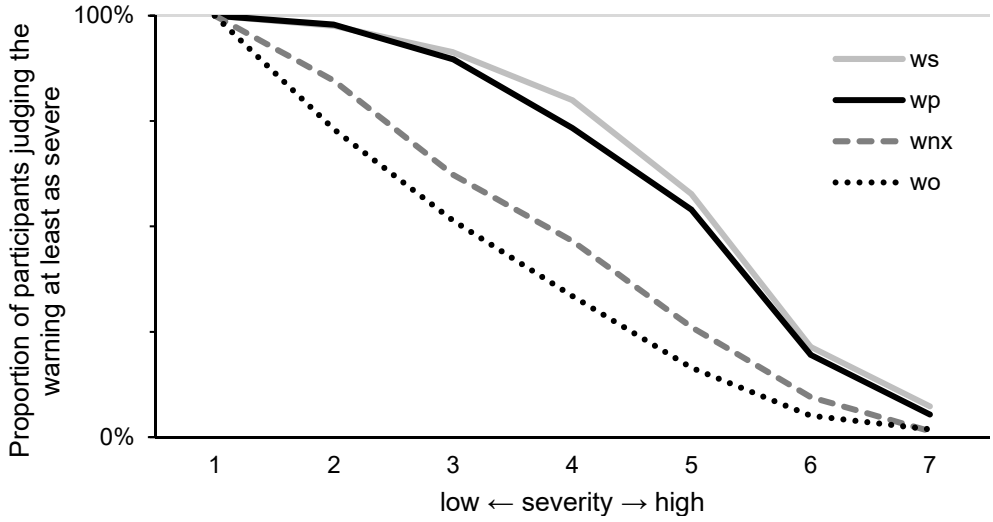

Figure 6.8: Perceived severity of the warning by condition, as rated by the participants in the follow-up questionnaire. The graph shows the proportion of participants who perceived the corresponding warning at least as severe.

of sensitivity: high ($p = 0.92$), medium ($p = 0.50$) and low ($p = 0.78$). H1b is therefore rejected.

H1c: Providing no explanation. It was hypothesised that a warning providing no explanation will have a lower deletion ratio than warnings giving an explanation. A comparison of WS and WNX showed that the proportion was more than double across all levels of item sensitivity. There were significant differences across all items (low: $p = 0.00001$; medium: $p < 0.00001$; high: $p < 0.00001$). H1c is therefore supported. Additionally, the warning that provided no explanation was rated as the least usable in the follow-up questionnaire, achieving an even lower score than the orthography one.

Although the analyses of results are based on participants who had read and understood the instructions, it is interesting to assess post-warning deletion patterns for

those who had not read the instructions. In the combined WS and WP conditions, deletion ratios are significantly lower amongst those who had not read the instructions (for date of birth: $p = 0.002$; for first name: $p < 0.001$; Fisher's exact test; n.s. for favourite colour). It could be argued that this might reflect overall lower engagement with the task at hand. Not only did participants not invest the time to read the instructions, but they also wanted to submit the form as quickly as possible, thereby not spending time on re-visiting their entries.

In the WNX condition, the difference in deletion ratios between those who had read the instructions and those who had not is similarly pronounced as for the privacy/security warnings that featured an explanation. The WNX warning stimulated a deletion action only for current city and date of birth.

H2: Item sensitivity. It was hypothesised that more sensitive items will be removed more often, forming a U-shaped relationship. Data sensitivity ratings were taken exogenously from a study by Malheiros et al. [162], where participants indicated their willingness to disclose a variety of data items online. Ratings were recorded on a four-point Likert scale (1 – happy to provide, 4 – unhappy to provide). The data items asked on the webform were mapped on those for which data item sensitivity ratings were available (e.g., 'good' was mapped to 'health'). The relationship was then approximated by a polynomial trend (Fig. 6.9). Highest deletion ratios were observed for items with medium willingness to disclose; low deletion ratios were observed for items with low or high willingness to disclose, such as favourite colour or health respectively. A parabolic fit explains a high 73% of the variance ($R^2 = 0.7317$). H2 is therefore supported.

H3a: Sunk cost fallacy. It was hypothesised that in line with the sunk cost fallacy the more time a participant spent on filling out the form, the less they will be likely to remove the information entered. On average, participants saw the warning after 86 seconds on the form (range between 82 seconds for WS1HO and 91 seconds for WO). As expected, there is no statistically significant difference across the conditions (two-tailed t-test: $p = 0.12$). The data across all conditions are therefore considered. Participants who took more time completing the form were more likely to delete information than those who had spent less time on the form ($p = 0.02$, G test of independence). This is the exact opposite of what was hypothesised, H3a is therefore rejected.

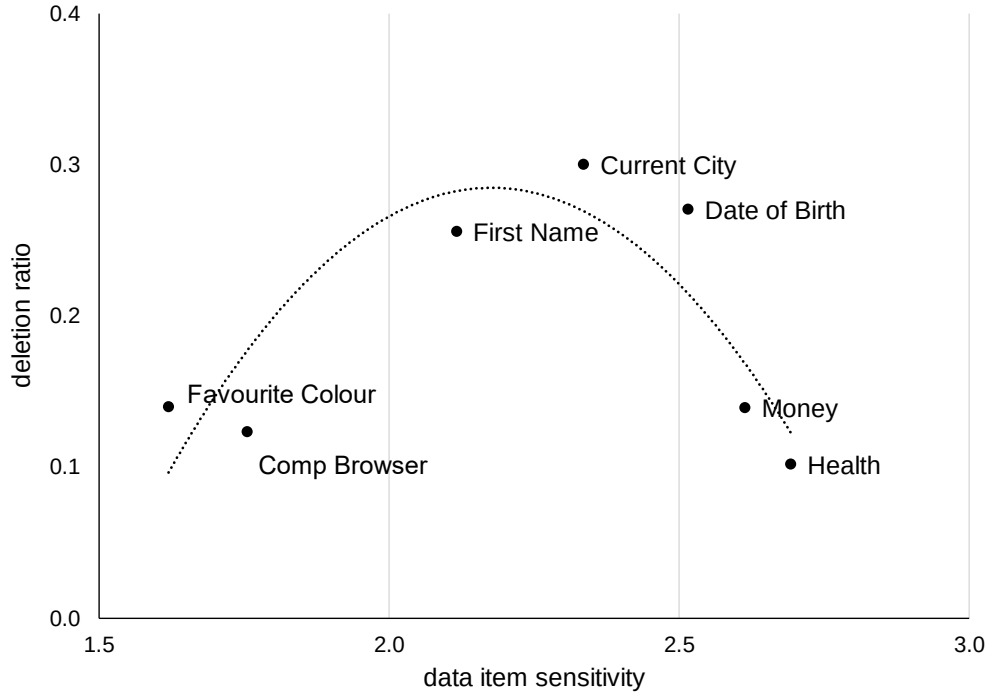

Figure 6.9: A parabolic (U-shaped) relationship between item sensitivity and deletion ratio.

H3b: Options minimising effort. It was hypothesised that the options that minimise user effort and save time will be preferred. From the warnings that varied the options on the buttons, the one with the highest click-through rate was 'highlight optional' (53%), 'followed by clear optional' (43%) and 'submit only mandatory' (34%). This came as a surprise since it was hypothesised that the warning that would minimise user effort would be preferred. H3b is therefore rejected.

H4a: Computer literacy. It was hypothesised that participants will differ in their deletion/altering behaviour depending on their levels of computer literacy. Computer literacy was measured with an eight-item question battery (reliability measured by Cronbach's alpha: $\alpha = 0.69$). Participants were asked whether they had ever performed tasks indicative of elevated computer literacy. This included changes to computer configurations, some which were security-related (various browser settings, firewall configuration), registered a domain or designed a website, or unscrewed anything on their laptop or PC. Participants generally showed a high level of computer literacy, with the lowest skill being ever having written some computer programme at 33%. As many as 95% respectively 92% of participants had changed the homepage or the default search engine in their browser. In the context of the study (webform completion behaviour and usage of browser-provided privacy-enhancing technologies), these are very high

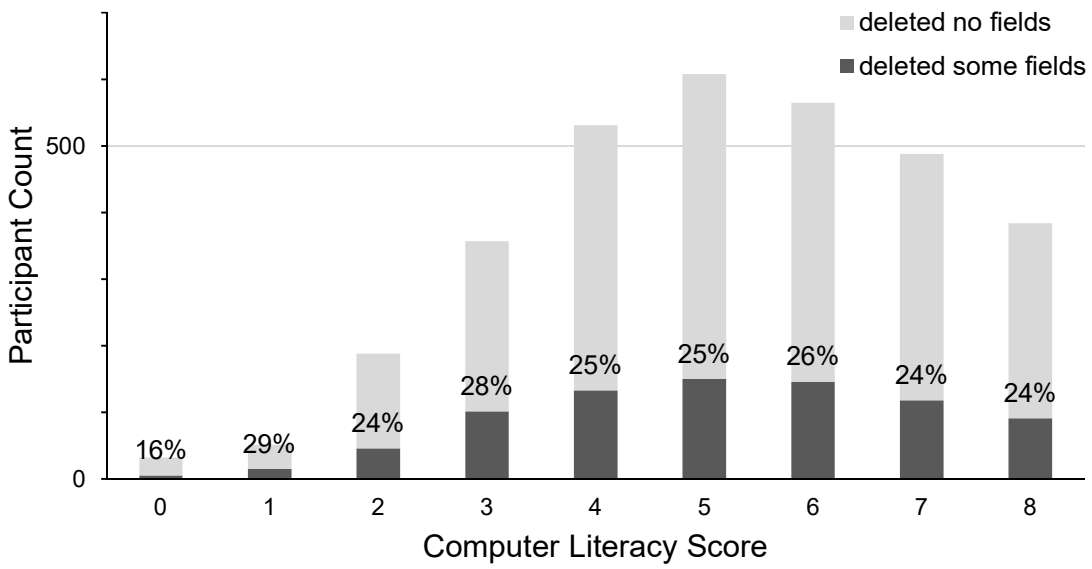

Figure 6.10: Deletion prevalence by computer literacy score.

values. Fewer than 1% of participants did not know whether they had done so, also indicating high literacy about these computer concepts.

Deletion behaviour was distributed uniformly over varying levels of computer literacy as presented in Figure 6.10. There was no statistically significant relationship between participants' deletion behaviour and their computer literacy. Hypothesis H4a is therefore rejected.

H4b: Cyberthreat exposure. It was hypothesised that participants will differ in their deletion/altering behaviour depending on their cyberthreat exposure. The term 'cyberthreat exposure' was adopted from Sunshine et al. [247], but expanded since their original item battery only contained four questions. A 14-item battery of yes/no questions (reliability measured by Cronbach's alpha: $\alpha = 0.70$) was used. Questions included: whether participants had ever been victims of cybercrime (e.g., hacked email or other online accounts), Internet-mediated fraud (e.g., phishing, spam), banking fraud (e.g., fraudulent transactions) and privacy crime in particular (e.g., data breaches, identity theft). As with computer literacy, there were two check questions here (after questions 4 and 9) to ensure the participants were paying attention. The individual scores were calculated for each participant by giving each affirmative answer the score of 1 and summing these. Figure 6.11 shows the results. When mapped on deletion behaviour, there was no clear trend there. Hypothesis H4b is therefore rejected.

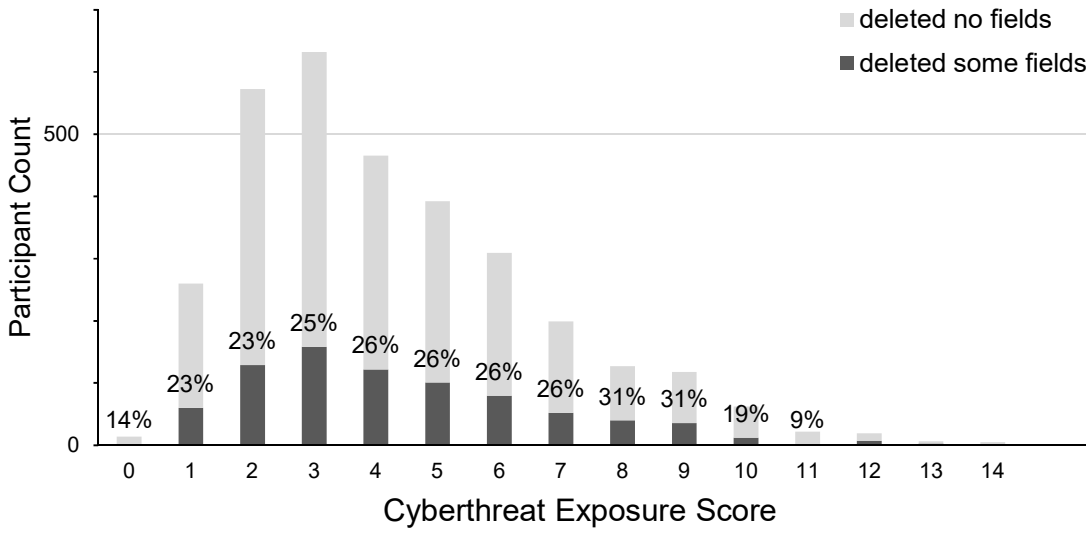

Figure 6.11: Deletion prevalence by cyberthreat exposure score.

6.3.4 Questionnaire results

Reasons for disclosure

Based on their disclosure in the experiment, participants were asked in the follow-up questionnaire for the reasons why they provided or did not provide their date of birth (DOB). Since these were free-text responses, participants were able to provide as many reasons as they wanted (or none at all) and there was no length restriction for their answers.

Reasons for disclosing DOB. Amongst those who provided their DOB, the most common reason for why they did so was that they did not mind (308 out of 1,612 responses, 19%). P76 explained: *"I did not mind revealing that information."* This was followed by participants saying they did not see any harm in providing it (297, 18%). Apart from that, 17% stated they did so to help research in some way. Also the sensitivity of the DOB was thematised by participants; 15% of participants stated the DOB was not a sensitive or personal piece of information whereas 16% that it was a sensitive piece of information but they gave another reason why it was fine to disclose it this time. For instance, P15 stated: *"While its sensitive information, I don't feel too concerned about sharing it."* Also, 10% stated they thought providing their DOB would bring some benefits with 87 (5%) emphasising a benefit to themselves, P397 stated: *"If i give you my date of birth, You may use it for helping me"*. Interestingly, 159 (10%) participants filled in their DOB to satisfy their completionist instinct, P91 explained:

"It would have bothered me more to have everything filled in but that one." Further, 130 (8%) participants argued that they their DOB was already out there so not providing it on the form would not have made much difference: *"Date of birth is easy to obtain through a simple google search, more information is typically available on the average users facebook page"* (P879). P118 also assessed the probability of something bad happening as low: *"My date of birth it public information that can be easily found. The likelihood of a negative outcome from giving this information was low."* Participants also reported on their coping strategies, 5% of participants stated they provided an incomplete DOB (e.g., only year or only day and month). Also time and effort were thematised, one answer was particularly interesting as it shows that users are attempting to economise their effort, P318 said: *"it took about as long to enter it as it would have to think about whether or not i wanted to enter it."* Finally, 43 (3%) participants stated it had become their habit to enter their DOB.

Reasons for not disclosing DOB. When asked why participants did not provide their date of birth, two responses dominated: 34% (693 out of 2,060) of participants stressed their DOB was a sensitive or personal piece of information and 32% emphasised the DOB was an optional form field. Further, 11% participants said that DOB can be used to identify them. Interestingly, 4% of participants mentioned that the DOB could help identify them when combined with other information they provided on the form or that someone could find on the Internet, P96 explained: *"I thought you might be able to use that, along with my first name, and my city to search other databases and get more information about me."* Also, 5% of participants stressed providing their DOB could harm them in some way; for example, P3585 stated: *"It can be used to steal my identity"*. Interestingly, 169 (8%) participants stated they heeded the warning that told them to limit disclosure. Also, 6% stressed they did not see the reason why their DOB was needed in this situation and 2% said they would have had no problem providing their age but not their full DOB.

