Usability Engineering for Code-based Multi-factor Authentication

Graeme Stuart Roy

A thesis submitted for the
Degree of Doctor of Philosophy

The University of Edinburgh

2013

Abstract

The increase in the use of online banking and other alternative banking channels has led to improved flexibility for customers but also an increase in the amount of fraud across these channels. The industry recommendation for banks and other financial institutions is to use multi-factor customer authentication to reduce the risk of identity theft and fraud for those choosing to use such banking channels. There are few multi-factor authentication solutions available for banks to use that offer a convenient security procedure across all banking channels. The CodeSure card presented in this research is such a device offering a convenient, multi-channel, two-factor code-based security solution based on the ubiquitous Chip-and-PIN bank card. In order for the CodeSure card to find acceptance as a usable security solution, it must be shown to be easy to use and it must also be easy for customers to understand what they are being asked to do, and how they can achieve it. This need for a usability study forms the basis of the research reported here.

The CodeSure card is also shown to play a role in combating identity theft. With the growing popularity of online channels, this research also looks at the threat of phishing and malware, and awareness of users about these threats. Many banks have ceased the use of email as a means to communicate with their customers as a result of the phishing threat, and an investigation into using the CodeSure card's reverse (sender) authentication mode is explored as a potential solution in regaining trust in the email channel and reintroducing it as a means for the bank to communicate with its customers.

In the 8 experiments presented in this study the CodeSure card was rated acceptably high in terms of mean usability. Overall, the research reported here is offered in support of the thesis that a usable security solution predicated on code-based multi-factor authentication will result in tangible improvements to actual security levels in banking and eCommerce services, and that the CodeSure card as described here can form the basis of such a usable security solution.

Declaration of Originality

Date: January 8th, 2013

This thesis is submitted for the Degree of Doctor of Philosophy. I declare that it has been composed by myself and that the work described was my own research.

Graeme S. Roy

Acknowledgements

I would like to thank Professor Mervyn Jack for his guidance, experience and support. Thanks must also go to Lloyds Banking Group, Visa Europe and Emue Technologies for their support during this research.

I would also like to express my thanks to Dr. Nancie Gunson, Dr. Gary Douglas, Nicholas Anderson and Diarmid Marshall for their constructive comments and assistance during the course of this research.

Finally, I would like to thank my dear wife Lynsey for giving me the support, confidence and encouragement I needed to finish this work, and Thomas, Alasdair and Fiona for being patient when their daddy was busy.

List of Abbreviations

ANOVA	Analysis of Variance
API	Application Programming Interface
ASP	Active Server Pages
ATM	Automated Telling Machine
BHO	Browser Helper Object
CAPTCHA	Completely Automated Public Turing test to tell Computers and Humans Apart
CGI	Common Gateway Interface
CNP	Card Not Present
COM	Component Object Model
CSC	Card Security Code
CVV	Card Verification Value
DCCN	Disposable Credit Card Number
DNS	Domain Name System
DoS	Denial of Service
EMV	Europay, Mastercard and Visa
GUI	Graphical User Interface
HTML	Hypertext Markup Language
HTTP	Hypertext Transfer Protocol
I-DT	Dispersion Threshold Identification
IE	Internet Explorer

IIS	Internet Information Services
IMAP	Internet Message Access Protocol
IP	Internet Protocol
IQ	Intelligence Quotient
IVR	Interactive Voice Response
LCD	Liquid Crystal Display
LED	Light-Emitting Diode
OTP	One-time Passcode
PAN	Primary Account Number
PIN	Personal Identification Number
POS	Point of Sale
SAW	Surface Acoustic Wave
SD	Standard Deviation
SDK	Software Development Kit
SET	Secure Electronic Transaction
SIM	Subscriber Identity Module
SMS	Short Message Service
SMTP	Simple Mail Transfer Protocol
SOAP	Simple Object Access Protocol
SSL	Secure Sockets Layer
TAN	Transaction Authentication Number
TCP	Transmission Control Protocol
TLS	Transport Layer Security
URL	Uniform Resource Locator
XML	Extensible Markup Language

List of Figures

List of Tables

Table of Contents

Chapter 1. Introduction

The thesis expounded in this research is that a usable security solution predicated on code-based multi-factor authentication will result in tangible improvements to actual security levels in banking and eCommerce services, and that the CodeSure card as described here can form the basis of such a usable security solution. The research reported here is presented as an extensive study into usability engineering aspects of such a usable security solution based around the CodeSure card.

Since the beginning of the 21st Century there has been a major increase in the use of online banking, with 6 in 10 of those who are connected to the Internet now checking their bank accounts online[1]. This is coupled with an increase in the number of banking 'channels' available for customers to use. While this has given more flexibility to their customers, it has also led to a corresponding increase in fraud across all such channels, leading to a perceived trade-off between convenience and security.

[1] Source: Ipsos (http://ipsos-na.com/news-polls/pressrelease.aspx?id=5573, 2012).

Early adopters of these new banking channels were technology enthusiasts, computer-savvy individuals who recognised the benefits of the new technologies but were also aware of the risks. Nowadays, customers have a broader range of competencies and many are not aware of the risks, or at least not completely aware if the responsibility for security lies with them or the bank. To reduce the risk of identity theft and fraud, it is now recommended that financial institutions use multi-factor (at least two-factor) authentication, where the factors should be two different aspects of what customers have, what customers know and what customers are.

Although there are many multi-factor authentication solutions currently being adopted, few offer a convenient security procedure across all banking channels, most being tailored for, or suited to, a limited set of channels. The CodeSure card presented in this research is a recent innovation that offers a convenient multi-channel, two-factor security solution based on the familiar, ubiquitous Chip-and-PIN bank card, modified to include extra electronics and functionality for code-based multi-factor security.

Very little research has been reported on the usability of such multi-factor security devices and in order for the CodeSure card to find acceptance, it must be shown to be easy to use and it must also be easy for customers to understand what they are being asked to do, and how they can achieve it. This need for a usability study forms the basis of the research reported here.

With the growing popularity of online channels, this research also looks at the growing threat of phishing and malware, and awareness of users about these threats. Email communication, for example, from banks has almost disappeared due to the prevalence of phishing emails and the resulting mistrust from customers. Indeed, many banks are now telling their customers that they will never communicate to them by email and if they receive any such emails then they must be fraudulent. Having reverse (sender) authentication where the bank authenticates itself with the user would perhaps reintroduce a feeling of trust where it no longer exists. The CodeSure card with its reverse authentication mode, and its inherent convenience, offers a solution for this problem.

1.1. Contribution

Although there have previously been attempts to define a methodology for assessment of usability of security systems, these have achieved only limited success. The work reported here contributes to knowledge with a detailed study and investigation into the usability assessment of the CodeSure card as the basis for a usable security solution. This research presents a usability assessment methodology based on the ISO definition (effectiveness, efficiency and satisfaction) along with measurements of ease of use, convenience, perceived security and associated performance metrics.

The main contribution to knowledge arising from this research is the empirical evaluation of a novel multi-factor and multi-channel authentication device in the domain of usable security. This exhaustive study of 8 experiments demonstrates for the first time the high usability scores for the CodeSure card and evidence that customers appreciate the need for a usable security solution. It also highlights the importance of sender authentication technology when using a device as a second factor for verification.

This work also offers new findings concerning the conflict between usability and perceived security with respect to multi-factor authentication, showing the limitations of social engineering approaches in combating phishing and, unlike other studies using students, is based on large scale experiments with banking customers.

1.2. Outline

Chapter 2 provides the background to this research. It begins with an introduction to banking channels and the problems associated with achieving their security. It includes an overview of authentication methods and the need for multi-factor authentication, followed by an introduction to the CodeSure card as a unique candidate for a two-factor solution across all banking channels. Malware and phishing are introduced with an emphasis on identity theft, and some of the various techniques used by fraudsters today are detailed. This is concluded with an introduction to usability engineering, along with some background on defining and measuring usability and the methodology that is used in usability experiments. There is also a description of gaze tracking, an increasingly popular data collection technique used in usability experiments.

In Chapter 3 a set of 3 experiments are described that investigate the usability of the CodeSure card. The first experiment investigates the usability of the card in a longitudinal study over 10 uses of passcode generation. Based on the findings of that experiment, the second experiment uses a simulation of a CodeSure card running on a tablet device to investigate some of the usability issues identified in the first experiment, comparing the baseline design (used in the first experiment) with an enhanced design. The third experiment uses a different simulation of the CodeSure card with a gaze tracking camera to investigate gaze behaviour during use of the CodeSure card.

Chapter 4 presents a usability experiment that investigates the use of the CodeSure card for multi-channel authentication, comparing use of the CodeSure card with use of the existing disparate security processes across three different channels: online shopping, Internet banking and automated telephone banking.

For Chapter 5, the problems of identity theft in online channels and the use of the CodeSure card in reverse authentication are investigated with a further 4 experiments. The first experiment looks at phishing awareness and the propensity for participants to click on fraudulent emails. The second experiment then investigates the role that education can play in reducing fraud rates. The third experiment investigates participant reaction to anti-malware software as a measure of protection that banks can offer customers during Internet banking sessions and 3 levels of intrusiveness of anti-malware software are investigated to see which participants preferred and trusted the most. The fourth experiment investigates the reverse / sender authentication capabilities of the CodeSure card for 3 banking channels.

Finally, Chapter 6 concludes with a discussion of the research findings and suggestions for future work.

Chapter 2. Research Background & Methodology

Historically, customers have had a limited choice of ‘channels’ with which to make contact with their bank or access their money: using the postal service, making a telephone call or interacting face-to-face in a branch with bank staff. Since the beginning of the 21st century, the number of banking channels has expanded significantly with the proliferation of Automated Teller Machine (ATM) facilities, automated telephone banking, text messaging, mobile phone banking (mBanking) and Internet banking (eBanking or online banking) (Centeno, 2004). Widespread adoption of new consumer channels such as Internet shopping (eCommerce) and mobile phone shopping (mCommerce) by consumers has also led to a massive increase in the number of (Card Not Present, CNP) Internet transactions. In 2012, more than 60%[2] of calls to the bank were handled by an automated service and more than 90%[3] of all consumer purchases on the Internet were CNP transactions. The proliferation of these new channels has resulted in a measured increase in fraud and

[2] Private communication with Lloyds Banking Group.

[3] Source: Visa Europe.

there is growing pressure on banks to offer customers improved security when accessing modern banking channels to minimise the impact of customer identity theft.

2.1. eBanking and Security

Offering more channels for customers to do banking and payment transactions presents an opportunity for banks to give their customers more flexibility in accessing their accounts whilst reducing business costs and in some cases deriving competitive advantage. However, use of such 'remote' or 'direct' channels comes at a price in terms of security. The perceived trade-off between customer convenience and security is at the heart of the issues facing banks today (Lichtenstein and Williamson, 2006; Weir et al., 2009) in their attempts to migrate customers towards more widespread use of automated services and their associated fight against the criminals who seek to take advantage of this. A study of 23,000 Internet users in Europe found that 40% are holding back from using online banking due to security worries (Ranger, 2005). An additional problem for banks is that having multiple channels means that there are more areas where they are vulnerable to attack and criminals will find the weakest link for their attacks (Jowitt, 2011). Criminal gangs can fraudulently obtain a targeted customer's account details and security details via email contact (phishing) or via telephone contact (vishing) and then use these details to gain access to the targeted customer's accounts via the bank's Internet banking site or telephone banking service.

It is not always clear who is responsible for losses occurring with (Internet) banking. In the USA a bank refused to accept any responsibility when a user's computer was infected with a key-logger trojan[4] and criminals stole $90,000 from the account (Henry, 2006). In Canada, where banks have a 100% reimbursement guarantee against online fraud, one study (Manna and van Oorschot, 2007) showed that 85% of users were unable to state any of the conditions required for the guarantee to be

[4] A trojan is a type of malicious software (malware) that masquerades as a legitimate file or helpful program with the ultimate purpose of granting a criminal unauthorized access to a computer.

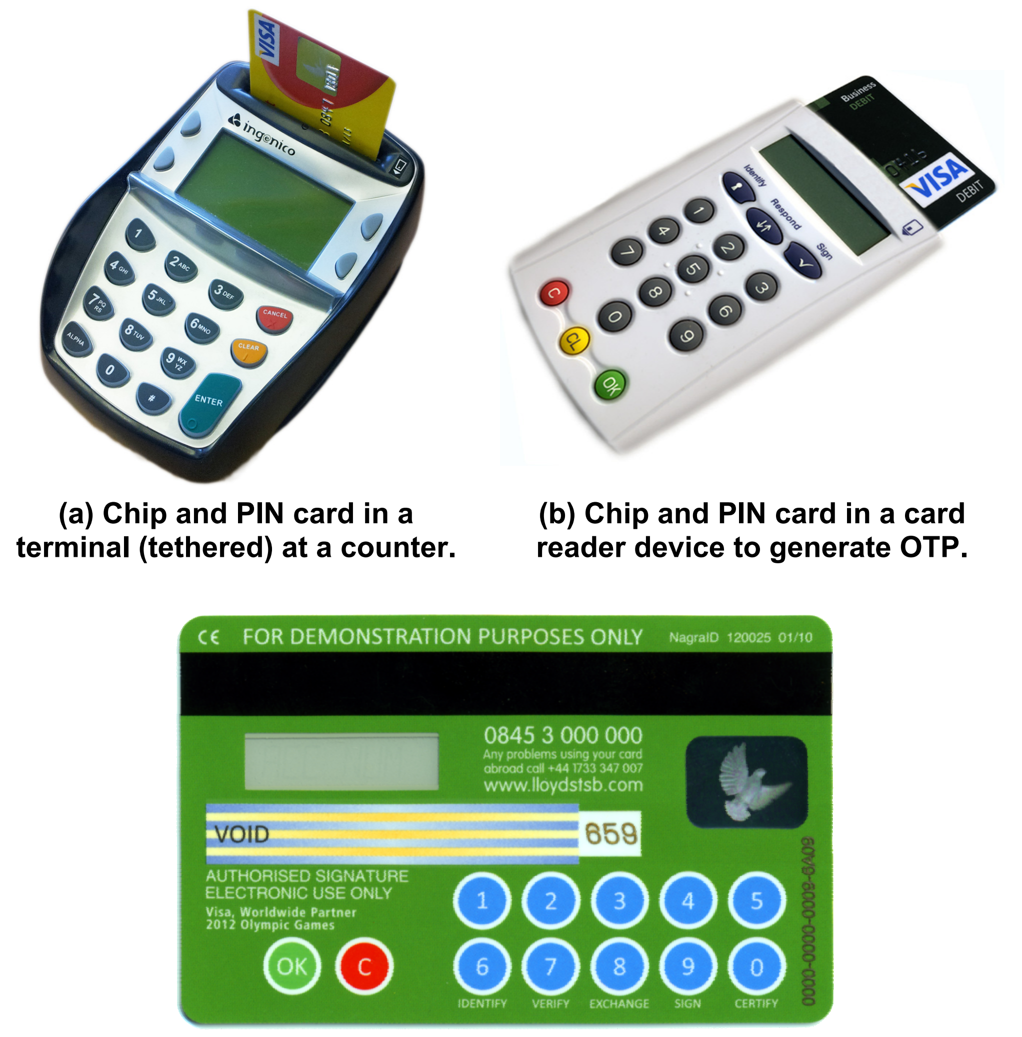
upheld. The UK has a similar 100% reimbursement policy. In China, victims must prove that the fault lies with the bank before they get any reimbursement (Manna and van Oorschot, 2007).

The early adopters of Internet banking were technology enthusiasts and computer-savvy individuals. With banks now pushing Internet banking to all of their customers to save on costs and provide added convenience, there is a much wider customer base, and therefore wider range of customer competencies, that needs to be catered for. The greatest business challenge is that of security; banks often have unrealistic expectations of the awareness of users when it comes to security (Manna and van Oorschot, 2007), specifically in terms of password policies, keeping security software up to date and reading banking agreements.

The role of a security procedure in a banking system is to prevent unauthorised individuals from accessing the system whilst allowing access for the genuine customer. Customer authentication (O'Gorman, 2003; Renaud, 2005) involves firstly identification, where a user will typically claim an identity by name or account number, or via a previously registered username or customer code. This is followed by a verification process which is the most critical and in which users can authenticate themselves by one or more of three factors (Smith, 2001): what they have (an object or device owned by or assigned to the user by the bank, such as a credit or debit card, a mobile phone or a token generator); what they know (a security token which is known only to the user and the bank, such as a password); or what they are (a measurable property of a user – a biometric, for example a fingerprint, a palmprint or a voiceprint).

Using only one of these factors to perform customer authentication is no longer recommended in online banking due to concerns about fraud, and financial services authorities and the security industry are now recommending the use of multi-factor authentication (Beaumier, 2006; FFIEC, 2005; Henry, 2006; Viega, 2005). By using more than one factor in authentication, the impact of the security limitations in each can be minimised (Renaud, 2005).

Interestingly, use of the ATM has required two-factor authentication since its introduction in the early 1960s. Customers are required to know a secret PIN number and have their ATM card in their possession. Since 2004, the introduction of Chip and PIN technology in the UK takes this same two-factor authentication approach (know a PIN and have a card) at counters in shops and in bank branches (Figure 2.1a) and has served to reduce fraud in CNP transactions. In the period 2004 to 2010, domestic fraud annual losses on UK-issued (Chip and PIN) cards fell by some 34% (King, 2012).



(a) Chip and PIN card in a terminal (tethered) at a counter.

(b) Chip and PIN card in a card reader device to generate OTP.

(c) The CodeSure card integrates OTP generation in the card.

Figure 2.1: The CodeSure Card for Code-based Security

The security limitations of a fixed PIN for a card have been addressed by a variety of solutions. One time passcodes (OTPs) can be used for customer or transaction authentication purposes. Several European banks send customers printed lists of Transaction Authentication Numbers (TANs) to use for authentication (Hiltgen et al., 2006; Reavley, 2005). After logging into their account (with information they know), customers must also enter one of their TANs (information they have) as an extra level of security against fraud. To provide additional safeguard against phishing attacks, banks also ask for a specific TAN in the list rather than just any one since once a TAN is used it can no longer be used for future transactions.

Rather than provide printed TANs to customers, banks are now moving towards the use of digital token generators which are able to automatically generate OTPs based on encryption algorithms aligned to the customer's account; or involve the use of a portable card reader (Figure 2.1b) requiring the customer to insert their bank card in order to generate an OTP. One study investigated the usability of digital OTP generators (Weir et al., 2009) and found that users placed higher importance on the convenience of such devices when compared to the increased security that they offer.

The disadvantages with OTP generators are that customers may end up acquiring several such devices for their various accounts and that these devices are costly for banks to develop and supply to customers (Claessens et al., 2002). An alternative is for customers to use a device that they already own, such as a mobile phone or mobile tablet for authentication purposes but these have security issues in that the banks have no control over any software applications installed on such a device.

Whilst two-factor (PIN based) OTP generators are recognised to be more secure than single-factor authentication alternatives, they can be inconvenient for customers to have more than one such device for different purposes. In addition, such devices are bulky and users have to plan in advance to take them to where they will be required. In contrast, the CodeSure card[5] (Figure 2.1c) bases two-factor customer authentication on an OTP generator integrated within the customer's bank card, and is activated by the card PIN.

The CodeSure card is a standard Chip and PIN debit or credit card on the front face whilst on the back face it presents a small keypad and an e-Ink display used for two-factor customer authentication and transaction signing (Figure 2.1c). The advantages of the CodeSure card are in customer convenience, requiring the customer to

[5] http://www.emue.com

remember only their card PIN number for all card transactions both in the physical world and in the digital world, and for common access to all bank channels; and in improved security, as a two-factor authentication device which is always to hand. The high levels of ‘usable security’ achievable with use of the CodeSure card promise an improvement in overall security levels since the convenience of having the card always to hand will serve to encourage positive customer engagement in security procedures.

The CodeSure card supports three main authentication modes: ‘Identification’, the generation of a one-time passcode to authenticate the user of the card; ‘Verification’, the input of a bank-supplied number to verify the authenticity of a service; ‘Challenge Response’, a combination of verification followed by identification. In addition, OTP generation can depend on additional variables such as a bank-generated reference number or transaction amount which can be used in verifying (‘Signing’) transactions.

Existing banking and eCommerce applications can be extended to support the CodeSure card by making a secure connection to a CodeSure card server, typically hosted by the issuer of the card. This holds a record of each individual card and its current PIN and is used to validate generated OTPs, generate challenge codes and handle maintenance tasks.

With the CodeSure card, one-time passcodes are implemented using a ‘moving window’ of codes rather than being time-dependent. The card server has a look-ahead list of expected codes for each CodeSure card which allows for a certain number of OTPs to be generated without transmission to the server[6]. If the number of untransmitted generated OTPs exceeds the window size on the server then the OTP will be rejected and the card will have to be re-synchronised with the server before it can be used further. This involves asking the user of the card to send 3 generated OTPs to the server so that the server can look ahead and re-compute a new window of codes.

[6] Typically around 50 to 100.

Choosing the most appropriate security procedure for customer authentication to achieve high levels of ‘usable security’ for a banking channel represents a challenge. There are three strategic options that exist in banking. Banks can use what is relevant and available for a channel: for example, fingerprints in a branch, voiceprints over the phone and passcodes over the Internet. Or they can rely on user or system forensics such as biometrics or IP address. Finally, they can use a code-based approach like Chip and PIN which can work across all channels.

The use of channel-specific security procedures has been tried and failed and the role of forensics continues to develop but is unlikely to gain user acceptance in the short term (Coventry, 2005). The research reported here addresses code-based approaches to customer and transaction authentication, an area where there has been considerable development but very few actual scientific usability studies (Piazzalunga et al., 2005).

Two critical areas that need to be addressed by banks are identified in this research. The first is the need to create usable security procedures, as discussed in Section 2.2; and the second is the need to help customers secure their security information from the possibility of identity theft in fraudster attacks by phishing and malware, as discussed in Section 2.3.

2.2. Usable Security

The idea of ‘usable security’ detailed above in terms of the CodeSure card is the key point of departure for the research reported here.

The premise behind the concept of usable security is that by furnishing customers with a familiar device, a familiar form factor, to be used in a familiar, common modality for all purchases, financial transactions and account access in the physical world as well as in the digital world, this familiarity will demolish the existing barriers where channel security procedures are seen by customers as being an imposition, thereby achieving improved levels of actual security because security procedures with usable security will become a natural part of everyday financial activity.

In online banking and commerce, as customer authentication moves from using a single fixed PIN or password to employing alternative or additional security procedures, security vulnerabilities often become evident in the new technologies (Braz and Robert, 2006; Du et al., 2011; Furnell, 2007). When new threats appear, security procedures need to evolve rapidly to counteract these threats and the need for usable security can be forgotten.

2.2.1. Designing for Usable Security

Problems with usability have been contributors to many high-profile security failures (Johnson and Willey, 2011). However, designing for usability must be balanced with achieving desired levels of security. In a recent case, a UK high street bank introduced an 'Instant Access Service' which was claimed to *"make it easier and more convenient for customers to log in to their account online"*. Using the service, customers were no longer required to remember a password and only required one of four items (their surname, date of birth, 16-digit card number or 3-digit card security code) to authenticate themselves. Since these could be considered public knowledge, this left a security vulnerability which fraudsters could easily exploit, resulting in the rapid withdrawal of the service (Smyth, 2010).

A common misconception regarding usable security is that a complicated security procedure will be inherently more secure than one that is easier to understand because criminals are more likely to target an easy banking channel than one that is harder to break into. However, this simplified philosophy fails to take into account the value of the data that is being stored, and more crucially does not take into consideration the behaviour of customers who would normally be using that channel and the human factors that are involved.

It was generally believed that designing for security was directly at odds with designing for usability (Kainda et al., 2010), and whilst it is true that the goals of a security procedure are often at odds with the goals of usability, failing to achieve a balance in a usable security procedure will either result in a secure banking channel that cannot be used or a usable banking channel that is insecure. Usability issues can lead to security problems (Piazzalunga and Salvaneschi, 2006) so making the effort to achieve the correct design balance is key to the design of a usable security procedure (Tognazzini, 2005). Improving the user experience will often improve the effectiveness of security procedures.

Several reasons have been identified why security procedures are often rendered ineffective due to the behaviour of users (Sasse and Flechais, 2005). Firstly, users often do not understand the importance of what is being protected or the importance of following the required security procedures, especially if they prefer to use an alternative procedure that is (in their view) just as secure. Further, they may not believe that what is being protected is at risk or they do not understand that their behaviour as users puts what is being protected at risk. And finally, they often have problems using the security procedures correctly.

In terms of banking, users will generally understand the need to protect their assets (their money) and accept that there will be certain security procedures being demanded since users are aware of the presence of criminals and threats. In banking, user behaviour in terms of prevention of identity theft can, to a certain extent, be overcome through education (addressed in Chapter 5).

Design of a usable security procedure requires (Sasse and Flechais, 2005) emphasis in ensuring that a user is not the weakest link by addressing the human factors issues surrounding the security procedure, the security context in which the channel and the user operate, and by educating the user to understand the importance of the security threat environment. A design balance of usability with security must be part of the whole solution, and not regarded as something that can be added on to a finished product (Yee, 2005).

The willingness of users to engage with a usable security procedure and to make the extra effort to be security conscious plays a vital part in the effectiveness of a channel (Weirich and Sasse, 2001). Users cannot be forced to be security-conscious; they have to be persuaded, typically by designing persuasiveness into systems and procedures. User resistance can also play a large part in the poor acceptance of security procedures (Schultz et al., 2001).

Usable security mandates that the presence of a security procedure should not make it harder to access a resource or perform an action than it would be if the security procedure were absent (Bishop, 2005). Although much research is focused on ease of use and clarity, the convenience of a security procedure also has an impact on how secure a user perceives it to be (Halderman et al., 2005; Hertzum et al., 2004). This is especially true for biometric authentication at an ATM (Coventry et al., 2003).

If a highly secure procedure is actually difficult to use, users will tend to move their account to more usable alternatives which may ultimately be less secure, or take security shortcuts to make their life easier, thus rendering the system less secure (Besnard and Arief, 2004; Johnston et al., 2003; Witt and Kuljis, 2006). Most users are security conscious but only as long as they understand or can see a need for it. Even if they accept a security procedure and continue to use it, efficiency will be reduced and the likelihood of user errors will increase (Schultz et al., 2001). Conversely, if a highly usable channel is not secure, it is unlikely to be very long before the security is breached with resulting fraud attack consequences, and it has to be rebuilt or replaced.

However, ensuring a secure banking channel is usable is sometimes difficult to accomplish. For example, the action of communicating security needs to the user can interrupt the flow of the user experience, providing unwelcome interaction between the user and the system. In addition, if users trust the channel they are using, highlighting extra security information about risks may actually seem to induce mistrust in the channel itself (Patrick et al., 2005).

Accessibility, especially for those who are visually impaired, is a major factor to take into consideration when designing usable security procedures (Jahankhani et al., 2010). The World Wide Web Consortium (W3C) has published guidelines for accessible web content making it possible for synthesised speech screen readers and Braille displays to convey the content of web pages to those who need it. However, screen readers have attendant security issues when users are requested to enter (masked) passwords into forms. Whilst some will say simply “star” or “asterisk” for each digit entered (which provides no feedback for the blind user), there are some that will speak each entered digit out loud thereby compromising security if others can hear.