Warnings have a bad reputation

There were several participants who expressed gratitude for being able to participate in the study saying it taught them a lesson: "*its to make sure to let you know that you don't have to put in all you info. i wish more websites had notifications like that*" (P2658). However, participants expressed negative attitudes towards warnings and were disillu-

sioned about the reasons why the industry used them. Overall, 12 participants in the study interpreted the warning as a means to transfer liability for keeping the data secure from the requester or mTurk onto the user. P719 stated the purpose of the warning was "*To try to protect the liability of the company or website it was displayed on*" and P2090 thought the warning was "*[a] disclaimer that the information [they] provided will be at [their] own risk.*"

There were three participants for whom the control condition, the orthography warning, did not work. P1525 described what they thought the purpose of the warning was by saying: "*It was a verification request to make certain I wanted to submit the form in the previous survey. It was slightly funny in that it pointed out the reasoning was because of spelling errors and not because of security issues.*" This shows that pop-up warnings have a strong connotation with security advice. Similarly, P1735 stated: "*I had never seen this before on turk. I thought it was strange. I felt like it meant that I hadn't filled out all of the fields, and so I should go back and check them.*"

Possibility of tapping into existing knowledge

Some participants also read things into the warning that were not included there. Without being prompted, 354 (9%) participants mentioned some type of crime that could happen as a consequence of over-disclosure, the most common was identity theft. P116 stated: "*any information you share online could be used against you; even information that is vaguely identifying could lead to potential theft of your identity*." The fact that US population seems to be particularly aware of identity theft could be used in the future to illustrate possible consequences of over-disclosure in the warning text.

6.3.5 Discussion

Warning text

The study found that warnings can be effective in reducing users' privacy exposure resulting from excessive disclosure of personal details on webforms. It is acknowledged that warnings, as a user experience principle, suffer from fundamental flaws as discussed in Section 6.3.2 but these would not matter for the research question at hand.

Warnings that evoke a privacy or security threat make users take privacy-protective measures significantly more often than an arbitrary warning not related to the threats of disclosing personal data. When comparing the conditions where users

were warned about over-disclosure with the condition notifying users about spelling mistakes, the former resulted in users going back to the form and removing previously entered details with much higher prevalence. Therefore, the fact that users disclosed fewer optional data items is not due to just any warning, but specifically due to the privacy and security warnings containing an explanation.

The results show that it made no difference whether the threat evoked was negative privacy or security consequences. Participants were nearly equally likely to delete previously entered information after having seen either of the warnings. However, providing an explanation increased the deletion ratio. This corroborates the findings by Egelman et al. [67] who found that participants were more tolerant of security delays if the threat was explained to them. Similarly here, when there was no explanation provided, participants rated the warning as less severe and less usable. The explanation is an aid for participants to make an informed decision about the real threat. They are thus more able to choose an action that better reflects their privacy preferences. One can argue such explanations are important for users even when the privacy indicator would not be deployed as an active pop-up warning.

Item sensitivity

Data sensitivity is a traditional operationalisation of privacy threats. The law enumerates a list of particularly sensitive data items and it is corporate practice that certain items require protection above and beyond the safeguards implemented for other items of personally identifiable information. However, data sensitivity does not seem to be a decisive factor for users' behaviour regarding disclosing or withholding their personal details; its influence is less significant than that of perceived relevance and fairness [162]. At the same time, disclosure rates were previously identified as a way to measure privacy concerns and those varied by data item sensitivity [191].

Users' deletion behaviours were found to vary significantly by data item sensitivity, but the relationship is not monotonous. One the one hand, users do not delete low sensitivity items they had previously entered, as the privacy threat may not exceed their desire to disclose. On the other hand, high sensitivity items cannot be deleted because users left those fields blank in the first place, knowing their optionality. Consequently, medium sensitivity items are deleted most often.

Contributions of the study

Material results. This study explored the uncharted territory of over-disclosure indicators and found that warnings can help users manage the disclosure of optional details through webforms. The main lessons learned include: the mention of privacy/security threats makes a privacy warning effective and an explanation of the threats associated with disclosure further heighten users' propensity to remove previously entered personal details. Actionable warnings are characterised by keeping the user in control as to which items they would like to remove from a webform: optimisation for reduced manual effort leads to lower acceptance than giving users the choice about which optional fields to submit and which ones to erase.

Methodological contributions. The study stands out from previous research into privacy by (i) being a true experiment (ii) in a naturalistic environment (iii) tested on a large number of participants (iv) using fully functional warnings. First, there have been numerous studies in both security and privacy showing that there is a discrepancy between users' reported and actual privacy and security choices (e.g., [24, 141]). Users' actual reactions were studied here: participants were not asked about their appreciation of privacy-enhancing tool support but the warnings were trialled under scrutiny and objective metrics were adopted to assess their effectiveness. Participants took part in the study in their homes as opposed to in a lab. If a study is conducted in a lab, participants might feel less comfortable and there might be a trust relationship between the experimenters and the participants. An online study at least partially removes such bias (there is still a potential impact of the requester's mTurk reputation). Third, each warning was tested on hundreds of participants which would not have been feasible if the study was lab-based. Fourth, in exploratory studies, the warnings or designs used are sometimes mock-ups which can give the participants the impression the interaction is not real. In the study presented here, the warnings were fully functional, they did what they said they would. For instance, if the 'clear' or the 'highlight' buttons were pressed, form fields were actually erased or flagged.

But aren't warnings bad? Previous research has shown that warnings are effortful and users are conditioned to ignore them. Feedback data from the study confirmed that users have negative associations with warnings. For these reasons, an over-disclosure warning should ideally be implemented as a browser plug-in that would highlight

mandatory and optional fields for the user while a webform loads on a website. In this way, one could achieve visible but non-disruptive warnings – users would not need to click them away to continue with their primary task. However, unlike previous passive indicators discussed in Section 2.5.4, they would not be placed on the periphery of the screen but in the centre of the user's visual attention. The plug-in could also be configurable and the user could pre-define what kind of information they are comfortable sharing with what type of website. This would pave the way for a context-aware auto-fill feature. The results of the study show that a more paternalistic approach, such as suppressing optional form fields, needs to be assessed carefully given users' quest for control.

Further research. Future research could juxtapose privacy indicators of over-disclosure over a longer period of time to measure how they would impact users' actions. Only such panel studies can assess habituation effects, counter-balance novelty biases and observe reactions across contexts. The development of robust success metrics will be key in guiding the development of such privacy-enhancing tool support. A long-running study would allow privacy indicators to become part of the web browsing routine, making it possible to investigate appropriation and habituation. The indicators would need to be implemented in real life which is not difficult as it used to be. HTML5 introduced advanced webform mark-up capabilities that include an attribute for form fields to indicate whether it is optional or required [264]. In addition, the browser could examine and interpret existing visual cues. Browsers have access to a combination of signals to sense the mandatoriness of fields and advise the user accordingly.

6.4 Conclusions

This chapter presented the results of two studies investigating user disclosure behaviours when completing webforms. The received wisdom is that filling in webforms is a nuisance; nevertheless, both studies presented in this chapter showed that users over-disclose at the expense of time, cognitive and physical effort. The first study varied the mandatoriness and incentives to assess their impact on disclosure behaviour. Once some fields were marked as mandatory, disclosure dropped for the remaining optional fields. Once there was a bonus for some fields, disclosure increased for the

remaining unincentivised fields. The message from both conditions is that users adapt their behaviour if there is some indicator as to what is expected of them. If there is none, they tend to just fill in all the blanks. Participants' over-disclosure could be attributed to the fact that it was easier just to fill in all the information than to try and submit a form with only some fields and then have it rejected. It could be a way of participants dealing with uncertainty that they completed everything "just in case". The second study addressed the fact that most webforms online do not provide mandatoriness indicators and it tested eight types of over-disclosure warnings. Users preferred warnings that provided an explanation and preferred options that left them in control even at the expense of effort.

6.4.1 Limitations

Both studies described in this chapter might suffer from the following limitations: (1) they observed one-off reactions, (2) the sample of participants might not have been representative, (3) the primary task was weak, (4) there might have been a trust bias and (5) participants might have falsified their webform entries.

First, both experiments focused on only one reaction to a webform or one instance of a warning and this might not be representative of the participants' general behaviour in varying situations. A longitudinal study to examine repeated reactions to the same warning is therefore necessary.

Second, both studies might have suffered from a sampling bias. By deploying on mTurk, the studies could have had a bias towards people who are used to filling in forms. The skills and the mindset of mTurk workers may be such that they are trained to complete forms quickly. However, when optimising for speed, uniformity and thereby over-disclosure may be more desirable than time-consuming selective disclosure. However, compliance with instructions was carefully checked and those participants who had not passed the check questions were removed from the analysis. The mTurk platform does not provide access to participant statistics, such as the number of previously completed tasks for each worker, which could have been a moderating variable.

When it comes to the representativeness of the sample, there have been studies that showed that the mTurk population is diverse across several demographic dimensions such as gender, age and income [210]. One might argue that workers on

the mTurk platform may deliberately over-disclose to increase their chances of future working opportunities. In this study, participants might have believed the webform was a pre-screen for future tasks. In the exit questionnaire, participants were asked about their motivation and there was only limited evidence for signalling behaviour. However, one should note that the quest for future tasks can hardly explain over-disclosure at this level of detail. Also, most mTurk workers are trained for compliance with instructions rather than for volunteering personal data.

Third, since security and privacy are secondary tasks that users encounter while performing a primary task online [220], a study has to mimic this to produce ecologically valid results. Although there was no explicit cover story provided for collecting the data, it is likely that the participants considered the form to be a pre-screen for future tasks. One could argue that the lack of an explicit primary task could have made participants more suspicious or on the contrary, more inclined to volunteer their details.

Fourth, the studies might have been biased by participants trusting a university as a requester more than they would trust a commercial entity. However, only a minority of participants named trust as a driver for participation (4% in the first study). It is important to note that well-known retailers or social networks may benefit from similar trust biases resulting from brand effects.

Fifth, it might be argued that the participants in the studies had no motivation to provide correct information and this could invalidate the results. Due to the deployment specifics of the experiments, it was not possible to perform data verification. However, commercial websites are similarly handicapped in their ability to test the accuracy of users' personal information, but the motives for voluntary over-disclosure work against lying. Also, misreporting only partially affects the effort of over-disclosure: the typing effort for an answer is independent of its truthfulness. Also in the exit questionnaire, participants were asked what they thought the purpose for collecting this information was and 287 (7%) answered it might be to pre-screen them for future tasks. Thus, the more accurate the information, the more likely they would be invited to the correct future tasks.

Chapter 7

Discussion

7.1 Summary of work

The six studies presented in this doctoral dissertation examined security and privacy behaviours through the lens of human effort. Two studies using authentication diaries and interviews aimed to capture the effort required to authenticate in an organisation and for online banking respectively. Further two studies examined the usability, effort and user acceptance of proposed mechanisms for authentication and verification. Finally, the last two studies investigated how users disclose information about themselves through webforms, with the second study assessing the impact of warnings on user disclosure behaviours.

7.2 How do users manage their effort?

Both authentication diary studies documented the various coping strategies participants employed to manage their authentication effort. These can be divided into facing authentication and avoiding it where avoidance can be permanent or temporary, depending on if the user postpones authenticating or removes the need to authenticate entirely. Postponing authentication means that users move the effort to a moment where it is more convenient to take the time, recall (or look up) their password and log in. This often happens when the authentication effort might interfere with the user's work as physical and cognitive effort might take the user's mind off their primary task. An important lesson here is that not all logins are equally effortful as some users backgrounded logins they performed from muscle memory but foregrounded ones that were effortful, disruptive and interfered with their work.

7.2.1 Habituation and familiarity

Both authentication diary studies showed that participants had become used to security, had developed coping strategies and facing the demands of security tasks was part of their daily routine. Participants reported being able to enter some of the passwords from muscle memory habitually without even noticing it. But it was not the case for 2FA where participants stressed they could not automate that process because they had to generate an OTP on each login. For completing webforms, a proportion of participants stressed they were used to filling in all the blanks and they did so in the study "without thinking".

Habituation is a response of the human brain to deal with environmental signals, it helps diminish cognitive and physical effort through automation. But in order for a security task to become automated, the user has to adopt a process or a mechanism and use it. The authentication in an organisation study showed that sometimes the user would simply give up or avoid the use of a technology if they consider it too much effort. The user must be motivated to give a mechanism a try, learn how to operate, not forget this skill between uses and gradually get into the habit of using it. This is a hurdle some security mechanisms might not overcome. Humans are goal-oriented, they want to "get their work done" and not to second guess what the designer or developer meant or apply thinking to figure out how to use technologies [149]. This holds true especially when there already is a competing mechanism the users are familiar with that might lead them to achieve the same goal. Switching to a new technology is always costly and users prefer the familiar.

Habituation is a natural evolutionary response however research studies have seen it as the enemy of security [10]. The advocated approach is that users should try harder to focus on warnings or the designers should make warnings harder to ignore. This approach has been called by Sasse "scaring and bullying people into security" [216]. Instead, systems and warning designers should focus on making indicators work for users taking their capabilities and natural responses into account.

Closely related to habituation is familiarity; relying on the familiar is a way for the user to save effort. From a decision making point of view the recognition heuristic works here in that users choose what appears familiar to them [92]. However, the danger is that the cues that users rely on might not be valid in the online world as users

do not realise how easy it is to imitate trust indicators [222] or the 'look and feel' of a reputable website [280].

7.2.2 Cognitive and physical effort

The studies showed that participants were bothered by both cognitive and physical effort. The problem of physical effort will become even more pronounced in the future as devices are becoming more ubiquitous and input methods more diverse. In the past, most devices just had a mouse and a physical keyboard, nowadays touchscreens are becoming more and more prevalent and virtual keyboards have usability flaws [99]. At the same time, the userbase is becoming more and more diverse including users with different motor and cognitive capabilities. One solution is that password entry or verification methods could be optimised for devices and/or user groups. The evidence from the studies presented here shows that participants were faster authenticating on a tablet's touchscreen using the VAG but slower verifying using NoBot.

7.3 The need for control

Throughout all the studies described in this dissertation, participants to some degree expressed the need to be in control of their tasks and security as part of these tasks. The coping strategies discussed above are a manifestation of users' desire to exercise control over their authentication effort. Participants adopted various coping strategies in order to move the effort (be it physical or cognitive) to where it was least disruptive to their primary task. They planned and batched activities to be in control of how they went about their work.

In the studies that evaluated verification mechanisms, some participants felt that using NoBot left them outside of control. Deciphering and entering reCAPTCHA text gave participants the sense of having more influence on the outcome of the verification than letting NoBot verify them with only a biometric input from them. Participants preferred to actively prove they were human in order to have a feeling that this was done correctly and they are not at the mercy of a technology that can lack accuracy.

Users' desire for control has been documented by several studies in usable security and privacy. Ruoti et al. [213] studied participants' use of different email encryption software that varied in how much of the underlying encryption process was visible to the user. Interestingly, participants preferred the system that required more effort as

encryption was done manually, they also made fewer mistakes and had more trust in the system. Although the researchers did not directly address this, the need for control might be behind it as one of their participants stated: *"I like that you can see all the steps of what you're doing, so you feel more in control of the process"* [213, p. 19].

Control has also been documented as a factor in users' choice of password managers. Karole et al. [130] studied how comfortable users were with giving up control over their passwords to a password manager. Participants felt they had the greatest control over their passwords when using USBs and their phone rather than an online application. The authors concluded that managing passwords locally on their devices gave participants a feeling of control and authority.

However, the relationship between control and user behaviour has been proven to be complex. Brandimarte et al. [40] investigated users' feeling of control and their willingness to disclose information. They found that the more control their respondents felt, the more likely they were to disclose information about themselves. This might have been due to the effect of a 'risk thermostat' as described by Adams [5] (previously mentioned in Section 2.4.3). If users felt they were protected, they were more likely to share information about themselves.

Also Internet service providers tap into users' desire to exercise control in the way they present their services. Vodafone, for example, advertise *"Family Time lets you control who can use broadband and when. Set up your own family network and control who uses what and when."*[1]

7.4 Views on effort and security

In the studies described in this dissertation, participants often expressed more general views on security that revealed their mental models of how they imagine the systems worked.

7.4.1 "Hard for me, so hard for an attacker!"

In both lab studies evaluating the usability of authentication and verification mechanisms, there was a theme of participants emphasising that security had to be hard to protect what is valuable to them. In the study examining verification mechanisms, some participants emphasised that an easy CAPTCHA gave them the perception that

[1] https://www.vodafone.co.uk/broadband/choices

anyone could access the system to steal their information and money. Others explained that CAPTCHAs had to be hard so that they do not make purchases too quickly.

In line with the fundamental security principle saying that the strength of a security measure should be proportional to the value of the asset it is protecting, the strength of authentication should be proportional to the importance and value of the accesses it is protecting. However, the need for stronger authentication should not mean increasing the burden on users. The old myth that there is a 'usability-security trade-off' leads security experts to assume that it is acceptable for stronger security to require more effort. The myth affected participants in the studies here where some consoled themselves that if the mechanism is demanding, it is secure. However, in the VAG study, the majority of participants were frank that they found the mechanism too demanding for regular authentication, and research to date has shown that authentication mechanisms that create too high a burden are circumvented, avoided or abandoned altogether by users as shown in the diary study examining authentication in an organisation.

Performing security tasks can give users the rewarding feeling that they have contributed to making their online interactions secure, but the effort has to be proportionate. The challenge is to be able to strike the right balance between providing users with sufficient reassurance and demanding their attention, time and effort.

7.4.2 "If there is no other choice, I'll do it"

In the authentication diary study and the VAG study, there were some participants who said, if they really needed the access, they would go through a cumbersome security procedure and get used to it if there were no other alternatives. This is not a sustainable approach since it will exhaust users' compliance budget. Research in organisations has shown that if forced into compliance with security policy, employees will start to resent it and will find ways to circumvent it [20].

7.4.3 Downplaying effort

On several occasions, participants downplayed the effort that they had to expend to use a security mechanism. There was a discrepancy between actual and perceived time needed to verify using NoBot. When filling in a form, participants often overestimated the time that they needed to complete it. Those who filled in more fields, tended to

enjoy the process more. In the evaluation of NoBot, one can speculate that the novelty of this verification mechanism overshadowed the fact that the verification process took longer than for reCAPTCHA and PlayThru. There might have been a similar effect in the second disclosure study examining the impact of warnings. Participants might have been more attentive to the warning since it was novel. If deployed in a real-life setting, this effect might diminish over time.

7.5 Effort for security and privacy

The studies in this dissertation also thematised the differences between security and privacy behaviours. The security tasks in this dissertation – authentication and verification – required explicit effort from users, privacy choices studied here were less explicit. This might reflect the nature of privacy choices since privacy is something that in the physical world happens naturally as disclosure is part of the development of human relationships [146]. It comes more naturally to users as every human assesses the situation and chooses the level of disclosure they deem appropriate in the given context. This often is implicit and requires little effort. This is different in the digital world. Facebook, for example, asks users upfront what type of information they want to share with whom and this is not in line with how humans go about information disclosure in reality. It is a challenge for both research and industry to create interfaces and controls to enable user to set privacy boundaries in ways that would be less disruptive and more natural.

This dissertation worked from the assumption that privacy preferences are diverse and disclosure is mediated by context. Being completely private is neither feasible nor desirable. In this respect, the difference between security and privacy can be described as a difference between vertical and horizontal preferences for products as theorised by economists Hotelling [118] and Lancaster [151]. Security preferences can be described as vertical as better security is considered to be desirable. Privacy preferences are horizontal as individuals moderate how private their information is based on context. To use the words economic theorists use, privacy is something users develop a "taste" for. Research into the economics of privacy has found evidence for horizontal privacy preferences [192, 25] and the fact that users develop "a taste for privacy", using the example of privacy settings on social media [156]. The disclosure studies presented

in this dissertation provide further evidence. They showed that users do not want the strictest privacy by limiting the amount of disclosed information to the mandatory minimum but choose to selectively disclose information about themselves based on different preferences, motivations and expected consequences.

7.6 Ways to minimise user effort

7.6.1 Learning from coping strategies

The long-term goal of studying user coping strategies is to understand what level of effort is acceptable and in what contexts. Some of the coping strategies (e.g., simplifying credentials to avoid resets) resembled workarounds known from previous studies in organisations (e.g., [20, 135]). Although it is beyond the scope of this dissertation to examine this, many of the coping strategies participants employed had the potential to undermine security. It is therefore crucial to learn about them and anticipate their development as new technologies and processes are being introduced. Research has postulated not to block workarounds but to learn from them and better support users in the future [134].

Rituals could be considered a special type of coping strategies, performing certain actions such as hiding a hardware token gave a participant the sense that they were doing the right thing, making them feel protected. Although this again lies beyond the scope of this dissertation, it is undeniable that some of them could be considered less effective in protecting users. One reason for this is that they are based on incorrect mental models, folk models as Wash [267] called them. To 'correct' these models, researchers and practitioners have suggested that users need to study IT security to both know what to do and appreciate the risks accurately. The approach that users need to learn about security is neither feasible nor desirable. Instead, the process or device design together with associated instructions should convey the correct mental models needed to operate them. Researchers have postulated to give users a set of *if-then rules* [1] that would make individuals use security tools effectively as is the case with driving a car.

7.6.2 Consolidating authentication effort: SSO

For the organisation examined in the first authentication diary study (Section 4.2), single sign-on (SSO) could be a way to reduce authentication effort, but as in the study

by Inglesant and Sasse [122], most participants reported that they had credentials for a system on which it had not been implemented. Participants expressed a preference to have one complex and longer password rather than many passwords with different requirements as is currently the case. P11 shared their dream to only have one password: *"I guarantee for me I would sit down and practice that thing and practice that thing and practice that thing until it's automatized and I wouldn't forget it and I would be totally happy to enter a 20-digit password if I could use the same one and not have to go through this hullabaloo of calling and resetting."* Properly implemented SSO could also solve the problem that participants in the study sometimes did not know which username and/or password to use to access which system.

7.6.3 "Technology should be smarter than this!"

In an age where there is so much emphasis on interactions being seamless and sleek and delivering users the best possible experience, users are still bothered with authentication which causes frustration and affects their productivity. Very often users are used to smooth interactions from personal accounts and devices and these are in a stark contrast with stagnated organisational systems. Study participants were knowledgeable about technology and they emphasised that technology should be smarter than this and should use different cues to know the user is still there but for example has not moved their mouse in a while. Also as described in Section 4.3.4, time of day made a difference to participants' authentication experience which could be used to adjust login requirements to user preferences and treat time of day as a contextual cue.

7.6.4 Shifting from explicit to implicit authentication

As cloud computing is becoming more and more powerful, even long and seemingly strong passwords now need less time to be cracked. Also attacks have been increasingly targeting password databases and in this situation, there is little individual users can do to protect themselves [112]. In this situation, companies push for two-factor authentication to achieve better security. It is predominantly tokens, smart-cards or mobile phones that are used to create the second factor. Although the solution may appear plausible, the findings from studies in Chapter 4 show it can be rather error-prone and creates additional work for the user. They have to remember to have the token with them and it can pose usability challenges due to screen size and confusing operating

instructions. In other contexts of use, for example the case of mobile phone-based two-factor authentication, one has to remember to have the phone with them and wait for the credential to arrive. Additional challenges include the phone running out of battery or not having reception. Here, the diary study of 2FA for online baking showed that the more effort was required, the lower participants' satisfaction was.

Because of the shortcomings of the state-of-the-art two-factor authentication mechanisms and in general mechanisms relying on the interrupt-authenticate model [132], a better solution would be to use implicit rather than explicit authentication. Existing work in this area has considered authenticating users based on either (i) them carrying a token or on (ii) behavioural biometrics.

Token-based implicit authentication requires the user to have the token on their person and for it to send a signal to a receiver that they have the right to access a system. Approaches like this have been under development, some use proximity sensors. The Pico project is working on creating a hardware solution, consisting of a main token and associated smaller tokens, to replace passwords, PINs and other knowledge-based credentials. Pico would provide continuous authentication resistant to various types of attacks [239]. Otherwise, a token for implicit authentication has been proposed for environments such as hospitals [18, 17] where time is critical to save human lives and secondary tasks that cost too much time and effort cannot be tolerated.

In turn, behavioural biometrics aim to authenticate the user based on their interactions with devices such as their typing rhythm, mouse movement or voice. For example, Shi et al. [233] explored the possibility of implicit authentication through observations of user behaviour such as their location, patterns of messaging and browsing. Killourhy and Maxion's [133] work on keystroke dynamics showed that users can be recognised by the way they type with high levels of accuracy. Also in the age of touchscreens and gesture-based interactions, systems could authenticate users based on the way in which they gesture or interact with touchscreens as part of their primary task. In this way, we could achieve what Sasse [223, 219] calls *0-effort, 1-step, 2-factor authentication*. The added advantage of behavioural biometrics would be increased security as they would recognise the presence of a human rather than the presence of a technology. In commercial products, Kinect Identity identifies the player based on several cues such as their height and clothing [204].

7.6.5 Accounting for a diverse userbase

In both authentication diary studies, there is evidence that systems were not designed with users in mind and users are asked to expend effort to compensate for the shortcomings of the systems and procedures. This can negatively affect satisfaction as well as organisational and individual productivity.

In the study assessing the user experience of 2FA in online banking, the participant sample consisted of relatively computer literate participants. Given that this group of participants struggled, it is likely that older, less technology-savvy users are put off online banking by its lack of usability which is an accessibility issue. It is also a missed commercial opportunity for banks seeking to lower transaction costs.

Authentication could be tailored to different user groups. Younger participants were faster authenticating using the Vernitski Authentication Grid than older participants. Sreeramareddy et al. [238] tested the usability of gesture-based passwords with younger and older participants and found that senior users entered their passwords with greater accuracy and fewer pauses. These findings show again how users differ and suggest that authentication should be tailored to individuals based on their capabilities and preferences.

7.6.6 Giving smart choice

Given the great diversity of users and their different goals, capabilities and needs, offering more choice of authentication technologies appears to be a plausible option. However, giving choice is not easy due to high administration and maintenance costs. Choice also requires user effort in needing to learn about options and adopt them. Consequently, research has shown that users hardly ever change their default settings [160, 273, 237]. At the same time, choice is not necessarily going to result in increased user satisfaction. A study by Korff and Böhme [138] showed that participants who had more privacy options to select from reported more negative emotions and were less satisfied with the choice they made.

7.6.7 Being sensitive to context

People's use of technology is embedded in a context – it matters who they are, what experience they have, what they need to accomplish, what device they are using etc. One cannot give a stamp of approval for a technology that it is effortless and usable

in every context because as the design mantra says "it depends" [236]. For example, the study on the Vernitski Authentication Grid showed that the effort needed to use the mechanism was seen as appropriate in some context more than in others. Authentication using the grid was faster on a tablet, so it might be more advisable to deploy it on touchscreens. Most users struggle with infrequently used passwords, and there the use of a more memorable password can offset the longer input times. An important lesson from all the security studies described in this dissertation is that there is no usable one-size-fits-all replacement for passwords – rather, mechanisms need to be selected to fit the devices, context of use (primary task, physical and social context), security requirements and preferences of individual users.

7.6.8 Providing extra benefits

Researchers in the area of usable security and privacy have argued that users should be educated to understand the benefits of security and follow the advice of experts regardless of how much effort it requires. This approach is neither feasible nor desirable, instead security researchers and designers should aim to make the first hurdle to adoption of secure and privacy-preserving tools lower, by making them intuitive and a pleasure to use. Additionally, it has been suggested that security mechanisms could be about more than just granting access but they could also provide additional benefits to their users as suggested by Sasse and Krol [223]. For example, the Nymi band[2] uses electrocardiogram (ECG) as a secure biometric to authenticate the wearer to systems. One can also imagine that this technology could monitor the users' heart rate and offer alerts if there were any irregularities.

7.7 Limitations of research

The limitations of each study were described in their respective sections. To summarise, the diary studies examined security behaviours in two specific contexts which might not be generalisable to other situations. Both studies had a small number of well-educated and highly computer literate participants and other demographic groups might have different capabilities, preferences and needs. Despite the use of diaries which were employed to ground participants' reflections in actual events, the studies relied on self-reports which are subject to biases.

[2] https://www.nymi.com/

The evaluations of authentication and verification mechanisms were conducted in a laboratory setting where participants might not have been entirely under real-life constraints. The studies therefore might have suffered from biases such as demand characteristics, volunteer bias, an artificially strong focus on the task and obedience to authority. The study on the VAG did not have a primary task to mimic a real-life situation. In the study on verification mechanisms, there was a primary task but it might have been not as meaningful to participants since it was not a product they wanted to buy and they did not receive if after the study session. Nevertheless, the studies made an effort to capture perceptions around the mechanisms and participants discussed the broader context of deployment of these mechanisms in real life.

The two studies that looked at information disclosure in webforms were conducted on the Amazon Mechanical Turk platform where again the sample of participants might not have been representative of the general population. Due to the nature of their work, mTurk workers could have been more willing to volunteer information. The studies also observed only one reaction to a stimulus (i.e., webform, warning) and this reaction might not have been representative of their general behaviour.

Chapter 8

Conclusions

8.1 The research problem restated

As more and more services go online, effortless, reliable and efficient security and privacy will become even more important in the years to come. Every day users are asked to expend effort to keep accounts and services secure. For privacy, minimising the amount of information users give away about themselves is difficult since the industry prefer to collect more data and hence they do not make privacy-preserving settings the default. This doctoral project addressed a research gap in that the impact of such developments and the high effort demanded have on users' security and privacy behaviours had not been rigorously studied yet.

8.2 Overview of findings

The studies described in this dissertation examined the role of effort in users' security and privacy behaviours online. The two diary studies (Chapter 4) found that users developed coping strategies to manage their effort and stay in control of their overall workload. While in some situations they decided to cope and expend the effort, in others they just disengaged and avoided using accounts or devices. Participants took these decisions based on factors such as frequency of use, urgency and the level of security they believed was needed.

The studies conducted on proposed mechanisms (Chapter 5) demonstrated participants saw authentication and verification very contextually, elaborating in which situations they would find the use of these mechanisms acceptable. Often, they translated increased authentication and verification effort into better security which might be a wrong inference to make. The fact that participants imagined difficult mechanisms were more secure and preferred these mechanisms for more valuable accounts,

shows that they need reassurance and they currently get best reassurance when they take care of their own security themselves. One of the reasons why participants disliked NoBot was because there were many things happening outside of their influence and they were unsure how they would recover if an error occurred.

For information disclosure in webforms (Chapter 6), the first study showed participants disclosed information that was not necessary for them to receive payment. Giving a monetary bonus led them to disclose more information, even for fields that did not have a bonus awarded for. Marking some fields as mandatory, however, did not have this effect. This might mean that the uncertainty over what was really required made participants fill in more information as it might have been more effort to try to submit some information and be rejected than to fill all in from the start. The second study investigated exactly this and deployed warnings to inform participants when they were providing non-mandatory information. In brief, participants preferred a warning that gave an explanation and left them in control over which form fields they submit.

In summary, authentication effort led participants to adopt different coping strategies which might affect their own and organisational productivity, some of them might also decrease the security of the systems and accounts. For both security and privacy choices, participants showed they wanted control and choice as their decisions and resulting behaviours are highly contextual depending on various factors.

8.3 Research questions answered

The overarching research question of this doctoral project referred to the role of effort in users' security and privacy behaviours online. The studies conducted show the effort required in performing security tasks such as authentication can alter users' use of technologies and users might try to move the authentication to a more convenient time where it would not interfere with actual work. This leads to re-organisation of individual and organisational work flows. Participants also showed that they have their own views on what level of security they require and these are not always in line with what policy makers mandate. As a response to authentication effort, users find coping strategies which arguably might undermine security. A factor in users' willingness to expend effort is the belief that the level of security required of them might not be too high for the risk they perceive they are facing.