Designers incorporating new security procedures into systems are not only faced with the technical challenge of making the changes required, but must also take the responsibility of designing security into the system: keeping information on a security procedure completely hidden can frustrate most users who wish security to be visible; making the details of a security procedure totally explicit can upset users who do not understand enough about the underlying details. One option is to have a ‘building bricks’ security model that can be easily modified as a system evolves, providing common features for users and making it easy for developers to add (Smetters and Grinter, 2002). An alternative is to have an event-driven model (Dourish and Redmiles, 2002) which presents security events to the user using familiar user interface components.

Since designers often design a security procedure to meet their own models and expectations as opposed to those of the users, security procedures can be more cumbersome and less effective than they should. Human factors design guidelines should be taken into consideration at all stages during the software or hardware life cycle (Karat et al., 2005; Zurko and Simon, 1996). For example, research on usable security focusing on identity theft (Bardzell et al., 2007) has shown that a security procedure will be more usable if the users’ conceptual model of the procedure required is aligned with the capabilities and process flow of that procedure. Users incorrectly assume that ‘secure’ means ‘trustworthy’ but security implies that there will be no unauthorised access to information whereas trust implies that a recipient of such information will not share it with others who are not authorised to read it. That research also noted that users will not comply with security procedures unless the system makes it easy to do so, the user’s perception of the value of complying is high, or the transactions they wish to conduct have a high perceived value. Users lack an adequate understanding of how criminals might be able to violate security and privacy and how quickly new security threats can appear. Most users also do not understand how to interact effectively with a security procedure and believe that they do not know enough about security to be able to make informed judgements on decisions affecting security.

Privacy should be considered separately from security in a usable security procedure and as such should be given separate treatment in the design process (Ackerman and Mainwaring, 2005). A highly secure system might be perceived as being intrusive and infringing on user privacy – for example a security card system that also tracks the owner’s location. Such privacy leakage, or exoformation (Brunk, 2005), can not only infringe upon the personal rights of the user, but may also lead to the leakage of personal information that can itself lead to security breaches.

User trust is also related to security in that a user must believe in the security of a system (Johnston et al., 2003). One study looked at the authentication methods used in banks in Sweden and the UK, in the context of the issues of security and trust (Nilsson et al., 2005). In the UK, where passwords are used to access online banking, using passwords was perceived to be significantly less trustworthy than the use of OTP generating devices which are used in Sweden. While both types of authentication procedures could be argued to be equally secure, the perceived trust in them varied significantly.

2.2.2. Choice of Authentication Method

Choice of authentication method has a central impact on the usable security of a procedure. The simplest and most common method of authentication in a security procedure is the password (Augier, 2007) and whilst the security aspects of passwords have been studied extensively (Zviran and Haga, 1999), the usability of passwords has not been studied in detail (Adams and Sasse, 2005). The most common problem with passwords is that they can be easily forgotten, especially if an individual has many different passwords to remember and associate with the correct system. Strong passwords are hard to remember (Barton and Barton, 1984) and the passwords that users find easiest to remember are also the easiest to guess (Yan et al., 2005).

Personal Identification Numbers (PINs) are a variation of passwords that have a fixed number of digits (typically four) and were originally chosen for use with systems that had only a numeric keypad, such as an ATM. Because of their easy memorability they are still popular with customers today and are used in a variety of authentication systems including Chip and PIN, mobile phone SIMs and most door entry systems. However, for online access they are generally used only as a second factor in two-factor authentication because of their low complexity and customers’ failure to follow good security practices when choosing and using their PINs (Bonneau et al., 2012).

Security procedures that impose tight restrictions on users regarding password strength and have strict policies on the lifetime of passwords may well have less chance of being broken into but in practice most users will undermine the security of the system by writing down their password in an easy to find location. In many cases there are no justifications for having such stringent policies. In one study (Florêncio and Herley, 2010), the password policies of 75 different web sites were analysed: the sites with the most restrictive password policies did not necessarily have greater security concerns than those that had less stringent policies.

Having a less stringent password policy might result in improved popularity with users but using mnemonic phrase-based passwords results in them being easy to guess or crack, given enough time (Kuo et al., 2006). Many users will re-use the same passwords across different accounts (Gaw and Felten, 2006; Ives et al., 2004) leading to potential security breaches in those accounts if the password is obtained for only one of them. One suggestion to improve memorability but also security of passwords is to allow users to propose a password and then shuffle other characters into it (Forget et al., 2008). Many web sites have now taken a slightly different approach to password authentication by asking only for certain characters or digits at varying positions in the password. While this reduces the chance of an eavesdropper obtaining a complete, reusable password, it is more difficult for users to mentally work out which characters or digits to use, and many will resort to writing the full password on a piece of paper, thus compromising security.

One alternative to passwords is the use of challenge questions. These ask the user for a piece of (private) information that is well known to them (and only them) in order to authenticate them. Several such questions can be used in place of passwords if they are deemed to be secure enough, or just one can be used to augment a password or provide one part (knowledge) of a two-factor authentication procedure that sends the forgotten password to their email address (ownership). The difficulty of guessing the answer to the challenge question has to be matched also by the ease of an attacker being able to retrieve or observe (Just, 2005) and the challenge questions themselves also have to be usable (Just and Aspinall, 2009). Another issue that is likely to become more relevant over time is that more and more personal information is becoming available to view online, and so it will be easier for the answers to challenge questions to be determined by fraudsters via phishing attacks (Rabkin, 2008).

An alternative to alphanumeric passwords is using graphical passwords (Bicakci and van Oorschot, 2011; Monrose and Reiter, 2005). These make use of the fact that humans are better at remembering pictures than they can text; and are useful for systems that do not have keyboards. However, if these become too prevalent then users are at risk of confusion, where the graphical password for one account gets confused for one to access another account. Another graphical password system uses distorted images based on an image that the user is familiar with. It is very difficult for an automated attack to bypass this system, and it is also difficult for a human who has not seen the original image to work it out (Hayashi et al., 2008). Rather than using images, CAPTCHAs[7] are distorted words used to authenticate a human but deny access to an automated attack. Despite their popularity with online registrations, they are often hard to read and their usability is poor (Yan and El Ahmad, 2008). CAPTCHAs can also be video based (Kluever and Zanibbi, 2009).

The use of biometrics as a means of authentication has been focused very much on getting the technology to work reliably rather than on the user experience (Coventry, 2005; Toledano et al., 2006). One study looked at using iris verification to authenticate a customer at an ATM (Coventry et al., 2003) and found that 90% of those who used it would prefer iris verification over PIN or signature, perceiving it as being more secure, more reliable and faster. However, this study did not look into using other biometrics technologies at an ATM, such as fingerprint verification.

While biometrics are in use today across a wide range of fields, they have yet to be adopted in online applications. This is mainly due to the need to have a trusted biometrics sensor for each user, something that is costly to do and difficult to maintain trust in over long periods of time. The move towards multi-factor authentication by banks may mean that biometrics will be adopted as part of a larger solution, but even then it is more likely to only be used where the customer is present, in a branch or at an ATM.

[7] Completely Automated Public Turing test to tell Computers and Humans Apart.

It is also acknowledged (Moody, 2004) that public perceptions of biometrics as having police or criminal connotations can hinder their acceptance, so usability research is very important before biometrics are introduced. Previous studies (Coventry et al., 2003; Furnell and Clarke, 2005; Moody, 2004) have shown that only a small percentage of a bank's customer base will have had any experience of biometrics, and their understanding of how a biometric procedure works will be limited. One study (Coventry et al., 2003) found that customers have difficulty believing the technology can work, and fear that it will fail to recognise them, that some biometrics like voiceprints will be easy to defraud, and that some have perceived health risks.

2.3. Identity Theft: Malware and Phishing

Banks continually seek to enhance security for their (Internet) banking channels. The presence of malicious software (or malware) such as trojans and key-loggers on a customer's computer or mobile phone/tablet can compromise the security of the customer's online banking account (by logging account details and passwords) with the intent of enabling the perpetrators of the malware infection to conduct theft from the customer's accounts. Some malware is self-propagating (viruses, worms) while other types are spread by unsuspecting users.

Another threat lies in the prevalence of phishing, an attempt to fraudulently capture a targeted customer's security details by pretending to be a trustworthy sender of an electronic communication, typically an email message (Myers, 2007a) or phone call (vishing).

2.3.1. Phishing

A customer who is deceived in a phishing attack would be directed to a fraudulent web site which would attempt to collect the information the attacker wishes to obtain. Such attacks, with their potential for unseen contamination of the computers that customers use for connection to the Internet banking sites could introduce malware applications which can compromise the security of the customer's computer and can compromise the security of the customer's online bank account (by logging account numbers and passwords) with the intent of conducting fraud (James, 2005; Lininger and Vines, 2005; Myers, 2007a).

Phishing first appeared in the early 1990s when hackers were able to use America Online (AOL) services for free, and registered with AOL using automatically-generated fake credit card numbers (Myers, 2007a). Although this in itself was not considered phishing, when AOL cracked down on the practice by contacting users to verify the credit card numbers already used in registrations, hackers posing as AOL employees attempted to steal real AOL accounts by emailing AOL users. In these phishing emails users were being asked to verify their password, not realising that this was for fraudulent purposes as it came from a legitimate source at AOL.

Phishers have now expanded into mimicking emails purporting to come from large financial institutions with the aim of obtaining credit card numbers and bank account details. Phishers are no longer simply competent hackers – organised crime and terrorism are now listed among those that are organising phishing attacks. Attack campaigns also span multiple countries (Emigh, 2007) and in many cases phishers do not use the stolen details directly but instead sell these on to other criminals.

In 2007, phishing was estimated to be costing US financial institutions in excess of $1 billion a year in direct losses (Emigh, 2007). However, this is not the full extent of the losses as there are other types of costs to consider when looking at the cost of phishing (Myers, 2007a): whilst direct costs represent the actual money that was stolen, indirect costs represent the costs involved in dealing with the attack, both by the victim and the bank / enterprise. In addition there is also the money lost in revenue due to users being unwilling to continue to use an online service due to mistrust and fear of further phishing attacks.

There are typically three key components of a phishing attack (Myers, 2007a). The first component is the ‘lure’, normally a fake / spoofed email that gets sent to a large number of users. Next there is the ‘hook’, which is typically a fake web site that mimics a legitimate web site and is the destination to where the victims are sent by the spoofed email of the lure. Finally there is the ‘catch’, where the attacker uses the fraudulently obtained information in criminal activities.

To make the lure convincing there are two main categories of tricks that phishers employ. Social engineering tricks involve the attacker offering a plausible reason for victims to click on the link in the email and give out their personal information on the website. Examples of social engineering include requiring security upgrades (or enrolment in an anti-fraud program), incomplete account information, financial incentives (Christin et al., 2011) or enticing offers where customers get things for free, problems with accounts (false account updates) or fictitious orders with a link to cancel.

Alternatively, the phishers will use technical tricks such as using copies of logos and images and a corporate font, using email spoofing to pretend the email comes from an authentic address; or will use URL hiding, encoding and matching or using bot nets. They may also use 'cousin domain attacks' where a similar sounding / looking name with a minor spelling change is used (Emigh, 2007), such as '`www.goggle.com`' instead of '`www.google.com`'.

To make the hook convincing the attacker must first mimic the legitimate web pages as faithfully as possible using the legitimate design schema (Myers, 2007a). Two things that cannot be changed however are the legitimate URL and lack of secure HTTP connection, which have to be worked around. This can be done using URL homograph attacks (URL hiding) to mimic the intended URL, such as '`www.paypa1.com`' where the digit '`1`' has been used to replace the letter '`l`'. This is a specialised case of a cousin domain attack and is intended to fool the user into thinking that the URL of a spoofed web site is in fact the URL of the legitimate web site (Fu et al., 2007). It is done by replacing some of the letters or numbers in the legitimate URL with ones that look identical (which is possible when using Unicode) or similar (such as digit '`1`' and letter '`l`', or digit '`0`' and letter '`O`') in the URL of the fraudulent web site.

The appearance of the user's web browser can also be modified (Raskin, 2007) using JavaScript or other web technologies to make it look as though there is a valid security certificate icon, change the URL which is displayed, or modify the rendered page. This is increasingly more difficult to do as web browsers become more security-aware with measures in place to prevent most of this behaviour, but there will no doubt be more loopholes uncovered.

Another way to make the hook convincing is for the attacker to make a new self-signed certificate that is not issued by a Certificate Authority. Browsers will warn that this is the case, but most of the time users ignore such warnings and just click through (Miller and Wu, 2005). However, rather than use elaborate technical tricks, some phishing attacks actually do not attempt to cover up their fraudulence, relying on the user failing to notice or ignoring the security warnings from the browser.

A phisher will not want to use their own computer to send out phishing emails or host the web site that victims will be directed to. For the emails, a bot net[8] will typically be used or a server with a compromised email server. For the web site, a computer must be hacked into and a new web server installed containing a copy of the spoof web site. The stolen credentials obtained from the web site will typically be posted onto Usenet newsgroups or web news and encrypted or obfuscated to prevent authorities tracing the attacker and anyone else from using the collected information. There are actually various tools available to a hacker to do all of this: a 'rootkit' is primarily used to obtain and maintain administrative privileges on a machine, employing various techniques to prevent discovery; a 'phishing kit' is used to set up the web server and web site on a compromised machine.

Another technique related to phishing, which eliminates the lure component is called pharming (Gupta, 2007a). It uses DNS[9] manipulation to map web site names to IP addresses of fraudulent servers. That way, users will still see the legitimate URL in their email client or web browser, but the server that they get connected to will be a fraudulent one. This can be done by hacking into servers or home routers (Tsow, 2007) and changing the default DNS server to a compromised one. The term

[8] A group of compromised computers connected to the Internet that can be made to do certain automated tasks without the knowledge of their owners.

[9] Domain Name Service – an Internet service used for mapping host names (such as '`www.google.com`') to IP addresses.

pharming also includes hosts file poisoning, which involves adding fraudulent entries to the local DNS 'hosts' file on a computer.