For the evaluation of mechanisms, participants found that time-consuming mechanisms should be used for infrequent and high-value transactions. Participants translated a high level of effort that was required to use a mechanism into a high level of protection it would give them. Again, a coping strategy of pointing at a grid might undermine the security of the mechanism.

In the case of information disclosure through webforms, habit, enjoyment and expectation of future rewards can override the effort needed to provide information. Making some fields mandatory and providing a bonus also impact disclosure. Once participants know some fields are mandatory, they are less likely to fill in other optional fields, the opposite effect holds for giving a bonus. For the proposed over-disclosure warnings, participants preferred being provided an explanation and having control even at the expense of effort. A longitudinal study would need to show if all these effects would hold over time.

8.4 Contributions of the research

The studies that form this dissertation make several important contributions to the field of usable security and privacy both in terms of findings and methodology.

8.4.1 Substantive contributions

The research contributes to the debate around security decisions and behaviours being based on a cost benefit calculation and a compliance budget [20]. It makes steps towards quantifying user effort which is important in both consumer and corporate contexts. The studies in this dissertation provide an insight into how users adopt security mechanisms and what their experience is with them over time. The two studies focusing on proposed mechanisms – the Vernitski Authentication Grid and NoBot – are the only user evaluations conducted on these mechanisms to date. Apart from providing specific recommendations for improving their usability, they also generate new knowledge around user perceptions, expectations and acceptability. The disclosure studies benchmarked the impact of mandatoriness and incentives on information disclosure in webforms. The subsequent study designed and evaluated warnings and examined their impact on user disclosure behaviours.

On many occasions, the ways in which participants managed increased security and privacy effort might have undermined their security and privacy. This shows that

usability and security are not opposing goals [224] and proves the mantra saying: "If it's not usable, it's not secure".

8.4.2 Methodological contributions

The studies described in this dissertation were designed and carried out in a way to maximise their ecological validity by either studying participant behaviour in real situations or approximating these conditions in a laboratory or through remote participation. The authentication diary studies furthered the use of diaries to capture authentication experiences in real life. The design of the diary was improved from previous research and from the first to the second study. The lab studies of authentication and verification mechanisms both not only captured performance data but also asked questions around acceptability and how these solutions would fit into users' every-day lives. The study on verification mechanisms provided participants with a realistic primary task to first make them focus on achieving a non-security goal as it happens in real life and second, to give them a context in which these mechanisms could be employed. The experiments studying disclosure in webforms were conducted outside a lab environment on a crowd-sourcing platform with participants taking part in an experiment in their usual environment likely using a familiar device.

8.5 Recommendations

8.5.1 Recommendations for researchers

To produce valid research findings, it is important to consider the context of use and test for use in different situations with different primary tasks if possible. The type of primary task can make a significant difference for how users perceive the security and privacy effort involved. What is acceptable in one context might not be acceptable in another.

Developing a solution, not only in security, is a multi-stage process and users should be involved at every stage of it. This includes testing mechanisms post-adoption to learn how users have incorporated them into their lives. First, this allows to change these mechanisms to improve the user experience. Second, the feedback can also benefit the design of future technologies.

It is advisable to use mixed research methods including quantitative and qualitative elements in a study. Had the studies in this dissertation only employed quantitative

measures, they would have missed explanations of user motivation and mental models that showed what kind of effort is acceptable and when. That said, it is important to be aware of participants making post-hoc rationalisations of their intuitive choices and reactions.

Following the designers' guideline that it is best to elicit feedback when users are shown designs side by side, this should be done for security and privacy too.

Finally, if it is a preliminary evaluation as opposed to in-the-wild study, it is interesting to gauge participants' intent to use a proposed technology. Although some participants liked NoBot in the study presented here, they said they were happy to continue using reCAPTCHA. Positive experience with a mechanism does not necessarily mean users will adopt it because most of them are likely to just continue using the established technology due to familiarity and habit.

8.5.2 Recommendations for practitioners

Advice for developers of technologies

It is crucial to involve users at every stage of the development of a product, starting from eliciting requirements and finishing with an evaluation post-adoption to see how users have appropriated the technology. As the study on verification showed, users tended to be more critical when they had a direct comparison between different products and they tried each one only once. As a general rule, acceptance improves over repeated use but this is not necessarily a good sign since users often learn to put up with mechanisms that are not usable. While a laboratory study often requires repeated use, in reality users might just give up after a first unsatisfactory experience. Even if they seemingly have no other options but to use a mechanism that is unusable and requires effort, they often find ways to circumvent it, sometimes meaning they will give up on devices or accounts and find alternatives.

Giving users choice is very important but the options that are presented to the user should not be a catalogue of every possible configuration but a selection of the most suitable choices. Setting up a security or privacy mechanism is likely to be a tertiary task for the user, meaning it takes place distantly from their actual goal. For this reason, it cannot cost too much effort as users will feel side-tracked and pushed further away from completing their primary task.

For privacy, users disclosed more when they were given a bonus, even on the fields that the bonus did not apply to. Making fields mandatory reduced the disclosure for the optional fields. Participants preferred control over effort and this might have something to do with the fact that users are all different, they had different attitudes to and expectations from completing the form. Rather than adopting the patronising approach of removing all optional information, users should be given choice.

It is advisable to study the existing products on the market to learn what users think about them. Even if a proposed product can be considered to require less effort, this has to be validated in a real-life setting and it is important to study user acceptance and readiness to switch. Established products even if resented, have the advantage of being well-understood by users and accepted as a necessary evil. A new product can often incur switching costs in that users need to learn how to use it and/or accommodate for new hardware.

Advice for decision-makers

Those mandating security solutions for organisations and consumers should consider how the introduction of a security measure would impact users' productivity, how it would interact with various primary tasks and also with other security tasks. Although initially users might like a new mechanism due to its novelty, this effect might wear off over time. It is therefore important not only to design well but also to monitor use and satisfaction over time.

8.6 Future work

Effort was a useful lens to consider security and privacy choices through. It brought up other interesting factors that affect behaviours such as the desire for control or the perceived level of security required. Future research could explore these issues in a systematic way to determine how exactly they impact the decisions users make.

8.6.1 On real-world authentication

Both diary studies were qualitative rather than quantitative in nature and were conducted on a small sample of participants. A larger study, for example a survey of the general population, would be needed to confirm if the established trends would hold.

Diaries have their shortcomings as they are disruptive and self-reports can be subject to biases. For this reason, the author's research group has been working on an

ethical and privacy-preserving tool [263] that currently collects data on the frequency of password entry, password strength and re-use. The idea is to extend it to capture authentication completion times, resets, mistakes, lockouts etc.

8.6.2 On proposed mechanisms

Future work could continue with further stages of a usability evaluation of the VAG. Participant responses showed in what kind of situations and for what types of systems the grid could be used and any future evaluations could focus on testing its deployment in these real-life contexts.

Authentication is different across contexts and the same level of effort can be seen as acceptable or unacceptable depending on the circumstances. To assess the suitability of security technologies for different contexts, it would be important to develop a range of standardised primary tasks to test the proposed mechanisms for. Authentication is contextual and to be valid, studies need to recreate use of these technologies under different conditions.

One of the limitations of the studies on the Vernitski Authentication Grid (Section 5.2) and verification mechanisms (Section 5.3) was that they were tested on a relatively small sample of participants who tended to be young and well-educated which might limit the generalisability of the findings. Testing the mechanisms in the wild with a wider range of services and in different primary tasks would be the next step.

For all studied security mechanisms, future work should focus on finding a better fit between the primary task and the authentication or verification task so that they are not competing for the same resources (cognitive or physical) and they do not interrupt or negatively affect the actual production task.

8.6.3 On helping users manage disclosure

Future work into tools helping users make an informed choice as to the disclosure of personal information could examine the usability of in-situ indicators that would interfere less with the primary task. They would need to appear at a suitable moment during the interaction in order not to disrupt the actual task.

Appendix A

Authentication in an organisation

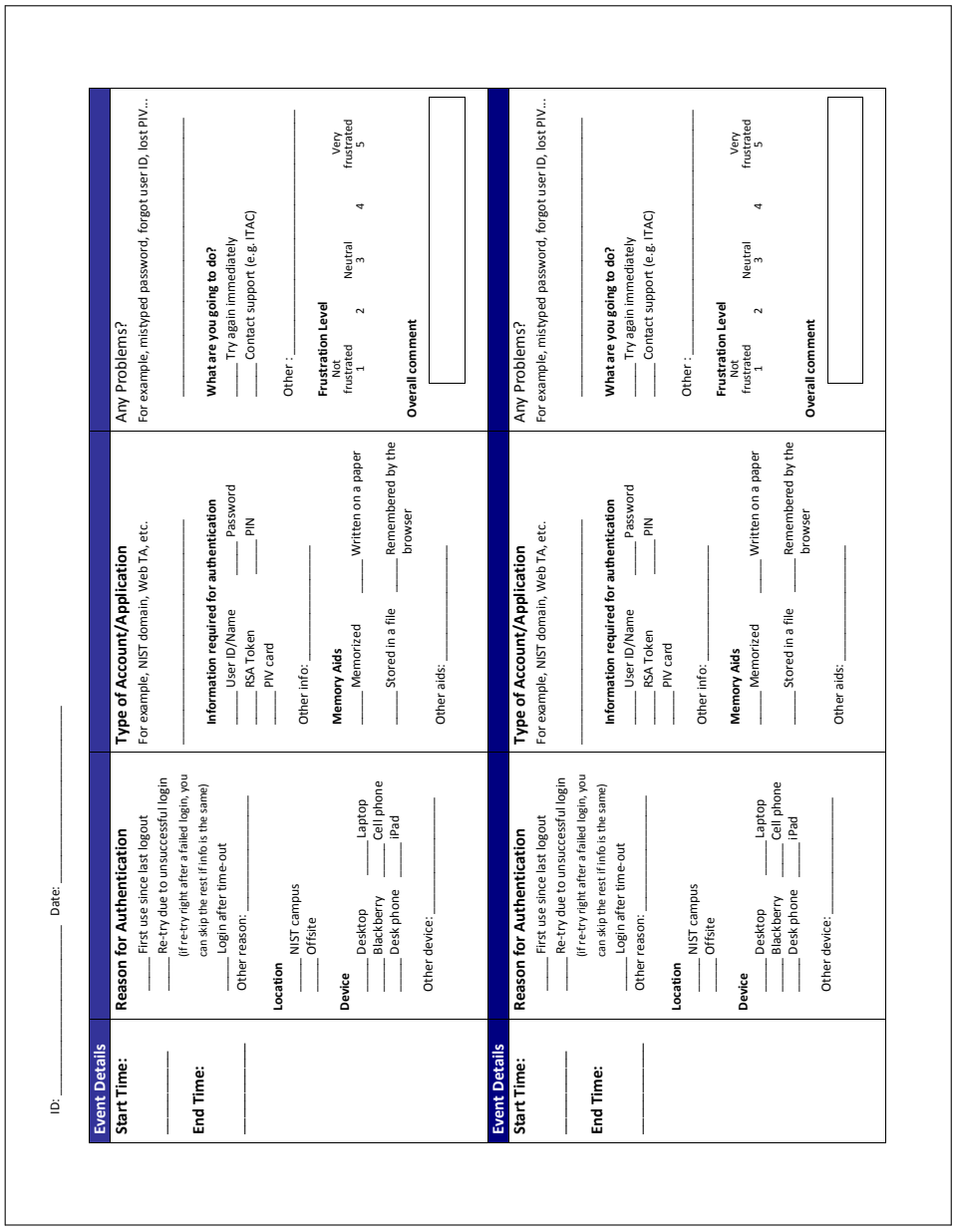

Figure A.1: An entry sheet from the diary used for the study on authentication in an organisation.

Step	Activity	GOMS-KLM Task Symbol	Time (in seconds)
1	Mentally prepare	M	1.35
2	Home hand on mouse	H	0.4
3	Position the cursor over the bookmark	P	1.1
4	Click mouse	P1	0.2
5	Position the cursor over the userid field	P	1.1
6	Click mouse	P1	0.2
7	Home hands on the keyboard	H	0.4
8	Recall userid	M	1.35
9	Enter userid (8 characters)	8 (K)	1.76
10	Home hand on the mouse	H	0.4
11	Position the cursor over the password field	P	1.1
12	Click mouse	P1	0.2
13	Home hands on the keyboard	H	0.4
14	Recall password	M	1.35
15	Enter password (8 characters)	8 (K)	1.76
16	Home hand on the mouse	H	0.4
17	Position the cursor over the submit button	P	1.1
18	Click mouse	P1	0.2
	Total Time		14.77

Figure A.2: Example Keystroke-Level Modelling sequence of steps for manually logging in to an application.

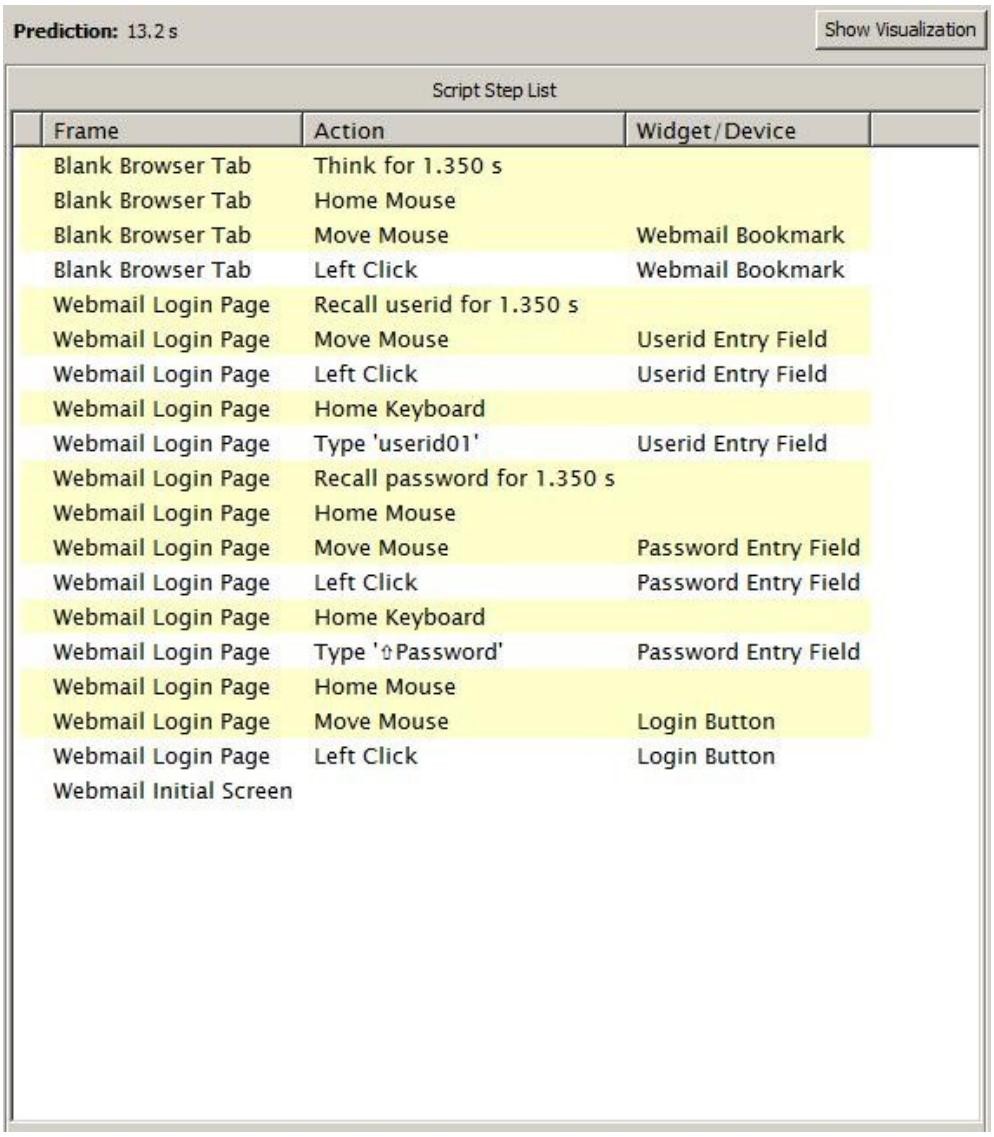

Figure A.3: A sample CogTool script for the Keystroke-Level Modelling sequence of steps for manually logging in to an application.

Appendix B

Two-factor authentication in online banking

Bank	Username	1st Factor	2nd Factor
Barclays-1	Surname + [Membership, Card, or Account No]	8-digit OTP (PINsentry or Mobile PINsentry	–
Barclays-2	Surname + [Membership Card, or Account No]	5-digit passcode	2 of 8+ characters memorable word
Halifax, Lloyds	User ID	8+ character password	3 of 6+ characters memorable information
HSBC, First Direct	User ID	memorable answer	6-digit OTP (Secure Key or Digital Secure Key)
Nationwide-1	Customer Number	memorable data	3 of 6 digits passnumber
Nationwide-2	Customer Number	8-digit OTP (card reader)	–
NatWest, RBS, Ulster	Customer Number (DOB ddmmyy + 4 digits)	3 of 4-digit PIN	3 of 6+ characters password
Santander-1	Personal ID	6+ character password	5-digit registration number
Santander-2	Personal ID	3 of 6+ character password	3 of 5-digit registration number

Table B.1: Credentials required for authentication to online banking with different UK banks.

Bank	ID always required?	Authentication Factor(s)
Barclays	No	5-digit PIN
Halifax, Lloyds	No	3 characters of memorable information
HSBC, First Direct	Yes No	Memorable answer + 3 characters of a password
Nationwide	No	3 of 6-digit passnumber
NatWest	No	5-8 digit PIN
RBS	No	5-8 digit PIN
Santander	Yes	3 characters of password, 3 digits of security number
Ulster	No	5-8 digit PIN

Table B.2: Credentials required for authentication to mobile apps with different UK banks.

My Authentication Diary

1. **Date**: _____ **Time**: _____
2. **Location: Where was I?**
   - □ Home
   - □ Workplace
   - □ Public place
   - □ On the move
   - □ Other: _____
3. **Reason: Why did I authenticate?**
   - □ Check account balance
   - □ Transfer funds between my accounts
   - □ Set up a payment to someone
   - □ Other: _____
4. **Device: What device did I use?**
   - □ Desktop Computer
   - □ Laptop
   - □ Mobile Phone
   - □ Tablet
   - □ Other: _____
5. **Information: What information was I asked to provide?**
   - □ Username
   - □ Password
   - □ Memorable answer
   - □ One Time Code
   - □ Other: _____
6. **Token: Did I use my authentication token?** □ Yes □ No
7. **Problems: Did I encounter any problems?**
   (forgotten credentials, mistypes, forgotten fields, token problems etc.)
   - □ Yes :
     _____
     _____
   - □ No
8. **How were these resolved?**
   _____
9. **The process left me:** Very Dissatisfied □—□—□—□—□ Very Satisfied
10. **Comments:**
    _____
    _____
    _____

Figure B.1: An entry page from the diary used for the study on two-factor authentication in UK online banking.

Appendix C

Human verification mechanisms

Easy to understand	Difficult to understand
Intuitive	Unintuitive
Exciting	Boring
Fast	Slow
Normal	Weird
Acceptable	Creepy
Fits well	Gets in the way
Great	Terrible
Harmless	Unhealthy
Trustworthy	Untrustworthy
Predictable	Unpredictable
Effortless	Effortful

Table C.1: A list of adjectives used in the study on human verification mechanisms. Each participant was asked to choose three adjectives that best described the verification mechanism(s) they used in the study. To shorten the reading time, positive adjectives were placed in the left-hand column while the corresponding negative ones were in the right-hand column.

Appendix D

Informing disclosure

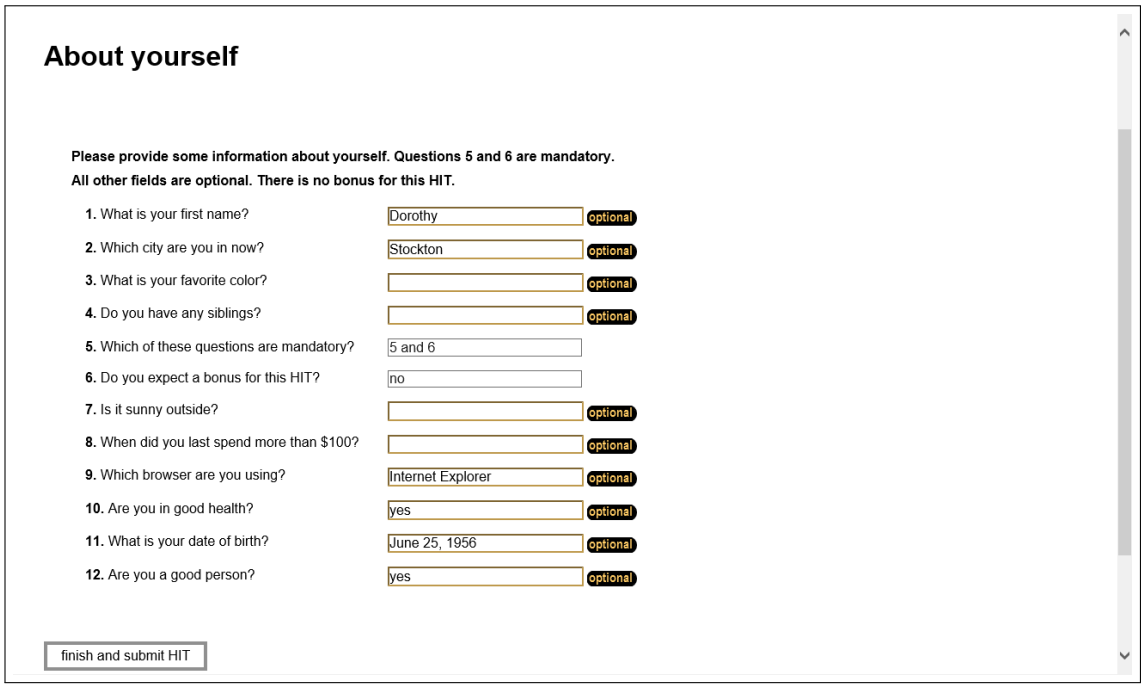

Figure D.1: A screenshot of the highlighting after the warning in condition WS1HO (*highlight optional fields*) was displayed.

References

[1] Ruba Abu-Salma, M. Angela Sasse, Joseph Bonneau, and Matthew Smith. POSTER: Secure Chat for the Masses? User-centered Security to the Rescue. In *ACM Conference on Computer and Communications Security (CCS 2015)*, pages 1623–1625. ACM, 2015.

[2] Mark S. Ackerman, Lorrie Faith Cranor, and Joseph Reagle. Privacy in e-commerce: Examining user scenarios and privacy preferences. In *ACM Conference on Electronic Commerce*, pages 1–8. ACM, 1999.

[3] Anne Adams and M. Angela Sasse. Users are not the enemy. *Communications of the ACM*, 42:40–46, 1999.

[4] Anne Adams and Martina Angela Sasse. Privacy in multimedia communications: Protecting users, not just data. In *People and Computers XV – Interaction without Frontiers*, pages 49–64. Springer, 2001.

[5] John Adams. Cars, cholera, and cows: The management of risk and uncertainty. *Cato Policy Analysis*, (335), 1999.

[6] Icek Ajzen. The theory of planned behavior. *Organizational behavior and human decision processes*, 50(2):179–211, 1991.

[7] Icek Ajzen and Martin Fishbein. *Understanding attitudes and predicting social behaviour*. Englewood Cliffs, NJ: Prentice-Hall, 1980.

[8] Devdatta Akhawe and Adrienne Porter Felt. Alice in Warningland: A Large-Scale Field Study of Browser Security Warning Effectiveness. In *USENIX Security*, pages 257–272, 2013.

[9] Eirik Albrechtsen. A qualitative study of users' view on information security. *Computers & Security*, 26(4):276–289, 2007.

[10] Bonnie Brinton Anderson, C. Brock Kirwan, Jeffrey L. Jenkins, David Eargle, Seth Howard, and Anthony Vance. How polymorphic warnings reduce habituation in the brain: Insights from an fMRI study. In *ACM Conference on Human Factors in Computing Systems (CHI 2015)*, pages 2883–2892. ACM, 2015.

[11] Ross Anderson. *Security engineering*. John Wiley & Sons, 2008.

[12] Panagiotis Andriotis, Theo Tryfonas, and George Oikonomou. Complexity metrics and user strength perceptions of the pattern-lock graphical authentication method. In *Human Aspects of Information Security, Privacy, and Trust*, pages 115–126. Springer, 2014.

[13] Dirk Annacker, Sarah Spiekermann, and Martin Strobel. E-privacy: Evaluating a new search cost in online environments. Technical report, Humboldt University of Berlin, Interdisciplinary Research Project 373: Quantification and Simulation of Economic Processes, 2001.

[14] Anthony. Why users fill out forms faster with unied text fields. `http://uxmovement.com/forms/why-users-fill-out-forms-faster-with-unified-text-fields/`, 2011. Accessed 12.11.2015.

[15] Albert Bandura. Self-efficacy mechanism in human agency. *American psychologist*, 37(2):122, 1982.

[16] Patti Bao, Jeffrey Pierce, Stephen Whittaker, and Shumin Zhai. Smart phone use by non-mobile business users. In *International Conference on Human Computer Interaction with Mobile Devices and Services (MobileHCI 2011)*, pages 445–454. ACM, 2011.

[17] Jakob E. Bardram. The trouble with login: On usability and computer security in ubiquitous computing. *Personal and Ubiquitous Computing*, 9(6):357–367, 2005.

[18] Jakob E. Bardram, Rasmus E. Kjær, and Michael Ø. Pedersen. Context-aware user authentication–supporting proximity-based login in pervasive computing. In *Ubiquitous Computing (UbiComp 2003)*, pages 107–123. Springer, 2003.

[19] Adam Beautement, Ingolf Becker, Simon Parkin, Kat Krol, and M. Angela Sasse. Productive security: A scalable methodology for analysing employee security behaviours. In *Symposium on Usable Privacy and Security (SOUPS 2016)*, pages 253–270. USENIX Association, 2016.

[20] Adam Beautement, M. Angela Sasse, and Mike Wonham. The compliance budget: Managing security behaviour in organisations. In *Workshop on New Security Paradigms (NSPW 2008)*, pages 47–58. ACM, 2008.

[21] Marios Belk, Christos Fidas, Panagiotis Germanakos, and George Samaras. Do cognitive styles of users affect preference and performance related to captcha challenges? In *Extended Abstracts on Human Factors in Computing Systems (CHI 2012)*, pages 1487–1492. ACM, 2012.

[22] Joey Benedek and Trish Miner. Measuring desirability: New methods for evaluating desirability in a usability lab setting. *Proceedings of Usability Professionals Association*, 2003:8–12, 2002.

[23] Zinaida Benenson, Gabriele Lenzini, Daniela Oliveira, Simon Parkin, and Sven Uebelacker. Maybe poor Johnny really cannot encrypt – The case for a complexity theory for usable security. In *New Security Paradigms Workshop (NSPW 2015)*, 2015.

[24] Bettina Berendt, Oliver Günther, and Sarah Spiekermann. Privacy in e-commerce: stated preferences vs. actual behavior. *Communications of the ACM*, 48(4):101–106, 2005.

[25] Alastair R. Beresford, Dorothea Kübler, and Sören Preibusch. Unwillingness to pay for privacy: A field experiment. *Economics Letters*, 117(1):25–27, 2012.

[26] Tim Berners-Lee and Dan Connolly. Hypertext Markup Language - 2.0. http://tools.ietf.org/html/rfc1866section-8, 1995.

[27] Rasekhar Bhagavatula, Blase Ur, Kevin Iacovino, Su Mon Kywe, Lorrie Faith Cranor, and Marios Savvides. Biometric authentication on iphone and android: Usability, perceptions, and influences on adoption. In *NDSS Workshop on Usable Security (USEC 2015)*, 2015.

[28] Robert Biddle, Sonia Chiasson, and Paul C. Van Oorschot. Graphical passwords: Learning from the first twelve years. *ACM Computing Surveys (CSUR)*, 44(4):19, 2012.

[29] Robert Biddle, Paul C. van Oorschot, Andrew S. Patrick, Jennifer Sobey, and Tara Whalen. Browser interfaces and extended validation SSL certificates: An empirical study. In *ACM Workshop on Cloud Computing Security*, pages 19–30. ACM, 2009.

[30] John M. Blythe, Lynne Coventry, and Linda Little. Unpacking security policy compliance: The motivators and barriers of employees' security behaviors. In *Symposium On Usable Privacy and Security (SOUPS 2015)*, pages 103–122, 2015.

[31] Niall Bolger, Angelina Davis, and Eshkol Rafaeli. Diary methods: Capturing life as it is lived. *Annual review of psychology*, 54(1):579–616, 2003.

[32] Sadie Boniface and Nicola Shelton. How is alcohol consumption affected if we account for under-reporting? A hypothetical scenario. *The European Journal of Public Health*, 2013.

[33] Joseph Bonneau. The science of guessing: Analyzing an anonymized corpus of 70 million passwords. In *IEEE Symposium on Security and Privacy (S&P 2012)*, pages 538–552. IEEE, 2012.

[34] Joseph Bonneau, Cormac Herley, Paul C. Van Oorschot, and Frank Stajano. The quest to replace passwords: A framework for comparative evaluation of web authentication schemes. In *IEEE Symposium on Security and Privacy (S&P 2012)*, pages 553–567. IEEE, 2012.

[35] Joseph Bonneau, Mike Just, and Greg Matthews. What's in a Name? Evaluating Statistical Attacks on Personal Knowledge Questions. In *Financial Cryptography and Data Security (FC 2010)*, pages 98–113. Springer, 2010.

[36] Joseph Bonneau and Sören Preibusch. The password thicket: Technical and market failures in human authentication on the web. In *Workshop on the Economics of Information Security (WEIS 2010)*, 2010.

[37] Joseph Bonneau, Sören Preibusch, and Ross Anderson. A birthday present every eleven wallets? The security of customer-chosen banking PINs. In *Financial Cryptography and Data Security (FC 2012)*, pages 25–40. Springer, 2012.

[38] Joseph Bonneau and Stuart Schechter. Towards reliable storage of 56-bit secrets in human memory. In *USENIX Security*, 2014.

[39] J. Efrim Boritz. Is practitioners' views on core concepts of information integrity. *International Journal of Accounting Information Systems*, 6(4):260–279, 2005.

[40] Laura Brandimarte, Alessandro Acquisti, and George Loewenstein. Misplaced confidences privacy and the control paradox. *Social Psychological and Personality Science*, 4(3):340–347, 2013.

[41] Virginia Braun and Victoria Clarke. Using thematic analysis in psychology. *Qualitative research in psychology*, 3(2):77–101, 2006.

[42] Alex Braunstein, Laura Granka, and Jessica Staddon. Indirect content privacy surveys: Measuring privacy without asking about it. In *Symposium on Usable Privacy and Security (SOUPS 2011)*, pages 1–14. ACM, 2011.

[43] Christina Braz and Jean-Marc Robert. Security and usability: the case of the user authentication methods. In *International Conference of the Association Francophone d'Interaction Homme-Machine*, 2006.

[44] British Parliament. Data Protection Act of 1998, 1998.

[45] Sacha Brostoff, Philip Inglesant, and M. Angela Sasse. Evaluating the usability and security of a graphical one-time PIN system. In *BCS Interaction Specialist Group Conference*, pages 88–97. British Computer Society, 2010.

[46] Sacha Brostoff and M. Angela Sasse. Are Passfaces more usable than passwords? A field trial investigation. *People and Computers*, pages 405–424, 2000.

[47] Alan Bryman. *Social research methods*. Oxford University Press, 2012.