There are typically three different areas of spoofing that phishers are able to employ (Gupta, 2007b): IP spoofing, email spoofing and web spoofing. The first involves modifying the low-level packets that are sent over IP networks by changing the sender address. This can be used in Denial of Service (DoS) attacks against commercial websites or in attempts to bypass authentication in a corporate network.

Email spoofing involves sending out emails that appear to be from one sender but are in reality sent from another. It is a technique used in spam emails and is possible due to a lack of security in the SMTP protocol (James, 2005). The IP address from which the email originated cannot be faked however, which is why spammers and phishers never send emails from servers they own. However, only the '`From`' address is ever visible when reading an email which is why so many people fall for this type of spoofing.

Web spoofing involves setting up a fraudulent web site that appears to the user to be the legitimate web site. As with email spoofing, there are various techniques currently being used, as described earlier. An additional technique is possible with the injection of malware into a system in terms of the 'man in the middle' proxy. This sets up a hidden intermediary to pass network traffic between a fraudulent service and the legitimate service, thus ensuring that the legitimate service works as expected and the user is never aware of what is going on (Emigh, 2007). By forwarding all data to the legitimate web site the user remains unaware that there is anything wrong since the legitimate web site behaves as expected and they could go on and give out further personal information.

The increased number of phishing attacks in the last few years has led to it being taken more seriously, with various attempts at countermeasures being put in place. Studies in the US (Graeber, 2004) and the UK (Ensor, 2005) highlight a growing concern for online security amongst Internet users, yet a large number of people have still never heard of the term phishing. While many Internet users have installed antivirus software and firewalls, lack of interest (and desire) in learning more about threats associated with identity fraud has led to customer complacency. Teenagers have been found to be less victimised by phishing attacks (15%) than have adults (22%), but they are willing to change their behaviour to protect against online fraud (Lopez, 2006), suggesting that education and awareness are as important as the security measures themselves. Banks may also need to make it clearer to customers where the liability resides for losses incurred due to phishing attacks.

Research that investigated the strategies used by phishing web sites (Dhamija et al., 2006) showed that one in four of the respondents did not look at security indicators (such as padlock icons) in the web browser, features designed to aid users in identifying suspect web sites. In addition, even experienced computer users were fooled by some of the fraudulent web sites that they were exposed to in the research. A study using gaze tracking (Whalen and Inkpen, 2005) showed that whilst the padlock icon in a web browser is commonly viewed, the site certificate information is rarely checked and most users actually stop thinking about security after having successfully logged in to a web site. Even browser toolbars specifically developed to convey security information to the user are actually ineffective at preventing phishing attacks (Wu et al., 2006) and many Internet users are unable to tell from web browser cues when an Internet connection is secure or not (Friedman et al., 2002).

Human-centred strategies to combat phishing attacks have included educating users about computer security and phishing (Bardzell et al., 2007; Robila and Ragucci, 2006); using training systems and games to teach users how to avoid phishing attacks (Kumaraguru et al., 2009; Sheng et al., 2007); training users in how to use online applications in a secure manner; encouraging good security practices in the workplace; and fostering cultures that encourage compliance. Explicitly warning users about risks during critical points in a security procedure can also help to heighten awareness, but unless users are stimulated by such warnings they are likely to have only a limited effect (Karlof, 2009). However, badly designed web sites can undermine the effectiveness of education in computer security and phishing. One study (Falk et al., 2008) analysed 214 web sites of financial institutions in the U.S. and found that 76% of these web sites had at least one security design flaw. Such flaws included presenting secure login options on insecure pages and offering to send users their statements and passwords through email.

Other, more design-centred strategies to improve security include having different security procedures for system administrators, developers and users in software; hiding security features in applications from the user if the application can automatically manage them; performing usability testing and experiments to learn more about what security policies an enterprise requires or expects, and what security policies a user is capable of properly complying with.

There are three main types of technology solutions that are available for enterprises like banks to use to combat phishing (Penn, 2004). Firstly, alerting services are available that continually search the Internet for fraudulent / spoof sites and will notify companies if their (legitimate) site has become the target of a phishing attack. This can be achieved by setting up 'honeypots' on the Internet (Viecco, 2007), systems that are deliberately left open to attack but are closely monitored so as to determine the behaviour and intent of any intruders on the system. Secondly, email validation procedures can check the validity of the content of an email through message signing, or can check the identity of the server that sent the email via domain validation (Myers, 2007b). Thirdly, web site validation can be achieved either through the web site itself (via two-step login procedures) or through the browser (using a phishing filter that compares the true IP address with those of known fraudulent sites). If a legitimate web site is discovered to also be being used as a phishing site, the owners of the web site will have to be convinced that their site has been compromised and should be taken down, and there may also be legal issues too. This 'takedown' can sometimes take too long, and some Internet service providers are reported to have launched denial of service (DoS) attacks on the phishing sites to prevent further users reaching them.

Users can also take an active role in the detection of spoofed emails by setting up spam filters in their email client to check the contents of each email for known keywords or content. Many email clients and spam filters also come with IP address blacklisting and whitelisting for additional control and whilst users are generally willing to accept a small amount of spam getting through to their email in-box, they are likely to be less forgiving about a small amount of valid emails being blocked as wrongly marked as spam. Because of the proliferation of spoofed emails an increasing number of online companies and banks have made a business decision to avoid use of email as a communication channel with customers. They instead offer local messaging systems on their web sites through which customers can be contacted. Google has recently added an authentication feature to their Gmail email client which displays a ‘key’ icon next to authenticated email addresses, but this icon will currently only be displayed for emails sent by eBay and PayPal and it is unclear how banks and other companies could register for this in the future.

When the World Wide Web first started, all communications between a web browser and the web site were unencrypted. With the advent of Secure Sockets Layer (SSL) and its successor Transport Layer Security (TLS) communications between web client and server had the option to be encrypted, in theory preventing the possibility of man-in-the-middle attacks. Both SSL and TLS involve the exchange of keys (in common with public key encryption and digital signatures) to set up a secure communications channel (Myers, 2007b), and avoid the issue of reliably distributing keys by requiring Certificate Authorities to issue them. However, man-in-the-middle attacks can still be achieved, either by installing a new fraudulent Certificate Authority via malware or by polluting the DNS cache which maps the correct URL to the wrong IP address.

There have been several attempts to define a more secure solution than SSL/TLS. Secure Electronic Transaction (SET) was a technology developed in 1996 that had the backing of many large companies, including MasterCard and Visa (Garfinkel and Spafford, 2002). This failed to gain popularity due to its level of complication, its reliance on requiring all parties to have digital signatures and because the cost per transaction was too high for small payments. Another alternative, Disposable Credit Card Numbers (DCCNs) are similar to one-time passcodes and remove the need for customers to divulge their credit card number to online merchants during an online transaction (Kadirire, 2010). A temporary, one-use credit card number is generated at checkout when the customer enters a password that was set up during registration. The merchant will be unaware that the credit card number is only temporary and the issuing bank can then associate that credit card number with the originating customer and that single transaction.

If a web site is displayed over a secure connection, the web browser should communicate the results of setting up the secure connection to the user. This is displayed in most browsers by having a clickable padlock button, and by having a URL that begins with '`https`'. Not many users are aware that they can click on the padlock icon for more information about the security certificate used to set up the connection.

Not all certificates are issued by Certificate Authorities. Many are self-signed which the browser will warn about, but the user will still have to accept if they want to use the legitimate site (Myers, 2007b). This has the effect of deconditioning the user to such warnings and accepting any certificates without much thought. Users are expected to be trained in knowing what a Certificate Authority is and this is clearly not true for most.

Attempts to use paper-based approaches involving lists of one-time passcodes for two-factor authorisation have been shown to be susceptible to phishing (Bennett, 2004) and a move to alternative passcode systems, such as using picture passwords (Fraser, 2007) or passfaces (Brostoff and Sasse, 2000), might prove more successful against password interception tools such as keyloggers. Other approaches aim to target web page spoofing that is used in most phishing attacks by making use of memorable images and employing visual matching (Dhamija et al., 2006) or comparing the content of a suspect URL with pre-registered content (Liu et al., 2006). Some two-factor authorisation mechanisms that use security fobs or other electronic devices offer some degree of enhanced protection from phishing (Ollmann, 2004).

The nature of the threat from phishing attacks has evolved as countermeasures struggle to keep up and consumers become (slowly) more aware of the threat and the issues surrounding it (Lopez, 2005). Phishing attacks are now targeting smaller financial institutions and have started to target individual account holders at specific banks using 'spear phishing' (Jakobsson, 2007). By better targeting victims, phishers can afford to make more effort in making the lure more realistic, can expect a higher success rate, and are less likely to get caught in a honeypot. However, technology is also moving to keep up, with spam filters and phishing filters reducing exposure to fraudulent sites, and with new standards emerging for authenticating email.

With more and more Internet users being online every day, it is relatively easy for fraudsters to obtain information on targeted individuals using publicly available data. Security challenges based on, for example, mother's maiden name are made ineffective as such knowledge becomes available on publicly accessible databases (Griffith, 2007). Other phishing / pharming attacks use social networking (Jagatic and Johnson, 2007), use of web browser history (Stamm and Jagatic, 2007), use of the autofill feature in forms (Menczer, 2007), retrieving data from deleted files on disks (Garfinkel, 2005), and even acoustic keyboard emanations (Zhuang et al., 2007) that can betray what is being typed by a potential victim.

2.3.2. Malware

Malware is also being used by fraudsters to enhance the effectiveness of phishing attacks. By using key logging software or trojans, phishers can now obtain anything that is typed in or displayed on the screen, not just in a web browser.

Malware is generally spread using either social engineering techniques (convincing someone to open an attachment in an email, for example) or through security vulnerabilities via use of a worm[10] or virus or a security loophole in a browser. Some examples of malware used in phishing include (Emigh, 2007) keyloggers (or keystroke sniffers) which are either installed as browser helper objects (BHOs) which record keystrokes when the user is at the target (legitimate) web site, or as device drivers which monitor all keyboard and mouse events; session hackers which wait until the user has logged into their account and then use the active session to perform background activities; web trojans which pop up over login screens to look exactly like the legitimate web site but collect credentials for the malware instead.

There are also lower-level system-wide malware techniques such as adding a proxy server through which all data gets rerouted; modification of DNS servers or hosts file

[10] A self-replicating form of malware designed to propagate across, and typically adversely affect, computer networks.

poisoning for use in pharming; scanning a user’s files on the computer looking for specific types of confidential information (data theft).

To combat malware on the customer’s computer, anti-viral, anti-malware (disinfection) solutions perform time-consuming (typically some 10 minutes) scans of a computer’s hard disk; and / or search for currently running broad coverage threats in the computer’s memory. These solutions are inherently time consuming since they search for all known viral threats, resulting in a huge database of malware signatures. They also require on-going updates to remain effective against recently discovered threats. One suggestion for reducing the overhead of such security solutions is to have sessions with higher security requirements running in a virtual machine which could then be closely monitored by the host computer (Sinclair and Smith, 2005), although the initial setup and maintenance of this solution would currently be beyond the skills of most users.

A relatively new approach for security software involves server-side detection, scanning for viruses which might be directed at specific enterprise sites rather than all-encompassing disinfection solutions. These server-side detection approaches are leaner in software terms and faster to download and run at log on. They allow the bank / enterprise to remotely initiate a scan of the customer’s PCs for malware when the customer is logging on to the (bank’s) Internet banking site, without demanding installation of a complete anti-malware solution on each customer’s computer. A similar proposal involves client-side logging with server-side auditing in order to detect the presence of malware on a customer’s machine (Jakobsson and Juels, 2009).

Having discussed the background research in the field of usable security, the rest of this chapter introduces the details of the methodology used in this research.

2.4. Usability Engineering of Security Procedures

Like other traditional engineering disciplines, usability engineering for security procedures involves procedures being designed, built and tested to achieve a high level of usability by focusing the process on the user and their tasks; providing techniques to support the management of resources in process design and development (Whiteside et al., 1998) and using principles of quality measurement (Faulkner, 2000), and relies on being able to define and measure usability in a quantifiable way.

2.4.1. Usability

There are many definitions of usability, but the key elements that they all share are summarised in this statement:

> *"The concept of usability means making artefacts easy, efficient and comfortable to use."* (Stanton and Young, 1999)

The typical benefits of having a highly usable system are increased productivity, decreased training costs and increased customer satisfaction. While some of these might not apply or be important for a specific system, usability is now beginning to be recognised as being as important as the performance of a system (efficiency of usage) and its robustness (effectiveness of coping with errors) in terms of system quality attributes.

The International Organisation for Standardization (ISO) Committee for the Ergonomics of Human-system Interaction gives usability a more formal definition as being:

> *"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."* (International Organisation for Standardization, 1998)

Here, effectiveness governs whether a typical user would be able to carry out their intended task and whether the system supports a range of user skills and needs. Efficiency describes the need for users to complete tasks quickly and with minimal effort, and satisfaction is measured by the attitudes of a person using the system and if it is comfortable for them to use. Successfully measuring all of these is key to measuring the overall usability of a system.

Such definitions of usability, admittedly generalised, have been criticised for placing too much emphasis on effectiveness and efficiency whereas some systems might not require these in favour of more specific usability goals (Quesenbery, 2003).

Taking a more detailed view, usability has been traditionally associated with five key dimensions, or attributes (Nielsen, 1993): ‘learnability’, ‘efficiency’, ‘memorability’, ‘low error rate’ and ‘satisfaction’. These five usability attributes are generalisations which should apply to most, but not all, systems. There are many dimensions to usability and variants of the five attributes listed above typically form the basis for a template upon which further attributes can be defined, either by broadening the scope of usability to other areas, or by splitting the main attributes into more specific detail. The key to a usable system is not only identifying these attributes but also their relative importance to one another. A well-defined set of usability attributes allows for their systematic measurement which can then be statistically analysed.