[48] Elie Bursztein, Steven Bethard, Celine Fabry, John C. Mitchell, and Dan Jurafsky. How good are humans at solving CAPTCHAs? A large scale evaluation. In *IEEE Symposium on Security and Privacy (S&P 2010)*, pages 399–413. IEEE, 2010.

[49] Stuart K. Card, Thomas P. Moran, and Allen Newell. The keystroke-level model for user performance time with interactive systems. *Communications of the ACM*, 23(7):396–410, 1980.

[50] Scott Carter and Jennifer Mankoff. When participants do the capturing: The role of media in diary studies. In *SIGCHI Conference on Human Factors in Computing Systems (CHI 2005)*, pages 899–908. ACM, 2005.

[51] J. Alberto Castañeda and Francisco J. Montoro. The effect of Internet general privacy concern on customer behavior. *Electronic Commerce Research*, 7(2):117–141, 2007.

[52] Kumar Chellapilla, Kevin Larson, Patrice Simard, and Mary Czerwinski. Designing human friendly human interaction proofs (hips). In *SIGCHI Conference on Human Factors in Computing Systems*, pages 711–720. ACM, 2005.

[53] Eun Kyoung Choe, Jaeyeon Jung, Bongshin Lee, and Kristie Fisher. Nudging people away from privacy-invasive mobile apps through visual framing. In *Human-Computer Interaction – INTERACT 2013*, pages 74–91. Springer, 2013.

[54] Robert B. Cialdini. Influence: The psychology of persuasion. *New York, NY: Morrow*, 1993.

[55] Kay Connelly, Ashraf Khalil, and Yong Liu. Do I do what I say?: Observed versus stated privacy preferences. In *Human-Computer Interaction – INTERACT 2007*, pages 620–623, 2007.

[56] Lynne Coventry, Antonella De Angeli, and Graham Johnson. Usability and biometric verification at the ATM interface. In *SIGCHI Conference on Human Factors in Computing Systems (CHI 2003)*, pages 153–160. ACM, 2003.

[57] Lorrie Faith Cranor. A framework for reasoning about the human in the loop. In *Proceedings of the 1st Conference on Usability, Psychology and Security*, volume 1, pages 1–15. USENIX Association, 2008.

[58] Mary J. Culnan. The Culnan-Milne survey on consumers & online privacy notices: Summary of responses. In *Interagency Public Workshop: Get Noticed: Effective Financial Privacy Notices*, pages 47–54, 2001.

[59] Mary J. Culnan and Pamela K. Armstrong. Information privacy concerns, procedural fairness, and impersonal trust: An empirical investigation. *Organization Science*, 10(1):104–115, 1999.

[60] Darren Davis, Fabian Monrose, and Michael K. Reiter. On user choice in graphical password schemes. In *USENIX Security*, volume 13, pages 11–11, 2004.

[61] Fred D. Davis. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS quarterly*, pages 319–340, 1989.

[62] Fred D. Davis, Richard P. Bagozzi, and Paul R. Warshaw. User acceptance of computer technology: A comparison of two theoretical models. *Management Science*, 35(8):982–1003, 1989.

[63] Emiliano De Cristofaro, Honglu Du, Julien Freudiger, and Greg Norcie. A Comparative Usability Study of Two-Factor Authentication. In *NDSS Workshop on Usable Security (USEC 2014)*, 2014.

[64] Norman K. Denzin. *The research act: A theoretical introduction to sociological methods*. Transaction publishers, 1973.

[65] Premkumar T. Devanbu and Stuart Stubblebine. Software engineering for security: a roadmap. In *Proceedings of the Conference on the Future of Software Engineering*, pages 227–239. ACM, 2000.

[66] Tamara Dinev and Paul Hart. Privacy concerns and levels of information exchange: An empirical investigation of intended e-services use. *E-Service Journal*, 4(3):25–59, 2006.

[67] Serge Egelman, Alessandro Acquisti, David Molnar, Cormac Herley, Nicolas Christin, and Shriram Krishnamurthi. Please Continue to Hold: An empirical study on user tolerance of security delays. In *Workshop on the Economics of Information Security (WEIS 2010)*, 2010.

[68] Serge Egelman, Lorrie Faith Cranor, and Jason Hong. You've been warned: An empirical study of the effectiveness of web browser phishing warnings. In *SIGCHI Conference on Human Factors in Computing Systems (CHI 2008)*, pages 1065–1074, 2008.

[69] Serge Egelman, Janice Y. Tsai, and Lorrie Faith Cranor. Tell me lies: A methodology for scientifically rigorous security user studies. In *Workshop on Studying Online Behaviour at CHI 2010*. ACM, 2010.

[70] Karin Eilers, Friedhelm Nachreiner, and Kerstin Hänecke. Entwicklung und Überprüfung einer Skala zur Erfassung subjektiv erlebter Anstrengung. *Zeitschrift für Arbeitswissenschaft*, 40(4):214–224, 1986.

[71] Carl Ellison, Chris Hall, Randy Milbert, and Bruce Schneier. Protecting secret keys with personal entropy. *Future Generation Computer Systems*, 16(4):311–318, 2000.

[72] Nicole B. Ellison, Charles Steinfield, and Cliff Lampe. The benefits of facebook "friends": Social capital and college students' use of online social network sites. *Journal of Computer-Mediated Communication*, 12(4):1143–1168, 2007.

[73] Katherine M. Everitt, Tanya Bragin, James Fogarty, and Tadayoshi Kohno. A comprehensive study of frequency, interference, and training of multiple graphical passwords. In *SIGCHI Conference on Human Factors in Computing Systems (CHI 2009)*, pages 889–898. ACM, 2009.

[74] Adrienne Porter Felt, Robert W. Reeder, Hazim Almuhimedi, and Sunny Consolvo. Experimenting at scale with Google Chrome's SSL warning. In *ACM Conference on Human Factors in Computing Systems (CHI 2014)*, pages 2667–2670. ACM, 2014.

[75] Ana Ferreira, Lynne Coventry, and Gabriele Lenzini. Principles of Persuasion in Social Engineering and Their Use in Phishing. In *Human Aspects of Information Security, Privacy, and Trust, HAS 2015, HCI International 2015*, volume LNCS 9190, pages 36–47, 2015.

[76] Guillaume Ferrero. L'inertie mentale et la loi du moindre effort. *Revue Philosophique de la France et de l'Étranger*, pages 169–182, 1894.

[77] Christos A. Fidas, Artemios G. Voyiatzis, and Nikolaos M. Avouris. On the necessity of user-friendly CAPTCHA. In *SIGCHI Conference on Human Factors in Computing Systems (CHI 2011)*, pages 2623–2626. ACM, 2011.

[78] Martin Fishbein and Icek Ajzen. *Belief, attitude, intention and behavior: An introduction to theory and research*. Reading, MA: Addison-Wesley, 1975.

[79] Karen E. Fisher and Lynne McKechnie. *Theories of information behavior*. Information Today, Inc., 2005.

[80] Paul M. Fitts. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of experimental psychology*, 47(6):381, 1954.

[81] Dinei Florêncio and Cormac Herley. A large-scale study of web password habits. In *International Conference on World Wide Web*, pages 657–666. ACM, 2007.

[82] Dinei Florêncio, Cormac Herley, and Paul C. Van Oorschot. Password portfolios and the finite-effort user: Sustainably managing large numbers of accounts. In *USENIX Security Symposium (USENIX Security 2014)*, pages 575–590, 2014.

[83] Will Friedman and Frank Chang. Collect Demographic Data More Easily with Internet Explorer 5. `https://msdn.microsoft.com/en-us/library/bb250414.aspx`, 1999. Accessed 08.08.2015.

[84] Masahiro Fujita, Yuki Ikeya, Junya Kani, and Masakatsu Nishigaki. Chimera captcha: A proposal of captcha using strangeness in merged objects. In *Human Aspects of Information Security, Privacy, and Trust, HAS 2015, HCI International 2015*, pages 48–58. Springer, 2015.

[85] Steven Furnell. Why users cannot use security. *Computers & Security*, 24(4):274–279, 2005.

[86] GAIN Publishing. Gator.com – Home. `http://web.archive.org/web/20031031020937/http://www.gator.com/home2.html`, 2003. Accessed 12.11.2015.

[87] GAIN Publishing. Gator.com – Home – Important information about GAIN software. `http://web.archive.org/web/20060630073649/http://www.gator.com/home2.html`, 2006. Accessed 12.11.2015.

[88] Dan Gardner. *Risk: The science and politics of fear*. Random House, 2009.

[89] Gerd Gigerenzer. *Gut feelings: The intelligence of the unconscious*. New York, NY: Viking Press, 2007.

[90] Gerd Gigerenzer, Ralph Hertwig, and Thorsten Pachur. *Heuristics: The foundations of adaptive behavior*. Oxford University Press, 2011.

[91] Gerd Gigerenzer, Peter M. Todd, and the ABC Group. *Simple heuristics that make us smart*. Oxford University Press, USA, 2000.

[92] Daniel G. Goldstein and Gerd Gigerenzer. Models of ecological rationality: the recognition heuristic. *Psychological review*, 109(1):75, 2002.

[93] Dieter Gollmann. *Computer security*. John Wiley & Sons, 2011.

[94] Google. Are you a robot? Introducing "No CAPTCHA reCAPTCHA". `https://googleonlinesecurity.blogspot.co.uk/2014/12/are-you-robot-introducing-no-captcha.html`, 2014. Accessed 08.09.2015.

[95] Rich Gossweiler, Maryam Kamvar, and Shumeet Baluja. What's up CAPTCHA? A CAPTCHA based on image orientation. In *International Conference on World Wide Web*, pages 841–850. ACM, 2009.

[96] John D. Gould and Clayton Lewis. Designing for usability: Key principles and what designers think. *Communications of the ACM*, 28(3):300–311, 1985.

[97] Tabreez Govani and Harriet Pashley. Student awareness of the privacy implications when using Facebook. *"Privacy Poster Fair" at the Carnegie Mellon University School of Library and Information Science*, 9, 2005.

[98] David Gragg. A multi-level defense against social engineering. *SANS Reading Room*, 13, 2003.

[99] Kristen K. Greene, Melissa A. Gallagher, Brian C. Stanton, and Paul Y. Lee. I Can't Type That! P@$$w0rd Entry on Mobile Devices. In *Human Aspects of Information Security, Privacy, and Trust*, volume LNCS 8533, pages 160–171. Springer, 2014.

[100] Joshua B. Gross and Mary Beth Rosson. Looking for trouble: understanding end-user security management. In *Symposium on Computer Human Interaction for the Management of information Technology*, page 10. ACM, 2007.

[101] Jeremiah Grossman. I know who your name, where you work, and live (Safari v4 & v5). `http://jeremiahgrossman.blogspot.com/2010/07/i-know-who-your-name-where-you-work-and.html`, 2010. Accessed 08.08.2015.

[102] Nancie Gunson, Diarmid Marshall, Hazel Morton, and Mervyn Jack. User perceptions of security and usability of single-factor and two-factor authentication in automated telephone banking. *Computers & Security*, 30(4), 2011.

[103] Seda Gürses. *Multilateral Privacy Requirements Analysis in Online Social Network Services*. PhD thesis, K.U. Leuven, 2010.

[104] Il-Horn Hann, Kai-Lung Hui, Sang-Yong Tom Lee, and Ivan P.L. Png. Overcoming online information privacy concerns: An information-processing theory approach. *Journal of Management Information Systems*, 24(2):13–42, 2007.

[105] Il-Horn Hann, Kai-Lung Hui, Tom S. Lee, and Ivan P.L. Png. Online information privacy: Measuring the cost-benefit trade-off. In *International Conference on Information Systems*, pages 1–8, 2002.

[106] Sandra G. Hart and Lowell E. Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. *Advances in psychology*, 52:139–183, 1988.

[107] Eiji Hayashi and Jason Hong. A diary study of password usage in daily life. In *ACM CHI Conference on Human Factors in Computing Systems*, 2011.

[108] David F. Heathfield. *Production functions*. Macmillan London, 1971.

[109] Rosa R. Heckle, Andrew S. Patrick, and Ant Ozok. Perception and acceptance of fingerprint biometric technology. In *Symposium on Usable Privacy and Security (SOUPS 2007)*, pages 153–154. ACM, 2007.

[110] Her Majesty's Revenue and Customs (HMRC). Guidance: Genuine HM Revenue and Customs contact and recognising phishing emails. `https://www.gov.uk/government/publications/genuine-hmrc-contact-and-recognising-phishing-emails/`, 2016. Accessed 02.12.2015.

[111] Tejaswini Herath and H. Raghav Rao. Protection motivation and deterrence: A framework for security policy compliance in organisations. *European Journal of Information Systems*, 18(2):106–125, 2009.

[112] Cormac Herley. So long, and no thanks for the externalities: The rational rejection of security advice by users. In *New Security Paradigms Workshop (NSPW 2009)*, pages 133–144. ACM, 2009.

[113] Cormac Herley. More is not the answer. *IEEE Security & Privacy*, 12(1):14–19, 2014.

[114] Cormac Herley and Paul Van Oorschot. A research agenda acknowledging the persistence of passwords. *IEEE Security & Privacy*, 10(1):28–36, 2012.

[115] Christine Hine. Privacy in the marketplace. *The Information Society*, 14(4):253–262, 1998.

[116] Chien-Ju Ho, Chen-Chi Wu, Kuan-Ta Chen, and Chin-Laung Lei. DevilTyper: A game for CAPTCHA usability evaluation. *Computers in Entertainment (CIE)*, 9(1):3, 2011.

[117] Godfrey Hochbaum, Irwin Rosenstock, and Stephen Kegels. Health belief model. *United States Public Health Service*, 1952.

[118] Harold Hotelling. Stability in competition. *Economic Journal*, 39(153):41–57, 1929.

[119] Jean-Pierre Hubaux and Ari Juels. Privacy is dead, long live privacy. *Communications of the ACM*, 59(6):39–41, 2016.

[120] Kai-Lung Hui, Hock Hai Teo, and Sang-Yong Tom Lee. The value of privacy assurance: An exploratory field experiment. *Mis Quarterly*, 31(1):19–33, 2007.

[121] Princely Ifinedo. Understanding information systems security policy compliance: An integration of the theory of planned behavior and the protection motivation theory. *Computers & Security*, 31(1):83–95, 2012.

[122] Philip G. Inglesant and M. Angela Sasse. The true cost of unusable password policies: Password use in the wild. In *SIGCHI Conference on Human Factors in Computing Systems (CHI 2010)*, pages 383–392. ACM, 2010.

[123] International Organization for Standardization. ISO 9241-11: Ergonomic requirements for office work with visual display terminals (VDTs – Part 11: Guidance on Usability). *The International Organization for Standardization*, 45, 1998.

[124] Caroline Jarrett and Gerry Gaffney. *Forms that work: Designing Web forms for usability*. Morgan Kaufmann, 2009.

[125] Jeffrey L. Jenkins, Mark Grimes, Jeffrey Gainer Proudfoot, and Paul Benjamin Lowry. Improving password cybersecurity through inexpensive and minimally invasive means: Detecting and deterring password reuse through keystroke-dynamics monitoring and just-in-time fear appeals. *Information Technology for Development*, 20(2):196–213, 2014.

[126] Carlos Jensen, Colin Potts, and Christian Jensen. Privacy practices of Internet users: Self-reports versus observed behavior. *International Journal of Human-Computer Studies*, 63(1-2):203–227, 2005.

[127] Nicola Jentzsch, Sören Preibusch, and Andreas Harasser. Study on monetising privacy: An economic model for pricing personal information. *ENISA Report*, 2012.

[128] Daniel Kahneman. *Thinking, fast and slow*. Macmillan, 2011.

[129] Daniel Kahneman and Amos Tversky. Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the econometric society*, pages 263–291, 1979.

[130] Ambarish Karole, Nitesh Saxena, and Nicolas Christin. A comparative usability evaluation of traditional password managers. In *Information Security and Cryptology-ICISC 2010*, pages 233–251. Springer, 2011.

[131] Auguste Kerckhoffs. La cryptographie militaire. *Journal des sciences militaires*, 9:5–38, 1883.

[132] Hassan Khan, Urs Hengartner, and Daniel Vogel. Usability and security perceptions of implicit authentication: Convenient, secure, sometimes annoying. In *Symposium on Usable Privacy and Security (SOUPS 2015)*, pages 225–239, 2015.

[133] Kevin S. Killourhy, Roy Maxion, et al. Comparing anomaly-detection algorithms for keystroke dynamics. In *IEEE/IFIP International Conference on Dependable Systems & Networks (DSN 2009)*, pages 125–134. IEEE, 2009.

[134] Iacovos Kirlappos, Simon Parkin, and M. Angela Sasse. Learning from "shadow security": Why understanding non-compliance provides the basis for effective security. In *NDSS Workshop on Usable Security (USEC 2014)*, 2014.

[135] Iacovos Kirlappos, Simon Parkin, and M. Angela Sasse. Shadow security as a tool for the learning organization. *SIGCAS Computers and Society*, 45(1):29–37, 2015.

[136] Kurt Alfred Kluever and Richard Zanibbi. Balancing usability and security in a video CAPTCHA. In *Symposium on Usable Privacy and Security (SOUPS 2009)*, pages 1–10. ACM, 2009.

[137] Hsiu-Chia Ko. The determinants of continuous use of social networking sites: An empirical study on Taiwanese journal-type bloggers' continuous self-disclosure behavior. *Electronic Commerce Research and Applications*, 12(2):103–111, 2013.

[138] Stefan Korff and Rainer Böhme. Too much choice: End-user privacy decisions in the context of choice proliferation. In *Symposium on Usable Privacy and Security (SOUPS 2014)*, pages 69–87, 2014.

[139] Isidora Kourti. Project FLAME social study report. Technical report, London School of Economics and Political Science (LSE), 2009.

[140] Kat Krol. "Wait: That's optional!" Designing helpful over-disclosure alerts. In *Designing Human Technologies (DHT) 2.0*, 2014.

[141] Kat Krol, Matthew Moroz, and M. Angela Sasse. Don't work. Can't work? Why it's time to rethink security warnings. In *International Conference on Risk and Security of Internet and Systems (CRiSIS 2012)*, pages 1–8. IEEE, 2012.

[142] Kat Krol, Constantinos Papanicolaou, Alexei Vernitski, and M. Angela Sasse. "Too Taxing on the Mind!" Authentication Grids are not for Everyone. In *Human Aspects of Information Security, Privacy, and Trust, HAS 2015, HCI International 2015*, volume LNCS 9190, pages 71–82, 2015.

[143] Kat Krol, Simon Parkin, and M. Angela Sasse. Better the devil you know: A user study of two CAPTCHAs and a possible replacement technology. In *NDSS Workshop on Usable Security (USEC 2016)*, 2016.

[144] Kat Krol, Simon Parkin, and M. Angela Sasse. "I don't like putting my face on the Internet!" An acceptance study of face biometrics as a CAPTCHA replacement. In *IEEE International Conference on Identity, Security and Behavior, Analysis (ISBA 2016)*, 2016.

[145] Kat Krol, Eleni Philippou, Emiliano De Cristofaro, and M. Angela Sasse. "They brought in the horrible key ring thing!" Analysing the Usability of Two-Factor Authentication in UK Online Banking. In *NDSS Workshop on Usable Security (USEC 2015)*, 2015.

[146] Kat Krol and Sören Preibusch. Effortless privacy negotiations. *IEEE Security & Privacy*, (3):88–91, 2015.

[147] Kat Krol and Sören Preibusch. Control versus effort in privacy warnings for webforms. In *ACM Workshop on Privacy in the Electronic Society (WPES 2016)*, pages 13–23. ACM, 2016.

[148] Kat Krol, Jonathan M. Spring, Simon Parkin, and M. Angela Sasse. Towards robust experimental design for user studies in security and privacy. In *The LASER Workshop: Learning from Authoritative Security Experiment Results (LASER 2016)*. IEEE, 2016.

[149] Steve Krug. *Don't make me think: A common sense approach to web usability*. Pearson Education India, 2005.

[150] Justin Kruger, Derrick Wirtz, Leaf Van Boven, and T. William Altermatt. The effort heuristic. *Journal of Experimental Social Psychology*, 40(1):91–98, 2004.

[151] Kelvin Lancaster. *Variety, equity, and efficiency: product variety in an industrial society*, volume 10. New York, NY: Columbia University Press, 1979.

[152] Caroline Lancelot Miltgen. Customers' privacy concerns and responses towards a request for personal data on the Internet: An experimental study. *Information Management in the Networked Economy: Issues and Solutions*, pages 400–415, 2007.

[153] Robert LaRose and Nora J. Rifon. 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, 2007.

[154] Jonathan Lazar, Jinjuan Heidi Feng, and Harry Hochheiser. *Research methods in human-computer interaction*. John Wiley & Sons, 2010.

[155] Pedro Leon, Blase Ur, Richard Shay, Yang Wang, Rebecca Balebako, and Lorrie Cranor. Why Johnny can't opt out: A usability evaluation of tools to limit online behavioral advertising. In *SIGCHI Conference on Human Factors in Computing Systems (CHI 2012)*, pages 589–598. ACM, 2012.

[156] Kevin Lewis, Jason Kaufman, and Nicholas Christakis. The taste for privacy: An analysis of college student privacy settings in an online social network. *Journal of Computer-Mediated Communication*, 14(1):79–100, 2008.

[157] John Leyden. iBank: RBS, NatWest first UK banks to allow Apple Touch ID logins. `http://www.theregister.co.uk/2015/02/19/natwest_mobile_banking_touch_id/`, 2015. Accessed 02.12.2015.

[158] May O. Lwin and Jerome D. Williams. A model integrating the multidimensional developmental theory of privacy and theory of planned behavior to examine fabrication of information online. *Marketing Letters*, 14(4):257–272, 2003.

[159] Robert L. Mack and Jakob Nielsen. *Usability inspection methods*. John Wiley & Sons, 1994.

[160] Wendy E. Mackay. Triggers and barriers to customizing software. In *SIGCHI Conference on Human Factors in Computing Systems (CHI 1991)*, pages 153–160. ACM, 1991.

[161] Miguel Malheiros and Sören Preibusch. Sign-up or give-up: Exploring user drop-out in web service registration. In *Symposium on Usable Privacy and Security Workshops (SOUPS 2013)*, 2013.

[162] Miguel Malheiros, Sören Preibusch, and M. Angela Sasse. "Fairly Truthful": The Impact of Perceived Effort, Fairness, Relevance, and Sensitivity on Personal Data Disclosure. In Michael Huth, N. Asokan, Srdjan Čapkun, Ivan Fléchais, and Lizzie Coles-Kemp, editors, *Trust and Trustworthy Computing*, volume LNCS 7904, pages 250–266. Springer Berlin Heidelberg, 2013.

[163] Stephen T. Margulis. Privacy as a social issue and behavioral concept. *Journal of Social Issues*, 59(2):243–261, 2003.

[164] Marcel Mauss. *The gift: Forms and functions of exchange in archaic societies*. Routledge, 1954.

[165] Aleecia M. McDonald and Lorrie Faith Cranor. Cost of reading privacy policies, the. *ISJLP*, 4:543, 2008.

[166] Miriam J. Metzger. Communication privacy management in electronic commerce. *Journal of Computer-Mediated Communication*, 12(2):335–361, 2007.

[167] Microsoft. Microsoft, Trustworthy Computing – Data Privacy Day. `http://www.microsoft.com/en-us/twc/privacy/data-privacy-day.aspx`, 2013. Accessed 08.08.2015.

[168] John Stuart Mill. On the definition of political economy, and on the method of investigation proper to it. In *Essays on Some Unsettled Questions of Political Economy*. London: Longmans, Green, Reader & Dyer, 1836.

[169] Kevin D. Mitnick and William L. Simon. *The art of deception: Controlling the human element of security*. John Wiley & Sons, 2011.

[170] David Modic and Ross Anderson. Reading this may harm your computer: The psychology of malware warnings. *Computers in Human Behavior*, 41:71–79, 2014.

[171] Stephen Monsell. Task switching. *Trends in cognitive sciences*, 7(3):134–140, 2003.

[172] Helen Nissenbaum. *Privacy in context: Technology, policy, and the integrity of social life*. Stanford University Press, 2009.

[173] Patricia A. Norberg, Daniel R. Horne, and David A. Horne. The privacy paradox: Personal information disclosure intentions versus behaviors. *Journal of Consumer Affairs*, 41(1):100–126, 2007.

[174] Donald A. Norman. Some observations on mental models. *Mental models*, 7(112):7–14, 1983.

[175] Jan M. Noyes and Daniel P.J. Bruneau. A self-analysis of the NASA-TLX workload measure. *Ergonomics*, 50(4):514–519, 2007.

[176] Andrew Odlyzko. Privacy, economics, and price discrimination on the Internet. In *International Conference on Electronic Commerce*, pages 355–366. ACM, 2003.

[177] Lawrence O'Gorman. Comparing passwords, tokens, and biometrics for user authentication. *Proceedings of the IEEE*, 91(12):2021–2040, 2003.

[178] Magda Osman. Does our unconscious rule? *PSYCHOLOGIST*, 28(2):114–117, 2015.

[179] Dennis Pagano and Bernd Brügge. User involvement in software evolution practice: a case study. In *Proceedings of the 2013 international conference on Software engineering*, pages 953–962. IEEE Press, 2013.

[180] Constantinos Papanicolaou. Novel authentication solution. Master's thesis, Department of Computer Science, University College London, 2013.

[181] Simon Parkin, Samy Driss, Kat Krol, and M. Angela Sasse. Assessing the user experience of password reset policies in a university. In *Passwords 2015*, 2015.

[182] Jeunese Payne. Misapplication of Psychology: The Case of Graphical Passwords. `http://www.jeunesepayne.co.uk/#!Misapplication-of-Psychology-The-Case-of-Graphical-Passwords/cg4/566961d10cf29cc3864a6260`, 2015. Accessed 12.12.2015.

[183] Jeunese Payne, Graeme Jenkinson, Frank Stajano, M Angela Sasse, and Max Spencer. Responsibility and tangible security: Towards a theory of user acceptance of security tokens. In *NDSS Workshop on Usable Security (USEC 2016)*, 2016.

[184] Joseph Phelps, Glen Nowak, and Elizabeth Ferrell. Privacy concerns and consumer willingness to provide personal information. *Journal of Public Policy & Marketing*, 19(1):27–41, 2000.

[185] Patrick Philippot and S. Canter. Fill in web forms automatically. *PC Magazine*, 18:16, 1999.

[186] David Pogue. Time to kill off CAPTCHAs. *Scientific American*, 306(3):23–23, 2012.

[187] Marc Pomplun and Sindhura Sunkara. Pupil dilation as an indicator of cognitive workload in human-computer interaction. In *Proceedings of the International Conference on HCI*, 2003.

[188] Chris Porter, M. Angela Sasse, and Emmanuel Letier. Designing acceptable user registration processes for e-services. In *BCS Conference on Human Computer Interaction*, 2012.

[189] Clay Posey, Tom L. Roberts, and Paul Benjamin Lowry. The impact of organizational commitment on insiders' motivation to protect organizational information assets. *Journal of Management Information Systems*, 32(4):179–214, 2015.

[190] Bruce Potter. User education – how valid is it? *Network Security*, 2006(4):15–16, 2006.

[191] Sören Preibusch. Guide to measuring privacy concern: Review of survey and observational instruments. *International Journal of Human-Computer Studies*, 71(12):1133–1143, 2013.

[192] Sören Preibusch and Joseph Bonneau. The privacy landscape: Product differentiation on data collection. In *Workshop on the Economics of Information Security (WEIS 2011)*, 2011.

[193] Sören Preibusch, Kat Krol, and Alastair R. Beresford. The privacy economics of voluntary over-disclosure in Web forms. In *Workshop on the Economics of Information Security (WEIS 2012)*, 2012.

[194] Jens Rasmussen. Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models. *IEEE Transactions on Systems, Man, and Cybernetics*, 13(3):257–266, 1983.

[195] Jens Rasmussen. Risk management in a dynamic society: A modelling problem. *Safety science*, 27(2):183–213, 1997.

[196] Kevin Rawlinson. Banks to allow account access using fingerprint tech. `http://www.bbc.co.uk/news/technology-31508932`, 2015. Accessed 02.12.2015.

[197] James T. Reason. *The human contribution: Unsafe acts, accidents and heroic recoveries*. Ashgate Publishing, 2008.

[198] Andrew P. Rebera and Emilio Mordini. Biometrics and ageing: Social and ethical considerations. In Michael Fairhurst, editor, *Age factors in biometric processing*, pages 37–62. The Institute of Engineering and Technology, 2013.

[199] Gary B. Reid and Thomas E. Nygren. The subjective workload assessment technique: A scaling procedure for measuring mental workload. *Advances in psychology*, 52:185–218, 1988.

[200] Karen Renaud. Quantifying the quality of web authentication mechanisms: A usability perspective. *Journal of Web Engineering*, 3(2):95–123, 2004.

[201] Karen Renaud and Wendy Goucher. Health service employees and information security policies: an uneasy partnership? *Information Management & Computer Security*, 20(4):296–311, 2012.

[202] Karen Renaud and Joseph Maguire. Armchair authentication. In *British HCI Group Annual Conference on People and Computers: Celebrating People and Technology*, pages 388–397. British Computer Society, 2009.

[203] Karen Renaud, Melanie Volkamer, and Joseph Maguire. Access: Describing and contrasting authentication mechanisms. In *International Conference on Human Aspects of Information Security, Privacy, and Trust*, pages 183–194. Springer, 2014.

[204] Microsoft Research. Kinect Identity: Player recognition in Xbox. `http://research.microsoft.com/apps/video/default.aspx?id=155583`, 2011. Accessed 27.11.2015.

[205] Gerardo Reynaga, Sonia Chiasson, and Paul C. van Oorschot. Exploring the usability of captchas on smartphones: Comparisons and recommendations. In *NDSS Workshop on Usable Security (USEC 2015)*, 2015.

[206] Alan Richardson. Mental practice: A review and discussion (Part II). *Research Quarterly. American Association for Health, Physical Education and Recreation*, 38(2):263–273, 1967.

[207] Ronald W. Rogers. A protection motivation theory of fear appeals and attitude change. *The journal of psychology*, 91(1):93–114, 1975.

[208] Ronald W. Rogers. Cognitive and physiological processes in fear appeals and attitude change: A revised theory of protection motivation. In John T. Cacioppo and Richard E. Petty, editors, *Social psychophysiology: A sourcebook*, pages 153–176. New York, NY: Guilford, 1983.

[209] Irwin M. Rosenstock, Victor J. Strecher, and Marshall H. Becker. Social learning theory and the health belief model. *Health Education & Behavior*, 15(2):175–183, 1988.

[210] Joel Ross, Andrew Zaldivar, Lilly Irani, and Bill Tomlinson. Who are the turkers? Worker demographics in Amazon Mechanical Turk. *Department of Informatics, University of California, Irvine, USA, Technical Report*, 2009.

[211] Volker Roth, Kai Richter, and Rene Freidinger. A PIN-entry method resilient against shoulder surfing. In *ACM Conference on Computer and Communications Security (CCS 2004)*, pages 236–245. ACM, 2004.

[212] Neil J. Rubenking. AutoComplete for Web Forms – Is It Safe? *PC Magazine*, 19(3):105–108, 2000.

[213] Scott Ruoti, Nathan Kim, Ben Burgon, Timothy Van Der Horst, and Kent Seamons. Confused Johnny: When automatic encryption leads to confusion and mistakes. In *Symposium on Usable Privacy and Security (SOUPS 2013)*, pages 1–19. ACM, 2013.

[214] Marshall David Sahlins. *Stone age economics*. Transaction Publishers, 1972.

[215] Jerome H. Saltzer and Michael D. Schroeder. The protection of information in computer systems. *Proceedings of the IEEE*, 63(9):1278–1308, 1975.