Additional usability concepts include ‘utility’ (Hartson, 1998) since it is just as important for a system to be useful as it is for it to be easy to use, how much knowledge people will be able to take from using a similar system to work out how to use a new or modified one, and ease of recovery when errors occur. The latter is of vital importance to the banking industry since usability engineering methods can be used to ensure that unrecoverable or unnoticed errors do not occur and minimise calls to the Helpdesk or via other, costly channels.

In addition to these, consideration must also be given to looking at the needs of the users when using the system: how much training is required, where will users be operating the system, when will they be operating the system, and what background knowledge users are expected to have. Once a working prototype is available further research can be carried out into what the most common errors users make are, how they recover from such errors and if they learn from their mistakes.

The importance of the usability of systems was recognised as far back as 1979 after the Three Mile Island nuclear accident. Inexperienced operators during the accident mistook a light on a control panel to represent the operation of a valve in the cooling system. However, the light only indicated if the solenoid in the valve was working correctly and this misunderstanding led to the operators wrongly assuming the valve was open. Such a mistake could have been prevented with proper training (the better trained operators in the next shift correctly diagnosed the real problem) but if the control panel had been easier to understand to someone with little or no training then perhaps the accident could have been avoided.

Research has shown (Karat et al., 2005) that enterprises that actively employ usability engineering techniques to design their products can not only improve the user experience of the product but can also bring about a significant return on investment.

There are two main ways to evaluate usability: expert-based inspection (Nielsen and Molich, 1990; Polson et al., 1992) and user experiments. Expert-based inspection can offer insights from usability experts and be used as part of the evaluation of a design before a working prototype is available. User experiments have the advantage of directly observing real users and can offer more insights, especially where the initial assumptions about users may prove unrealistic (Karat, 1998; Lewis, 2001), but have the disadvantage of usually being more costly than expert-based inspection.

With user experiments, a large proportion of any usability engineering work involves observing user behaviour and how the relevant usability attributes affect the user in performing the required task. Usability engineering also involves research in human performance and capabilities, anthropometrics[11] and design (Norman, 1988).

Data collected during user experiments can be quantitative, such as the time taken to complete a task, task completion rates or error rates. Qualitative data can also be collected, such as having the user observed by a usability expert, having them think aloud during the session, or by taking part in a structured interview session or focus group. Think aloud sessions tend to cause certain groups of users to behave differently in that they concentrate more on the process (rather than the task) than they would normally, so non-intrusive observations are generally preferred.

Administering questionnaires (Hornbaek, 2006) can be a good alternative to structured interviews and can provide quantitative data on attitudes and feelings that can be analysed using statistical techniques.

[11] Anthropometrics – measurement of the human individual.

The research reported here is based on such user experiment approaches permitting statistical analysis of (quantitative) attitudinal data on usability attributes derived from questionnaires.

2.4.2. Gaze Tracking Cameras in Usability Research

The research presented here employs gaze tracking cameras as one of the key tools used to capture usability and usage data during usability experiments.

It is assumed with gaze tracking that visual attention gives a direct insight into attention in general, i.e. what a user is focused on and currently thinking about. There has been a lot of research in this area (Duchowski, 2007) and it is generally accepted that this is the case, although users can relatively easily covertly direct their gaze away from their focus of attention, or even daydream.

Foveal vision describes *what* is being looked at (the focus of attention) and is the scene being projected onto the fovea, the high-resolution part of the retina. Parafoveal (or peripheral) vision highlights *where* is being looked at and indicates the periphery of the scene being projected onto the low-resolution movement-sensitive part of the retina; this is where the next focus of attention is likely to come from.

The retina is not like a digital camera with a grid array of pixels. It has various physiological limitations that have to be corrected by the visual processing centres in the brain, which often account for various optical illusions, resulting in unexpected visual perception. The chemical processes used to convert photons to electrical impulses in the photoreceptors also mean that they are motion sensitive and have to have a continually moving scene (however small) in order to be stimulated. Even if the eye is fixated on a particular object, ‘microsaccades’ occur in order to provide enough tiny movements to keep an image being generated. If these microsaccades did not occur and the image on the retina was motion-compensated then the viewed scene would become blank within about a second if no other eye movement occurred.

This behaviour is related to ‘persistence of vision’, a phenomenon where the eye is unable to sample rapidly changing intensities above approximately 50Hz. This is how television and cinema are able to achieve the illusion of a lifelike scene as opposed to the constant flickering frames that are really being shown. They also make use of the ‘phi phenomenon’, or stroboscopic motion, where light flashes give the illusion of motion above 60Hz.

Motion detection is the main task for peripheral vision processes and provides an early warning system for moving targets about to enter the visual field. As such, peripheral vision is most likely to be processed in parallel fashion within the brain in order to achieve a high response rate, whereas foveal vision is more likely to be processed in serial fashion. However, peripheral vision perceives the velocity of a moving target to be slower than would the fovea, and is twice as sensitive to horizontal axis movement than vertical axis movement, something that should be taken into consideration when gaze tracking with moving targets.

Fixations are eye movements lasting between 150 to 600ms that stabilise the fovea over a stationary object of interest. During this time, the oculomotor system is generating microsaccades (the tiny movements that ensure that the image does not disappear on the retina) and is getting ready to reposition the fovea to another fixation through a rapid eye movement called a saccade. Saccades may be voluntary or involuntary, last from 10 to 100ms and appear to be preprogrammed in that once they are executed, they cannot be altered. Programming saccades is performed by visual processes in the brain and can take up to 200ms to prepare. During a saccade, a person is effectively blind due to the rapid eye movement; however, 90% of viewing time in humans is devoted to fixations.

Saccades are used to move from fixation to fixation and, when voluntary, are an indication of the desire to change the focus of attention, but there are other movements generated by the oculomotor system as well. Smooth pursuits are eye movements that follow a moving target and match its velocity and direction. A feedback mechanism is used that ensures the target is always centred, otherwise a saccade occurs that attempts to correct the motion and continue the smooth pursuit. Nystagmus movements compensate for the movement of the eyeball or head during fixations or smooth pursuits and vergence movements focus the eyes on a target in the distance for depth perception. However, only fixations, saccades and smooth pursuits need be modelled or detected during eye movement analysis and all three can be characterised by linear filters.

The gaze tracking cameras (Tobii T/X[12]) used in this research are video-based, utilising both pupil and infrared corneal reflection to provide point of regard binocular tracking with free-head movement, designed to be non-intrusive and easy to use, ideal for use in usability experiments.

The cameras use infrared-transmitting LEDs to generate reflection patterns on the corneas and lenses of the user's eyes (Purkinje images). A built-in video camera collects these infrared reflection patterns and also measures the position and distance of the eyes relative to the camera, thus allowing for free head movement and non-intrusive measurement, ideal for a usability experiment. Real-time image processing is then used to calculate the gaze point of each eye on the screen. Because the eye tracker operates in the infrared spectrum, it can be sensitive to near-infrared interference such as direct sunlight, which must be controlled during an experiment.

Because eyeballs and corneas can vary in shape between individuals it is essential that each user calibrates the camera settings to their eyes before use. This involves asking the user to gaze at a set of points (usually between 5 and 9 points are used, spaced across the area of the screen), which appear sequentially on the screen for about 2 to 3 seconds each. This enables the gaze tracker camera to correct for any aberrations in subsequent gaze tracking with that individual.

The Tobii gaze tracker provides a stream of gaze tracking data packets (typically every 10 milliseconds). Data contained within each packet include: time stamp (in microseconds); gaze target (for current calibration) on screen for each eye ranging between (0, 0) and (1, 1); validity (or confidence) score that the gaze target is correct for each eye; pupil size for each eye, in millimetres. Using the validity scores for each eye it is possible to determine whether the average gaze target of both eyes, the gaze target for just one eye or no gaze target can be used for each eye tracking event. Knowing the resolution of the screen, the gaze target can be converted to screen pixels. Such logs can then be processed at the end of the experiment, filtering the

[12] http://www.tobii.com

gaze events through a saccade-fixation filter and converting screen co-ordinates to points of interest.

The client application used in the research reported here communicates with the gaze tracking server via a TCP/IP connection, which allows such applications to run on different operating systems. This is all achieved by building applications with a special set of libraries that are provided in the Tobii Software Development Kit (SDK). These software libraries expose a set of functions that can be used to perform various gaze tracking tasks, and perform all underlying TCP/IP communication with the server in the background.

The development libraries supplied with the SDK include a low-level API for performing basic gaze tracking functions, and also a component API built on top of that which provides a COM interface with some ready made GUI components.

2.4.3. Usability Research Methodology

In scientific and engineering research, controlled experiment-based evaluation of user interfaces can provide robust data suited to hypothesis testing. This form of research involving the direct manipulation of variables results in more pertinent data than purely correlational (observational) research techniques which involve observing natural events and do not allow manipulation of variables (Field, 2009). Critics of this approach argue that it is too artificial and that any findings may not generalise beyond the sample population. However, the biggest benefit of a laboratory setting is that more confounding variables can be controlled, allowing any effects on dependent variables to be identified and measured.

Experiment-based research methods involve the direct manipulation of an independent variable (such as the version of a prototype) and measure the effect on one or more dependent variables (such as usability attribute scores or error frequency whilst performing a task). Such experiments can be used to test theories, and the results of such experiments can be used to support or reject hypotheses (predictions about a theory). Null hypotheses assume independent variables have no effect on dependent variables, whereas alternative hypotheses oppose this view and state that there will be an effect. Statistical tests are used to determine whether such effects are of statistical significance.

In the usability experiments carried out in this research, participants were given tasks and asked to interact with a functional prototype system. After each experiment session, questions were posed about a wide range of relevant usability attributes, with regard to cognitive, fluency, quality and engagement characteristics. These measures allow the resulting analysis to isolate areas where improvements are needed before the development of a final design. The evaluation process characterises user attitudes and behaviours (Howell, 1985).

Each of the experiments reported here were repeated-measures (Landauer, 1998) (within subjects) experiments, in which all participants experienced all versions of the designs being tested, the benefit being to allow comparisons to be made for each participant, reducing the effect of individual differences. When participants experience more than one version of a design there can be an order effect, either through habituation through learning, or fatigue causing bias. As such, the order of experience was carefully balanced across the sample in controlled experiments (Preece et al., 2002; Robson, 1983).

Experiment procedures were standardised: each participant received minimal instruction (priming) and followed the same session blueprint. In this way, the data collected can be used for statistical comparisons (Coolican, 1990; Whiteside et al., 1998).

The participants in usability experiments were recruited so as to be representative of the intended users of a system. Age and gender, background and experience were all taken into account during the recruitment process and the spread of version and order of design experienced were carefully balanced. Participants were invited to attend the experiment session for a set period and were given an honorarium payment in return for their participation.

A variety of data were collected in the usability experiments presented in this research, pertinent to the requirements for each experiment. Quantitative data on participants' attitudes to usability attributes were captured by computer-based, self-completion questionnaires. Usability questions were randomised to avoid contextual effects (Oppenheim, 1992) and were carefully worded to avoid bias or careless prompting. Quantitative data were also obtained from the tasks as they were being performed, such as timing data or success rates. Qualitative data were also obtained from individual interviews at the end the hands-on session for each participant.

Immediately after experiencing each design variant, each participant was asked to complete a usability questionnaire, designed to objectively measure their attitudes across a range of usability attributes. The use of such questionnaires to evaluate services and user interfaces has a long history (LaLomia and Sidowski, 1990; Root and Draper, 1983) and the questionnaires used here have been developed and refined over many years (Dutton et al., 1993; Jack et al., 1993; Love et al., 1992). They have been adapted and used in a wide range of experiments (Davidson et al., 2004; Foster et al., 1998; Gunson et al., 2011; Larsen, 1999, 2003; Morton et al., 2004; Sturm and Boves, 2005; Weir et al., 2009; Weir et al., 2010), giving weight to the reliability of the questionnaires and also to their validity (they measure what they purport to measure).

The questionnaires are comprised of a set of statements representing an attitude, each with a set of tick-boxes on a seven-point Likert format scale (Likert, 1932; Rossi et al., 1983) ranging from ‘strongly agree’ through ‘neutral’ to ‘strongly disagree’. The statements cover a range of usability-related attributes such as ease of use and complexity, level of frustration or stress experienced, efficiency of the design, and other issues specific to an experiment. Participants find such questionnaires natural to respond to, and they have been shown to have high reliability (Oppenheim, 1992). However, they are best used for relative comparisons between designs rather than as a tool to provide absolute measures.

To counteract the problem of acquiescence response bias, which is the general tendency for respondents to agree with the statement offered (Gross, 2001), statements in the questionnaire are equally balanced, positive and negative, and are presented in a randomised order. During analysis, participant responses are converted into numerical values from 1 (most unfavourable) to 7 (most favourable) allowing for the polarity of the statements: for example, a ‘strongly agree’ response to a negative statement is converted to a value of 1. A resulting score higher than 4 therefore represents a positive attitude and scores below 4 represent a negative attitude, with 4 being the neutral point. Taking the mean of these normalised numbers across all of the questions gives a measure of each participant's overall attitude to the usability of a design and a measure of the overall attitude to the design can then be obtained by taking the mean of all the questionnaire results for participants who experienced that design, giving the overall usability metric (Coolican, 1990), an objective measure of system appeal (Nielsen, 1993).

The mean scores for individual usability statements in the questionnaire can also be examined to highlight any aspects of the design which were particularly successful or which require improvement. Results can also be analysed according to demographic groupings of participants (such as gender or age) and any significant differences between groups can then be identified.

In each of the usability experiments reported here, all participants completed the questionnaires following exposure to each of the experiment conditions. An overall attitude score was determined by calculating the overall mean for all of the attributes by all participants. Individual attributes were also analysed separately to identify any specific issues that arose. These data were analysed using parametric tests involving analysis of variance (ANOVA).