[216] Angela Sasse. Scaring and bullying people into security won't work. *IEEE Security & Privacy*, 13(3):80–83, 2015.

[217] M. Angela Sasse. Usability and trust in information systems. *Trust and Crime in Information Societies*, page 319, 2005.

[218] M. Angela Sasse. Red-eye blink, bendy shuffle, and the yuck factor: A user experience of biometric airport systems. *IEEE Security & Privacy*, 5(3):78–81, 2007.

[219] M. Angela Sasse. "Technology Should Be Smarter Than This!": A Vision for Overcoming the Great Authentication Fatigue. In *Secure Data Management*, pages 33–36. Springer, 2014.

[220] M. Angela Sasse, Sacha Brostoff, and Dirk Weirich. Transforming the 'weakest link': A human/computer interaction approach to usable and effective security. *BT Technology Journal*, 19(3):122–131, 2001.

[221] M. Angela Sasse and Ivan Fléchais. Usable security: Why do we need it? How do we get it? In Lorrie F. Cranor and Simson Garfinkel, editors, *Security and Usability: Designing secure systems that people can use*, pages 13–30. O'Reilly, 2005.

[222] M. Angela Sasse and Iacovos Kirlappos. Familiarity breeds con-victims: Why we need more effective trust signaling. In *Trust Management V*, pages 9–12. Springer, 2011.

[223] M. Angela Sasse and Kat Krol. Usable biometrics for an ageing population. In Michael Fairhurst, editor, *Age Factors in Biometric Processing*, pages 303–320. IET, 2013.

[224] M. Angela Sasse, Matthew Smith, Cormac Herley, Heather Lipford, and Kami Vaniea. Debunking security-usability tradeoff myths. *IEEE Security & Privacy*, 14(5):33–39, 2016.

[225] M. Angela Sasse, Michelle Steves, Kat Krol, and Dana Chisnell. The great authentication fatigue – and how to overcome it. In *International Conference on Cross-Cultural Design, HCI International*, volume LNCS 8528, pages 228–239, 2014.

[226] Florian Schaub, Ruben Deyhle, and Michael Weber. Password entry usability and shoulder surfing susceptibility on different smartphone platforms. In *International Conference on Mobile and Ubiquitous Multimedia*, pages 13:1–13:10. ACM, 2012.

[227] Stuart E. Schechter, Rachna Dhamija, Andy Ozment, and Ian Fischer. The emperor's new security indicators. In *IEEE Symposium on Security and Privacy (S&P 2007)*, pages 51–65, 2007.

[228] Noah Schiffman and Suzanne Greist-Bousquet. The effect of task interruption and closure on perceived duration. *Bulletin of the Psychonomic Society*, 30(1):9–11, 1992.

[229] Bruce Schneier. Key Length and Security. `https://www.schneier.com/crypto-gram/archives/1999/1015.html#KeyLengthandSecurity`, 1999. Accessed 27.11.2015.

[230] Bruce Schneier. *Secrets and Lies: Digital Security in a Networked World*. John Wiley & Sons, 2000.

[231] Tali Sharot. *The Optimism Bias: Why we're wired to look on the bright side*. Hachette UK, 2012.

[232] Kim Bartel Sheehan and Mariea Grubbs Hoy. Dimensions of privacy concern among online consumers. *Journal of Public Policy & Marketing*, pages 62–73, 2000.

[233] Elaine Shi, Yuan Niu, Markus Jakobsson, and Richard Chow. Implicit authentication through learning user behavior. In *Information Security*, pages 99–113. Springer, 2011.

[234] Herbert A. Simon. Rational choice and the structure of the environment. *Psychological Review; Psychological Review*, 63(2):129, 1956.

[235] Sarah Spiekermann, Jens Grossklags, and Bettina Berendt. E-privacy in 2nd generation e-commerce: Privacy preferences versus actual behavior. In *ACM Conference on Electronic Commerce*, pages 38–47. ACM, 2001.

[236] Jared Spool. UIEtips: Deriving Design Strategy from Market Maturity – Part 2. `https://www.uie.com/brainsparks/2009/08/19/uietips-deriving-strategy-2/`, 2009. Accessed 12.11.2015.

[237] Jared Spool. Do users change their settings? `https://www.uie.com/brainsparks/2011/09/14/do-users-change-their-settings/`, 2011. Accessed 12.11.2015.

[238] Lakshmidevi Sreeramareddy, Pewu Mulbah, and Jinjuan Heidi Feng. Investigating the use of gesture-based passwords by the seniors. In *Human Aspects of Information Security, Privacy, and Trust*, pages 107–118. Springer, 2015.

[239] Frank Stajano. Pico: No more passwords! In *Security Protocols XIX*, pages 49–81. Springer, 2011.

[240] Frank Stajano and Paul Wilson. Understanding scam victims: Seven principles for systems security. *Communications of the ACM*, 54(3):70–75, 2011.

[241] Barry M. Staw. Knee-deep in the big muddy: A study of escalating commitment to a chosen course of action. *Organizational behavior and human performance*, 16(1):27–44, 1976.

[242] Michelle Steves, Dana Chisnell, M. Angela Sasse, Kat Krol, Mary Theofanos, and Hannah Wald. Report: Authentication Diary Study. National Institute of Standards and Technology (NISTIR) 7983, 2014.

[243] Elizabeth Stobert and Robert Biddle. The password life cycle: User behaviour in managing passwords. In *Symposium On Usable Privacy and Security (SOUPS 2014)*, pages 243–255. USENIX Association, 2014.

[244] Anselm Leonard Strauss and Juliet M. Corbin. *Basics of qualitative research*, volume 15. Sage. Newbury Park, CA, 1990.

[245] Dennis D. Strouble, G.M. Schechtman, and Alan S. Alsop. Productivity and Usability Effects of Using a Two-Factor Security System. In *Annual Conference of the Southern Association for Information Systems*, 2009.

[246] Lucy A. Suchman. *Plans and Situated Actions: The Problem of Human-Machine Communication*. Cambridge University Press, 1987.

[247] Joshua Sunshine, Serge Egelman, Hazim Almuhimedi, Neha Atri, and Lorrie Faith Cranor. Crying wolf: An empirical study of SSL warning effectiveness. In *USENIX Security*, pages 399–416, 2009.

[248] David G. Taylor, Donna F. Davis, and Ravi Jillapalli. Privacy concern and online personalization: The moderating effects of information control and compensation. *Electronic Commerce Research*, 9(3):203–223, 2009.

[249] Richard H. Thaler and Cass R. Sunstein. *Nudge: Improving decisions about health, wealth, and loss aversion*. London: Penguin, 2009.

[250] Mary Theofanos, Ross Micheals, Jean Scholtz, Emile Morse, and Peter May. Does habituation affect fingerprint quality? In *Extended Abstracts on Human Factors in Computing Systems (CHI 2006)*, pages 1427–1432. ACM, 2006.

[251] Shari Trewin, Cal Swart, Larry Koved, Jacquelyn Martino, Kapil Singh, and Shay Ben-David. Biometric authentication on a mobile device: A study of user effort, error and task disruption. In *Annual Computer Security Applications Conference*, pages 159–168. ACM, 2012.

[252] Janice Y. Tsai, Serge Egelman, Lorrie Cranor, and Alessandro Acquisti. The effect of online privacy information on purchasing behavior: An experimental study. *Information Systems Research*, 22(2):254–268, 2011.

[253] James Turland, Lynne Coventry, Debora Jeske, Pam Briggs, and Aad van Moorsel. Nudging towards security: Developing an application for wireless network selection for Android phones. In *British HCI Conference*, pages 193–201. ACM, 2015.

[254] Joseph Turow. *Americans and online privacy: The system is broken*. Annenberg Public Policy Center, University of Pennsylvania, 2003.

[255] Amos Tversky and Daniel Kahneman. Judgment under uncertainty: Heuristics and biases. In Dirk Wendt and Charles Vlek, editors, *Utility, Probability, and Human Decision Making: Selected Proceedings of an Interdisciplinary Research Conference, Rome, 3–6 September, 1973*, pages 141–162. Springer Netherlands, Dordrecht, 1975.

[256] Sebastian Uellenbeck, Markus Dürmuth, Christopher Wolf, and Thorsten Holz. Quantifying the security of graphical passwords: the case of android unlock patterns. In *ACM Conference on Computer and Communications Security (CCS 2013)*, pages 161–172. ACM, 2013.

[257] United Press International (UPI). UPI Poll: Concern on health privacy, 2007.

[258] US Census Bureau. Tabulation and Processing. `http://www.census.gov/history/www/innovations/technology/tabulation_and_processing.html`, 2015. Accessed 08.08.2015.

[259] Sarah Uzzaman and Steve Joordens. The eyes know what you are thinking: eye movements as an objective measure of mind wandering. *Consciousness and cognition*, 20(4):1882–1886, 2011.

[260] Paul C. van Oorschot and Julie Thorpe. On predictive models and user-drawn graphical passwords. *ACM Transactions on Information and System Security (TISSEC 2008)*, 10(4):5, 2008.

[261] Viswanath Venkatesh, Michael G. Morris, Gordon B. Davis, and Fred D. Davis. User acceptance of information technology: Toward a unified view. *MIS Quarterly*, pages 425–478, 2003.

[262] Michael A. Vidulich. The use of judgment matrices in subjective workload assessment: The subjective workload dominance (SWORD) technique. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 33, pages 1406–1410. SAGE Publications, 1989.

[263] Nikolai Vorkinn. Ethical password monitor. Master's thesis, Department of Computer Science, University College London, 2015.

[264] W3C, Ian Hickson. HTML5. A vocabulary and associated APIs for HTML and XHTML, Section 4.10 Forms, 2011. W3C Working Draft – 25 May 2011.

[265] Yang Wang, Pedro Giovanni Leon, Kevin Scott, Xiaoxuan Chen, Alessandro Acquisti, and Lorrie Faith Cranor. Privacy nudges for social media: An exploratory Facebook study. In *International Conference on World Wide Web*, pages 763–770, 2013.

[266] Steven Ward, Kate Bridges, and Bill Chitty. Do Incentives Matter? An Examination of On-line Privacy Concerns and Willingness to Provide Personal and Financial Information. *Journal of Marketing Communications*, 11(1):21–40, 2005.

[267] Rick Wash. Folk models of home computer security. In *Symposium on Usable Privacy and Security (SOUPS 2010)*, pages 1–16. ACM, 2010.

[268] Rick Wash, Emilee J. Rader, Kami Vaniea, and Michelle Rizor. Out of the loop: How automated software updates cause unintended security consequences. In *Symposium on Usable Privacy and Security (SOUPS 2014)*, pages 89–104, 2014.

[269] Luc Wathieu and Allan Friedman. An empirical approach to understanding privacy valuation. *HBS Marketing Research Paper*, (07-075), 2007.

[270] Catherine S. Weir, Gary Douglas, Martin Carruthers, and Mervyn Jack. User perceptions of security, convenience and usability for ebanking authentication tokens. *Computers & Security*, 28(1), 2008.

[271] Catherine S. Weir, Gary Douglas, Tim Richardson, and Mervyn Jack. Usable security: User preferences for authentication methods in ebanking and the effects of experience. *Interacting with Computers*, 22(3), 2010.

[272] Matt Weir, Sudhir Aggarwal, Michael Collins, and Henry Stern. Testing metrics for password creation policies by attacking large sets of revealed passwords. In *ACM Conference on Computer and Communications Security (CCS 2010)*, pages 162–175. ACM, 2010.

[273] Ryan West. The Psychology of Security. *Communications of the ACM*, 51:34–40, 2008.

[274] Alan F. Westin. Privacy and freedom. 1967. *New York, NY: Atheneum*, 1970.

[275] Tara Whalen and Kori M. Inkpen. Gathering evidence: Use of visual security cues in web browsers. In *Proceedings of Graphics Interface 2005*, pages 137–144. Canadian Human-Computer Communications Society, 2005.

[276] Alma Whitten and J. Doug Tygar. Why Johnny Can't Encrypt: A Usability Case Study of PGP 5.0. In *USENIX Security*, pages 169–184, 1999.

[277] Gillian M. Wilson and M. Angela Sasse. Listen to your heart rate: Counting the cost of media quality. In *Affective interactions*, pages 9–20. Springer, 2000.

[278] Michael S. Wogalter. Communication-Human Information Processing (C-HIP) Model. *Handbook of warnings*, pages 51–61, 2006.

[279] Luke Wroblewski. *Web form design: Filling in the blanks*. New York, NY: Rosenfeld Media, 2008.

[280] Min Wu, Robert C. Miller, and Simson L. Garfinkel. Do security toolbars actually prevent phishing attacks? In *SIGCHI Conference on Human Factors in Computing Systems (CHI 2006)*, pages 601–610, 2006.

[281] Jeff Yan and Ahmad Salah El Ahmad. Usability of CAPTCHAs or usability issues in CAPTCHA design. In *Symposium on Usable Privacy and Security (SOUPS 2008)*, pages 44–52. ACM, 2008.

[282] Yulong Yang, Janne Lindqvist, and Antti Oulasvirta. Text entry method affects password security. In *Learning from Authoritative Security Experiment Results (LASER 2014)*, pages 11–20. USENIX Association, 2014.

[283] Bluma Zeigarnik. Das Behalten erledigter und unerledigter Handlungen. *Psychologische Forschung*, 9:1–85, 1927.

[284] George Kingsley Zipf. Human behavior and the principle of least effort. 1949.

Glossary

Authentication is the act of proving that the user is who they claim they are. Often, the term is used synonymously with 'login'.

Disclosure refers to the act of sharing information. In this dissertation, the term 'over-disclosure' is also used, it describes a situation where a user provides more information than what is mandatory to submit a webform.

Effort describes the amount of work that an individual has to complete in order to achieve a goal. In this dissertation, 'effort' is used synonymously with 'workload'.

Heuristics are simple decision-making rules that enable fast and frugal decision making. The decisions are not guaranteed to be perfect or optimal but the aim is for them to be good enough in the given circumstances [90].

Information Sensitivity refers to how vulnerable the revelation of a certain piece of information would make an individual.

Privacy describes an individual's right and/or ability to protect their personal information [11]. It can also be summarised as the right and/or ability to determine flow of information mediated by social norms and governed by context [172, 119]. This term has a particularly rich history of being difficult to define. Gürses [103] provided a comprehensive overview of the different definitions of privacy. Based on these definitions, she identified three research paradigms into privacy: privacy as confidentiality (hiding), privacy as practice (identity construction) and privacy as control (informational self-determination).

Security refers to a state where a system and/or data is free from compromise. When speaking of user security, this dissertation means their assets (e.g., devices, accounts) are protected from interference. Traditionally, three aspects of security are mentioned: confidentiality, integrity and availability [93]. *Confidentiality* refers to the prevention of unauthorised access of information. *Integrity* describes the prevention of unauthorised changing of information. Data must be accurate, complete, timely and valid [39]. Finally, *accessibility* refers to the fact that a system has to be available to legitimate users when they need it. This aspect, thus, refers to *"the prevention of unauthorised withholding of information or resources"* [93, p. 34].

Two-Factor Authentication refers to logging in users with the use of two factors where factors can be based on knowledge (e.g., password), possession (e.g., access card) or a biometric (e.g., fingerprint). For example, banking customers wanting to withdraw money from a cash machine are usually asked to authenticate with their bank card which is a possession-based factor and a four-digit

PIN which is a knowledge factor. In the study of authentication in an organisation (Section 4.2), for certain logins, employees are asked to provide their username, a PIN and a one-time password (OTP) generated by the SecureID token.

Usability This dissertation adopts the ISO 9241-11 [123] definition of usability. According to this definition, usability is the *"extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use"*(p. 2).

Verification In this dissertation, 'verification' is used to describe the process of a user verifying they are human in order to access services or buy goods online. The most commonly used term for it is CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart) but also the acronym HIPS (Human Interaction Proofs) has been used in the literature [52]. In this dissertation, the term 'verification' was chosen instead as one of the mechanisms studied, NoBot, might not be considered to be a CAPTCHA.