Although it was originally considered (Likert, 1932) that data from such Likert scales are ordinal data as opposed to being interval data, and so should be analysed by non-parametric techniques, there is considerable evidence (Kim, 1975; O'Brien, 1979) to suggest that parametric tests are in practice sufficiently robust against violations of their underlying data assumptions, providing these are not too extreme, and can therefore be used with Likert scale data.

Statistical methods are used to analyse the quantitative results collected during the experiments reported here. The output from such analyses can be used to draw conclusions about whether the hypotheses used in the experiment are likely or unlikely. There are two ways to test a hypothesis – observe what naturally happens (correlational research) or manipulate one variable to see its effects on another (experiment-based research). Using correlational research does not guarantee that cause comes before effect and the impact of external influences (confounding variables) can never be fully ruled out (Field, 2009). The research reported here uses only experiment-based techniques, where an experiment collects data from a cohort of the population (a sample) and uses these data to infer things about the population as a whole (Field, 2009).

The choice of statistical tests used to analyse the data depend on the type of the dependent variables that are measured during the experiment. Categorical data can be binomial (2 categories), nominal (more than 2 categories) or ordinal (ordered categories) whereas interval data are measured on a continuous scale between a maximum and minimum value. The usability questionnaires used in this research use a seven-point Likert scale (Likert, 1932), which measures participant attitude with an ordinal variable. There has been some debate about whether data from such a questionnaire can also be considered to be on an interval scale (Martilla and Garvey, 1975; Munshi, 1990). While there is no resolution on this issue, it is generally accepted that if the differences between the responses are perceived to be equal then the advantage of the more powerful statistical tests with known sampling errors offsets the small amount of error that may be accompanied with the treatment of ordinal variables as interval data. (Labovitz, 1970).

Background on the statistical tests used in this research, and on working definitions of levels of statistical significance for the work, can be found in Appendix A.

The next chapter exploits the usability research methodology introduced here in a series of three usability experiments with different cohorts of users, to report on usability performance of the CodeSure card as a proposed universal usable security procedure to combat phishing and malware attacks.

Chapter 3. CodeSure Card Usability

Results from three experiments to provide a detailed characterisation of the usability of the CodeSure card are reported in this chapter.

The first experiment (The CodeSure Card Baseline Usability Experiment) involves 112 participants in a longitudinal study of usability metrics for the (baseline design) CodeSure card used in two-factor customer authentication.

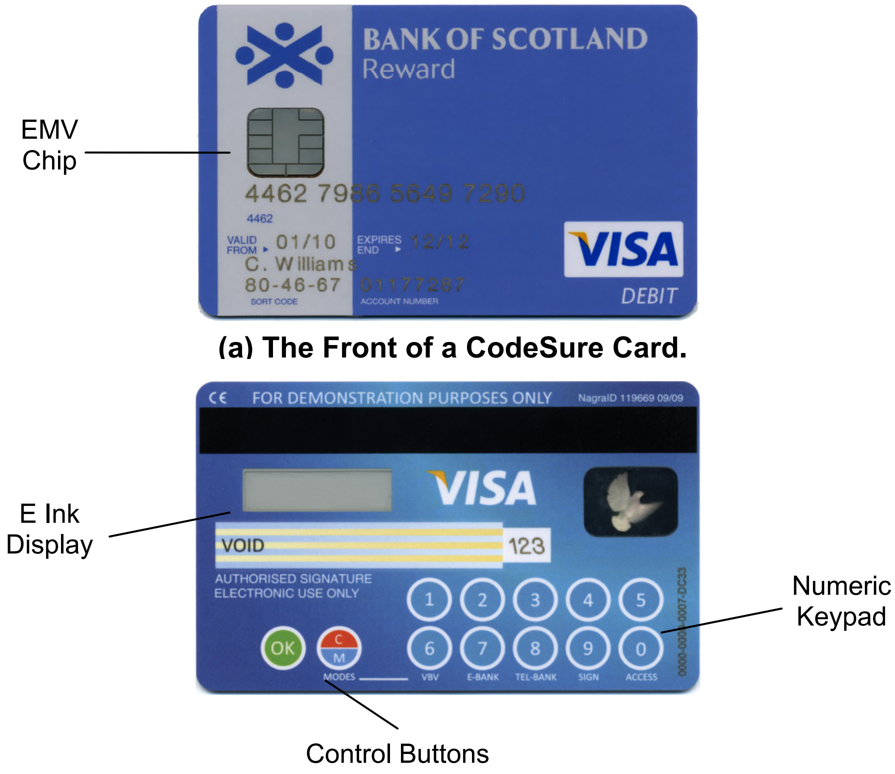
Identifying areas for improvements from that experiment, the second experiment (The Enhanced CodeSure Card Usability Experiment) with 70 participants used a simulation of the CodeSure card running on a handheld touch screen tablet device to investigate the usability and performance of a proposed enhanced user interface design. The tasks for the experiment included using the CodeSure card to generate one-time passcodes and to sign financial transactions.

The third experiment (The CodeSure Card Keypad and Display Synchrony Experiment) involved 40 participants to investigate issues with the display and the buttons of the CodeSure card by using a different simulation of the CodeSure card, this time running on a desktop computer with touch screen, operating in conjunction with gaze tracker cameras to investigate gaze behaviour during use of the CodeSure card.

Taken together, the findings from these three experiments offer a detailed characterisation of the usability attributes of the CodeSure card as the foundation of a solution for usable security procedures.

3.1. The CodeSure Card Baseline Usability Experiment

This experiment collected usability and performance data from participants using the CodeSure card (Figure 3.1) in the context of a two-factor customer authentication procedure with an Internet banking site. Participants were all using the CodeSure card for the first time, and for five uses in a session: and also in a repeated session after seven days had elapsed. The aim of the experiment was to produce data on user attitudes to usability and on user learning behaviour.



(a) The Front of a CodeSure Card.

(b) The Obverse of a CodeSure Card.

Figure 3.1: The CodeSure Card

The CodeSure card is constructed in the form of a standard ISO/IEC 7813 bank card with a two-track magnetic stripe and a standard EMV[1] chip. For security, the EMV chip has no connections to the CodeSure card components. The CodeSure card contains a 6502 8-bit embedded microprocessor running at 5MHz, surface mounted on a flexible printed circuit board, a 12-button keypad (each button uses copper electrodes separated by an air gap) and an 8-digit alphanumeric E Ink (Electrophoretic Ink) display, which has the advantages of low power consumption (power is only required to change the display rather than refresh it), and a wide viewing angle. Power for the CodeSure microprocessor and the E Ink display is supplied by a thin-film Li-ion battery that has been proven to be resistant to the types of flexing experienced by the card with general usage. The battery has a projected operating life of around three years which is the typical replacement frequency for card issuers in issuing replacement cards to customers.

The CodeSure card can operate in several authentication modes depending on the application required. In Identify mode, the primary mode for card use in two-factor authentication procedures, entering a 4-digit card PIN generates an 8-digit OTP for use in two-factor authentication. In Exchange mode, a challenge-response exchange between bank and customer provides enhanced validation (for transactions), entering a 4-digit PIN and an 8-digit reference challenge number (which may be a coded form of the account numbers involved in a transaction) generates an 8-digit OTP for the challenge-response exchange. In Verify mode, customers can reassure themselves that a phone call, text message or email purporting to be from the bank is legitimate by checking that the bank provide a (reverse authentication) reference number that is valid only for that customer's CodeSure card. Entering this (8-digit) reverse authentication reference number generates a PASS/FAIL display for that reference number. In Sign mode, transaction signing is possible by entering the target account number for a transaction and then the amount to be transferred, followed by the 4-digit PIN to generate an 8-digit OTP. The numeric keypad serves a dual role, allowing mode selection involving pressing a control button followed by a mode (digit) button, as well as digit data entry.

[1] Europay, MasterCard and Visa, whose collaboration resulted in a global standard for Chip and PIN.

3.1.1. Experiment Design

A sample of 112 banking customers was recruited in Edinburgh to take part in the experiment sessions. Participants attended for their individual experiment session and were introduced to the CodeSure card and given a (paper) User Guide for the device (see Figure B.1 in Appendix B). They then used the CodeSure card to generate a one-time passcode in order to login to a typical Internet banking page – a total of five uses in succession during their 60-minute experiment session.

After each use of the card, participants completed a usability questionnaire. The questionnaire was designed to explore attitudes to device usability and security and was presented as 25 randomised attitudinal statements allowing users to respond on a 7-point Likert format that ranged from 'Strongly Agree' to 'Strongly Disagree'. The usability questionnaire statements are listed in Table 3.1, which also shows the statement polarities (needed in the data analysis) and abbreviated forms used in the discussion here.

The 18 core usability statements used (Love et al., 1992) were:

Cognitive attributes – concentration, feeling flustered, stress, frustration.

Fluency attributes – complication, knew what to do next, feeling in control, quick.

Quality attributes – use again, reliable, needs improvement, matched expectations.

User engagement attributes – user-friendly, appearance, liked using, enjoyed using, trustworthy, secure.

To these were added 7 device-specific usability statements (Table 3.1):

Device attributes – button layout, ease of use of buttons, screen response, display size, ease of reading, ease of retrieving information, need for user instructions.

Questionnaire Statement (polarity)	Abbreviated Form
I had to concentrate hard to use this device (-)	Concentration
I felt flustered when using this device (-)	Flustered
I felt under stress when using this device (-)	Stressed
Using this device was very frustrating (-)	Frustration
Using this device was too complicated (-)	Complication
When using this device I didn’t always know what to do next (-)	Knew what to do next
I felt in control when using this device (+)	In control
Using this device was quick (+)	Quick
I would be happy to use this device again (+)	Use again
I felt this device was reliable (+)	Reliable
I felt this device needs a lot of improvement (-)	Needs improvement
The operation of this device didn’t match my expectations (-)	Matched expectations
I found this device user-friendly (+)	User friendly
I liked the appearance of this device (+)	Appearance
I liked using this device (+)	Liked
I did not enjoy using this device (-)	Enjoyment
Using this device felt trustworthy (+)	Trustworthy
Using this device felt secure (+)	Security
The layout of the buttons on this device was clear (+)	Button layout
I found the buttons on this device easy to use (+)	Easy to use buttons
The screen on this device was slow to respond to button presses (-)	Screen response
The display on this device was too small (-)	Display size
The display on this device was difficult to read (-)	Easy to read
Retrieving the details I needed from this device was straightforward (+)	Information retrieval
I would have liked more instructions on the use of this device (-)	Instructions

Table 3.1: Usability Questionnaire Statements

Participants were asked to access ‘their’ bank account, and in order to do so they would need to use the CodeSure card to generate, and then enter a passcode into the screen. To generate the passcode they needed to enter their (supplied) PIN number into the CodeSure card. If the PIN was entered incorrectly on the CodeSure card, this was not reported to the participant until checked by the banking site.

In common with all other Chip and PIN cards, but unlike other OTP generators, the CodeSure card has no record of the PIN actually stored on the card itself. If the PIN is entered incorrectly by the user then an OTP will still be displayed and the user will not be notified that the OTP is actually invalid. All of the checking is done via the bank's CodeSure card server, which maintains a record of the PIN for each card and is able to verify that the OTP was generated using the correct PIN for that card. The CodeSure chip on the card is unable to communicate with the EMV chip on the card.

The programming interface for the CodeSure card server was accessed using SOAP (Simple Object Access Protocol), an XML-based messaging system transferred over HTTP (Hypertext Transfer Protocol). After registering each of the CodeSure cards with the CodeSure card server it was possible to check the OTPs generated in the experiment.

A **Perl** program was written to interface with the card server using the **`SOAP::Lite`** module. Three software functions were written: **`isalive`**, which was used to check if the card server was available, and **`webservicewithserial`** and **`webservicewithid`**, both of which were able to run an individual CodeSure card server function by its specified name (along with any supplied parameters) and then wait for the response from the card server itself. Identification of a CodeSure card to the card server could either be done with a 16-digit hexadecimal serial number (using **`webservicewithserial`**) unique to each card, or could be achieved by using a 6-digit ID (using **`webservicewithid`**) tied to the account used to initially login to the card server which uniquely identified cards within the domain of the account used to login to the service.

For this experiment, a web page to simulate user login to an Internet (banking) site was written in **Perl** using the CGI module, and ran on a Microsoft IIS (Internet Information Services) web server. The web page consisted of an HTML form containing only two fields: '**`userid`**' and '**`passcode`**'. A prompt was displayed just above these fields: "*To log on, please enter your User ID and generate a passcode using your card.*" After entering the requested details and pressing the submit button, the passcode was checked with the card server.

If the card server reported an error with the entered passcode, an error message was displayed requesting the fields be completed again: "*I'm sorry, that passcode does not match your details. To log on, please enter your User ID and generate a* *passcode using your card.*" All data entered in the forms were sent to a log file for later analysis along with the number of attempts each participant took in each task.

The cohort of participants was recruited by quota, balanced by gender and age, recruited in equal numbers in three age groups:

Ages 18 to 35 'Generation Y', the 'Net generation' who have grown up in an online world, a world with small form factor consumer devices, often as early adopters.

Ages 36 to 55 'Generation X', parents of the techno-savvy Net generation, who are more usually 'followers' rather than early adopters of new technologies.

Ages 56 and older The 'Baby Boomers' who are generally (hesitant) traditionalists when it comes to new technologies, but with more wealth to be secured.

These three consumer segments were balanced in the experiment design so that robust statistical analysis could be performed on the data collected to identify differences in attitudes and reactions between various groups.

Experiment sessions took approximately 60 minutes and were conducted in a research suite at the University. The details of the experiment are summarised in Table 3.2.

In order to determine if usability attitudes change with time, participants carried out two sessions, one week apart (to the day) to repeat the experiment treatment. In the second visit, they were not given the User Guide to the CodeSure card that they had been given in the first week.

To encourage participants back for the second week, they were rewarded with a £20 cheque for taking part during the first week and a £40 cheque for participation during the second week.

Data collected included: demographic and technographic characteristics of participants (e.g. age, gender, mobile phone and text message usage); attitudes to device usability per-use; quality rating; task completion rates and timing data; debriefing interview responses with qualitative comments on specific issues.

The CodeSure Card Baseline Usability Experiment	
Experiment purpose	Exploration of user attitudes to the usability of a CodeSure card one-time passcode in two-factor authentication.
Null hypothesis (H_0)	There will be no differences in metrics of participant attitude to usability, or performance (usage times, success rates) over time.
Experiment design	Participants use the CodeSure card 5 times, entering a PIN to generate a one-time passcode for two-factor authentication each time. Session repeated after 7 days. Repeated measures, longitudinal study.
Dependent variables	Usability questionnaire (7-point Likert scale). Quality rating data (linear scale).
Other data	Demographic and technographic data. Performance data: times, error rates. Exit interview data.
Independent variables	Participant: gender (2 genders, balanced), age group (3 groups, balanced, split 18 to 35 / 36 to 55 / 56 and over).
Cohort	96 (3 age groups × 2 genders × over-sampling ratio 16:1) = 96.
Honorarium	Personal cheque for £20 for the first session and a cheque for £40 for the second session.
Session time	60 minutes.

Table 3.2: Experiment Design Details

After each session, a one-on-one interview sought to gather further opinions and suggestions about aspects of the usability of the CodeSure card, security and behavioural intention to use, after the repeated measures experiment; and usefulness and convenience of the device for authentication.

3.1.2. Results

3.1.2.1. Demographic and Technographic Results

A good balance for gender (50.5% male and 49.5% female) and for age groups (32.3% aged 18 to 35, 38.4% aged 36 to 55 and 29.3% aged 56 and over), was achieved in the final sample of 112 participants.

Participants can be judged as a valid cross-section of the customer base for the experiment, with 92.9% having visited a branch in the four weeks prior to the start of the experiment; 68.7% having phoned their bank; 62.6% having used Internet banking; and 93% having used an ATM. Their preferred method of banking was Internet banking (46.9%), visits to the branch (24.0%), ATM (13.5%) and automated telephone banking (12.5%). The mean duration of their banking relationship was 22 years.

Almost all the cohort (94%) owned a mobile phone. Of these, 99% receive text messages on their phone whilst 98% had sent a text message. The average number of texts sent in a week was 34, an important skill for an experiment involving usage of devices with small buttons and screens.

Some 61% of participants regularly shop online. The majority (68%) of the cohort claimed to do their own research before purchasing technology products rather than relying on recommendations from others – representing a good mix of 'leaders' and 'followers' for adoption of new products.

Finally, 8% of the cohort reported to being left-handed which is within the expected range (10%) for the general population.

These data confirm that the cohort was a suitable sample for the purposes of this experiment.

3.1.2.2. Usability Results

Analysis of the questionnaire responses for the 95 experiment participants who completed all 10 uses show that the CodeSure card achieves a mean usability score (grand mean over all 25 statements) of 4.49 on first use, increasing to a score of 4.66 after the tenth use, Table 3.3. Whilst the means and standard deviations are presented in tables, the discussion focuses on the (more powerful) results from the analysis of variance reported for the ANOVA tests. These results indicate an acceptable usability score, (being greater than 4.0 on this scale[2]) for the baseline CodeSure card design.

[2] The median on the scale of 1 to 7 and the minimum acceptance criteria for the work presented here.

Use	Mean Score	Std. Dev.	N	ANOVA Results
1	4.4888	1.18510	95	
2	4.6463	1.19305	95	Use 2 > Use 1 (*df*=1; *F*=9.089; *p*=0.003)
3	4.7785	1.23404	95	Use 3 > Use 2 (*df*=1; *F*=12.309; *p*=0.001)
4	4.7718	1.24574	95	
5	4.7655	1.31335	95	
6	4.4358	1.30145	95	Use 6 < Use 5 (*df*=1; *F*=11.843; *p*=0.001)
7	4.5183	1.32344	95	Use 7 > Use 6 (*df*=1; *F*=5.294; *p*=0.024)
8	4.6021	1.30171	95	Use 8 > Use 7 (*df*=1; *F*=6.001; *p*=0.016)
9	4.6581	1.29847	95	Use 9 > Use 8 (*df*=1; *F*=4.889; *p*=0.030)
10	4.6623	1.30710	95	

Table 3.3: Mean Usability Scores for the CodeSure Card after Each Use

Table 3.3 reflects a strong learning behaviour with an improvement in overall mean usability scores between successive uses. This improvement was especially marked between first and second use and between second and third use in each week. Statistical tests of usability scores from successive pairs of uses (repeated measures ANOVA with age group and gender as between-subjects independent variables) show (Table 3.3) that the increase in usability scores between first use and second use was statistically significant, *F*(1,89)=9.089, *p*=0.003: the increase between second use and third use was statistically significant, *F*(1,89)=12.309, *p*=0.001: as was the decrease between fifth use and sixth use, *F*(1,89)=11.843, *p*=0.001: the increase between sixth use and seventh use, *F*(1,89)=5.294, *p*=0.024: the increase between seventh use and eighth use, *F*(1,89)=6.001, *p*=0.016: and the increase between eighth use and ninth use, *F*(1,89)=4.889, *p*=0.030. Similar ANOVA tests confirm that the increase in usability scores between the first use and fifth use was statistically significant, *F*(1,89)=14.851, *p*<0.001, as was the increase between sixth use and tenth use, *F*(1,89)=14.847, *p*<0.001.

The difference (fall) in usability scores between fifth use (4.77) and sixth use (4.44) after a gap of seven days, was statistically significant, *F*(1,89)=11.843, *p*=0.001, reflecting a regression after the gap of seven days. There were no statistically significant effects by age group on overall mean usability scores but males tended to score the card usability higher than did females after the gap of seven days, for the sixth use, *F*(1,89)=4.736, *p*=0.032, and seventh use, *F*(1,89)=4.976, *p*=0.028.

The mean scores for the individual usability attributes by successive uses are shown in Table 3.4. Scores for ‘security’, ‘reliable’ and ‘appearance’ were high (greater than 4.90) at the outset and remained high throughout. For the first use four usability attributes failed to achieve scores higher than 4.0: ‘concentration’, ‘screen response’, ‘easy to use buttons’ and ‘needs improvement’. The usability scores for ‘concentration’ showed a significant difference (increase) between first use and second use, $F(1,89)=35.356$, $p<0.001$, and again between second use and third use, $F(1,89)=4.980$, $p=0.028$: and scores for ‘easy to use buttons’ showed a significant difference (increase) between first use and second use, $F(1,89)=5.001$, $p=0.028$ – as might be expected as users become familiar with operating the card.

Usability scores for attributes associated with ‘easy to use buttons’ on the card and the associated ‘screen response’ remain low throughout, especially in the second week and are identified as core usability issues. Scores for ‘needs improvement’ remained low throughout at less than 4.0. Interestingly, ‘button layout’ scored consistently high for the first week but much lower for the second week compared to the first week, although uses in both weeks consistently score above 4.0 for this attribute. This might be due to the fact that participants did not have access to the User Guide for their second session. In general, males consistently scored each of the usability attributes higher than did females.

	Week 1 Uses					Week 2 Uses				
Usability Statement[3]	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Concentration	3.35	**4.22**	**4.47**	4.46	4.54	**3.94**	**4.35**	**4.54**	4.57	4.75
Flustered	4.38	**4.64**	4.84	5.01	5.00	**4.55**	**4.78**	4.92	4.94	4.99
Stress	4.62	**4.85**	4.98	4.91	4.94	4.73	4.82	**5.01**	5.07	5.05
Frustration	4.34	**4.58**	4.71	4.69	4.67	4.53	4.49	4.53	4.63	4.73
Complication	4.75	4.99	5.03	5.08	5.12	4.86	4.74	**4.96**	5.03	5.03
Knew what to do next	4.55	**4.95**	5.15	5.11	5.18	**4.09**	**4.76**	4.94	5.12	5.00
In control	4.56	**4.99**	5.03	5.04	4.87	4.66	4.83	4.87	4.89	5.05
Quick	4.66	4.74	4.81	4.81	4.92	**4.52**	4.66	4.69	4.75	4.75
Use again	4.68	4.63	4.77	4.77	4.74	4.61	**4.51**	4.58	4.64	4.65
Reliable	4.93	5.01	4.99	5.05	4.96	4.93	4.95	4.97	4.99	4.96
Needs improvement	3.67	3.82	3.95	3.87	3.96	**3.61**	3.67	3.78	3.76	3.72
Matched expectations	4.32	4.42	4.61	4.59	4.55	4.53	4.39	**4.65**	4.60	4.54
User-friendly	4.37	4.36	4.54	4.51	4.46	4.21	4.28	4.29	4.38	4.37
Appearance	4.93	4.89	5.04	5.06	4.98	4.81	4.92	4.96	4.97	4.99
Liked	4.43	4.36	**4.62**	4.53	4.45	4.25	4.24	4.32	4.45	4.35
Enjoyment	4.29	4.31	**4.66**	**4.48**	4.61	**4.02**	**4.23**	4.27	4.40	4.42
Trustworthy	4.80	**5.08**	5.12	5.06	5.04	4.94	5.04	5.08	5.05	4.99
Security	5.06	5.16	5.14	**5.03**	5.05	5.03	5.01	5.02	5.13	5.04
Button layout	5.08	5.17	5.13	5.15	5.17	**4.77**	4.62	4.74	4.78	4.83
Easy to use buttons	3.56	**3.89**	3.93	3.94	3.91	**3.48**	3.62	3.61	3.60	3.63
Screen response	3.77	3.88	4.04	4.01	4.01	**3.68**	3.60	3.52	3.60	3.56
Display size	4.56	4.57	4.64	4.73	4.73	**4.41**	4.54	4.56	4.56	4.61
Easy to read	4.96	4.82	**5.06**	4.96	4.93	4.75	4.66	4.73	4.78	4.73
Information retrieval	4.89	4.98	5.19	5.31	5.20	**4.75**	4.85	4.89	5.02	5.07
Instructions	4.72	4.84	5.02	5.14	5.17	**4.24**	4.39	**4.63**	4.75	4.76

Table 3.4: Mean Usability Scores for Individual Usability Statements – by Use

Using the data from all of the ten usability questionnaires involved in the experiment, reliability analysis for each (calculation of Cronbach's α) shows a minimum of α=0.960 (this was for the first questionnaire in week 1), indicating that the reliability of all ten questionnaires was exceptionally high.

[3] Underscore bold entries represent a statistically significant difference (p<0.05) from the previous use.

3.1.2.3. Quality Rating Results

Participants were asked to rate the CodeSure card on a scale of 0 to 30, in terms of overall quality and for three attributes, ‘convenience’, ‘security’ and ‘ease of use’, Table 3.5. This was done after all tasks had been completed in each of the weeks.

Attribute	Week 1	Week 2
Overall quality	16.01	17.33
Convenience	21.54	22.43
Security	19.54	21.04
Ease of use	16.42	16.25

Table 3.5: Mean Quality Rating Results

The CodeSure card was rated highest for ‘convenience’ in both weeks but lower for ‘overall quality’ and ‘ease of use’.

Participants who gave a high overall quality rating for the CodeSure card were generally impressed by the design and portability: “*Like it, high tech, exciting*”, “*I liked the availability and mobility of the card*”. Reasons for giving a low overall rating ranged from comments on the responsiveness of the device “*The CodeSure card screen response was a bit slow*” to the difficulty of using the buttons “*The buttons were so difficult to use*”.

For ‘convenience’, positive comments for the CodeSure card focused on its portability when compared to other security devices “*The card is so easy to carry and light*”. Those that gave negative comments tended to focus on reliability: “*The CodeSure card looks like it would be easy to damage*”.

A few participants thought that the buttons on the CodeSure card would wear with use and that would aid thieves when attempting to guess the owner’s PIN: “*The CodeSure card isn’t trustworthy because it seems like it could wear and you can see where people have been pressing the keys and have left an indent corresponding to the PIN number*”.

Most negative comments about the ‘ease of use’ for the CodeSure card related to the buttons: “*The buttons on the card are difficult to press and the screen is slow to respond*”, “*The buttons on the card need improved*”.

In general these comments underscore the usability attribute trends shown in Table 3.4.

3.1.2.4. Task (Login) Times and Success Rates

Tasks were timed from the point at which the login page was displayed until the timer was stopped when the participant pressed the submit button on the login page. No attempt was made to validate the fields before submitting the form, although all entered data were logged for later analysis. The submitted User ID was not checked as the experiment's primary focus was on the generated passcodes.

All login details were checked using the card server since the card itself generated a passcode regardless of whether the PIN was entered correctly or not. The timing data reported here do not include the time spent contacting the card server (typically 3 seconds). If a participant failed validation at their first login attempt when using the CodeSure card, the time to login for their second attempt was added to the overall time for that task.

The mean login times for the login task are shown in Table 3.6 across both weeks. Statistical analysis of this data (repeated measures ANOVA with age group and gender as between-subjects independent variables) confirms (Table 3.6) that there was a significant learning behaviour up to the third use in each week.

Use	Mean Login Time (s)	Std. Dev.	ANOVA Results
1	133.88	111.914	
2	71.62	42.674	Use 2 < Use 1 (*df*=1; *F*=30.917; *p*<0.001)
3	57.31	29.254	Use 3 < Use 2 (*df*=1; *F*=11.503; *p*=0.001)
4	53.85	25.911	
5	49.22	22.299	Use 5 < Use 4 (*df*=1; *F*=8.558; *p*=0.004)
6	85.47	43.474	Use 6 > Use 5 (*df*=1; *F*=74.988; *p*<0.001)
7	63.06	46.261	Use 7 < Use 6 (*df*=1; *F*=17.661; *p*<0.001)
8	52.38	29.343	Use 8 < Use 7 (*df*=1; *F*=8.795; *p*=0.004)
9	53.37	38.571	
10	49.57	30.981	

Table 3.6: Mean Task (Login) Times with the CodeSure Card – by Use

The gap of 7 days between fifth use and sixth use produced a large increase in login times which was statistically significant, $F(1,89)=74.988$, $p<0.001$, but the sixth use still took much less time than for the first use. There were no significant differences in login times associated with gender but there was a statistically significant effect of age group on login times, with the difference in times for the youngest age group and the oldest age group being statistically significant ($p<0.05$) across all uses, and especially evident between the fifth use and sixth use, $F(2,89)=12.609$, $p<0.001$. The youngest age group participants were fastest overall with a mean login time on first use of 95 seconds whereas the oldest age group participants were slowest overall with a mean login time on first use which was twice as long at 185 seconds. The gap between the age groups narrowed after the first use in both weeks, Table 3.7.

	Age 18 to 35		Age 36 to 55		Age 56 and over	
Use	Mean	Std. Dev.	Mean	Std. Dev	Mean	Std. Dev
1	95.37	67.875	125.11	101.769	184.62	141.333
2	60.00	26.024	70.53	43.562	85.00	52.044
3	48.47	25.205	55.39	28.559	68.83	31.137
4	45.67	22.690	48.67	20.007	68.76	29.731
5	41.97	22.101	48.86	21.981	57.17	20.915
6	64.63	25.281	80.53	38.810	113.17	50.088
7	49.50	31.053	58.31	40.495	83.00	59.146
8	46.60	27.932	50.78	28.352	60.34	31.192
9	42.90	26.987	56.47	51.243	60.34	27.888
10	40.37	24.656	45.86	22.206	63.69	40.761

Table 3.7: Mean Task (Login) Times – by Age Group

The number of successful first-attempt and second-attempt logins using the CodeSure card was measured and the results are shown in Table 3.8. As with login times, there was a noticeable learning effect up to the third use of the CodeSure card in each session with a major regression in first-attempt success rate between the last use of the first week (use 5) and the first use of the second week (use 6) after a gap of 7 days.

Use	1 Attempt	2 Attempts
1	76.77%	92.93%
2	89.90%	98.99%
3	96.97%	100.00%
4	94.95%	96.97%
5	96.97%	98.99%
6	87.88%	97.98%
7	92.93%	97.98%
8	94.95%	98.99%
9	96.97%	100.00%
10	94.95%	96.97%

Table 3.8: CodeSure Card Login Overall Success – by Use

Although only 77% of participants successfully managed to login on their first attempt with their first use of the CodeSure card, Table 3.8, all subsequent uses (allowing for two attempts) had success rates better than 96%.

The login data (6-digit User ID and 8-digit passcode) entered by each participant was recorded during the experiment for subsequent analysis to permit characterisation of login errors. These data show that (data for first use, first attempts only) the reasons for entering the wrong User ID were using the PAN[4] of the card (3% of attempts) or a simple typing error (1% of attempts), although there were no User ID errors after the seventh use.

Errors with passcode entries were due to users entering the card PIN instead of the passcode (7% of attempts on the first use, first attempt), however, this improved markedly after the first use in both weeks, falling to around 1% of attempts. On first use, first attempt, some 8% of attempts were failures due to passcodes entered with an incorrect length and many of these were 7 or 9 digits in length, possibly highlighting a problem transcribing the string of 8 passcode digits from the CodeSure card display to the computer keyboard.

[4] Primary Account Number, the (typically) 16-digit number that is printed on the front of credit and debit cards.

3.1.2.5. Interview Responses

The individual interviews with participants sought to gather comments and preferences on a wide range of issues related to the CodeSure card experienced by the participants. The responses gathered are presented here with illustrative comments.

Some 45% of respondents appreciated the fact the CodeSure card could fit in their wallet and afford them the convenience of not having to carry a separate reader or similar device: “*I quite like having it on your card because I hate having to carry a card reader around*”, “*It’s one piece of equipment, its innovative having everything on the same card*”, “*I liked the concept, it would fit into my wallet*”.

Almost all participants mentioned that the responsiveness of the CodeSure card was poor: “*The buttons were a bit slow*”, “*The buttons are not very responsive. You have to push hard to work them*”, “*Sometimes I was typing and the screen didn’t seem to respond*”.

When asked for suggestions for improvements, performance of the buttons was consistently mentioned: “*The buttons could be made to respond a bit faster*”, “*The numbers should stay on screen for longer before they get masked*”, “*The screen should be quicker to respond and something should tell me when it’s activated so I know what mode I’m in*”.

Generally, participants were happy with the display on the CodeSure card although some mentioned it was too small or not bright enough: “*It was fine, it was clear enough and the font size was good*”, “*It was a wee bit small, but other than that I liked it*”, “*I’d like to see the PIN digit I’d typed on the screen rather than just a star*”.

The CodeSure card was appreciated by participants as a security solution that could be applied across a range of channels. Of the participants who use Internet banking, 82% expected to be able to use the CodeSure card to access their bank’s Internet banking service. For those who use Telephone banking, 73% expected to be able to use the CodeSure card to access their bank’s Telephone banking service. Of the participants who do online shopping, 88% expected to be able to use the CodeSure card for online shopping. Some 67% of participants said they would use the CodeSure card as their main card.

3.1.3. Discussion on Experiment Findings

The null hypothesis (H_0) for this experiment was:

> *"There will be no differences in metrics of participant attitude to usability, or performance (usage times, success rates) over time."*

The evidence presented here is sufficient to refute this null hypothesis since usability scores, success rates and usage times for the CodeSure card exhibited significant improvements between successive uses.

Usability scores for the CodeSure card at 4.49 (on a 7-point scale) for first use rising to 4.66 on tenth use confirm the device usability as suited for use in a usable security procedure. Participants recognised that the CodeSure card was convenient to use as reflected in the quality ratings and comments; and appreciated the role of the card as a multiple channel, usable security solution where 82% expected to be able to use it to access their Bank's Internet banking service, 73% to use it for Telephone banking and 88% to use it for online shopping.

The results obtained from this usability experiment highlight a range of usability issues with the CodeSure card, where improvement might further boost card usability and which are addressed in the next two experiments.

Dual mode operation. Although only one CodeSure card mode (Identify) was used in the experiment, users still had problems coping with the button sequence required for mode selection to set the device into Identify mode: "*Something should tell me when it's activated so I know what mode I'm in*". Modifying the control button sequences and better use of the display in prompting and separating mode selection from data entry might be an improvement here, although this would increase power consumption for the card.

Navigation cues. Users were often unclear about how the control buttons were meant to be used at various points in the sequence and when to enter PIN digits and other data. Better use of the display to show more navigation cues (mode entry instructions, when to press OK, which data to enter) might be an improvement here.

PIN masking. The digit masking used on the display (straight to star) also caused some problems: "*I'd like to see the PIN digit I'd typed on the screen rather than just* *a star*". In contrast to the digit masking for this baseline design of CodeSure card, which displayed each PIN entry masked as a star, a scheme where the keyed digit is displayed briefly before being masked to a star might be an improvement here.

Key-ahead. Users also had problems with the response time of the keypad, not so much with the delay between button press and screen display, but more when they tried to type ahead of the display: "*Sometimes I was typing and the screen didn't seem to respond*". A keypad buffer allowing type-ahead may be an improvement here, although it would again increase power consumption for the card.

Display format. The format of the (passcode) display as a single string of 1x8 digits may have produced read-remember-type errors for passcodes during tasks. Displaying the passcode in a 4-space-4 digit format may be an improvement here. This would require a longer display on the CodeSure card, at increased cost for the card.

Button sensitivity. The small size of the buttons and their closeness on the card were identified as being less of a usability problem than the contact pressure required in getting the buttons to 'press': "*The buttons are not very responsive. You have to push hard to work them*". Other button technologies (for example piezoelectric or capacitive button technology instead of the air gap / resistive button technology used in the baseline CodeSure card design) might be an improvement here in making buttons easier to press.

These usability issues are addressed in the next experiments which compare the usability of the baseline design of CodeSure card with an enhanced user interface design.

3.2. The Enhanced CodeSure Card Usability Experiment

In the previous usability experiment with the (baseline design) CodeSure card, issues which were impacting usability were identified and enhancements to address these issues were proposed. This experiment investigates the effectiveness of some of these enhancements in improving the usability of the CodeSure card.

Dual mode operation. An alternative method for mode selection was used for the enhanced design to give the user more feedback when selecting the card mode. In the enhanced design, pressing the OK button when the card is inactive displays a prompt message '**MODE _**' and when the mode number button is pressed, the number for the selected mode is displayed on the screen for one second or until OK is pressed.

Navigation cues. PIN entry was also changed in the enhanced design with use of a display prompt '**PIN** ____' to prompt the user to enter their PIN after mode selection.

PIN masking. Two different masking styles for the PIN digit display were also investigated, comparing the existing straight-to-star PIN masking design with an alternative digit-to-star PIN masking design.

Key-ahead. A keypad buffer is used with the enhanced design to store any key presses made before the display was updated, to alleviate the issue where users felt the keypad was not responsive enough when trying to key ahead.

Display format. Two different passcode display styles were also investigated, comparing a 4-space-4 digit passcode display design with the existing passcode display design of a single string of 1x8 digits.

The enhanced CodeSure card design featured the alternative method for mode selection and improved navigation cues to provide visible feedback to the user during mode selection, and incorporated a keypad buffer to allow type-ahead. The different PIN masking styles and passcode display formats were not tied to a specific card design and were investigated separately.

3.2.1. Experiment Design

For this experiment a 1:1 size simulation of the CodeSure card was created on a small touch screen tablet device to investigate some of these usability issues in more depth. There was no requirement to use the CodeSure card server for this experiment as there were no physical CodeSure cards being used.

The experiment involved two tasks (two CodeSure card modes) – login using Identify mode, and transaction signing using Sign mode. A web page very similar to that used in the previous experiment was used to simulate participant login and transaction signing. This was written in **Perl** using the CGI module and ran on Microsoft IIS. For participant login, the form contained two fields, '`User ID`' and '`Passcode`' and the prompt was changed to "*To log on, please enter your User ID and generate a Passcode using IDENTIFY mode on your card.*" For transaction signing, the form had an additional two fields labelled '`Account No`' and '`Amount`' and the prompt was changed to "*To proceed with the transfer of £*`AMOUNT` *to account number* `ACCOUNT`*, please enter your User ID and generate a Passcode using SIGN mode on your card.*", where `AMOUNT` and `ACCOUNT` were dynamically filled depending on the payment details in the task.

After entering the requested details and pressing the submit button, the details were logged for later analysis. No live checking on the content of the fields was done other than ensuring that the '`User ID`' or '`Passcode`' fields were not empty. If that was the case then a **JavaScript** alert box prompted them to "*Please enter values for both User ID and Passcode.*"

To characterise card usability and learning effects, participants were asked to attend the usability labs at the University. In their individual session they were asked to use the CodeSure card a total of 16 times (4 times with each of Identify and Sign modes with baseline and enhanced designs), generating a new and unique OTP to either login to a typical Internet page or sign a typical transaction. Participants were given paper User Guides on how to use each design (Figure B.2 through Figure B.4 in Appendix B). Task time [time to derive OTP + login] and success rates were measured for each use. After the first and fourth repeat of each task with each user interface design participants completed a usability questionnaire.

A cohort of representative banking customers (Edinburgh) took part in the experiment, using a CodeSure card simulator to login and sign transactions. None of the participants had taken part in the previous CodeSure experiment and they were recruited on a balanced quota basis with equal numbers of men and women: and equal numbers in two age groups – under 45's and those 45 and over[5].

[5] Results of the previous experiment showed no significant effect involving the intermediate age group so the 3 age groups of that experiment were simplified to 2 age groups for this second experiment.

The same usability statements used in the previous experiment were completed by each participant after their first and fourth uses of each design for each of the two tasks / modes. The attributes being measured did not change but the statement wording changed from ‘*this device*’ to ‘*this process*’ to focus interest on the usability of the overall task. As before, the questionnaire was presented as 25 randomised statements in a 7-point Likert format that ranged from ‘Strongly Agree’ to ‘Strongly Disagree’, Table 3.1.

The experiment aimed to address three aspects: the relative usability of the two card designs; the error rates for the two alternative passcode display modes (1x8 versus 4-space-4); and preferences for digit PIN masking approaches.

For each use, keystrokes were logged, time was measured, and the frequency of occurrence of key-ahead attempts was monitored. The order of experience for design and for mode was balanced across the cohort. Participants also experienced the two passcode displays (4-space-4 or 1x8 digits) alternating between uses to investigate if the enhanced display format improves transcription of the passcode from the card to the computer.

Individual sessions lasted 60 minutes. Participants were thanked with honorarium cheques of £30. The details of the experiment are summarised in Table 3.9.

The Enhanced CodeSure Card Usability Experiment	
Experiment purpose	Exploration of usability issues with enhanced CodeSure card design.
Null hypothesis (H_0)	There will be no differences in metrics of participant attitude to usability, task timings or error rates between different designs for the CodeSure card user interface.
Experiment design	Participants experience baseline design of CodeSure card and enhanced design which seeks to better partition mode selection from card usage; participants also experience two different display formats to assess error rates; participants give preference for PIN masking techniques. Repeated measures (balanced order), within subjects.
Dependent variables	Usability questionnaire (7-point Likert scale). Error rates with the different display formats.
Other data	Demographic and technographic data.
Independent variables	Experiment: treatment order (2 design orders × 2 mode orders, balanced). Participant: gender (2 genders, balanced), age group (2 groups, balanced).
Cohort	64 (2 age groups × 2 genders × 4 treatment orders × over-sampling ratio 4:1) = 64. Not the same participants as the previous experiment.
Honorarium	Personal cheque for £30.
Session time	60 minutes.

Table 3.9: Experiment Design Details

3.2.2. CodeSure Card Simulator

The CodeSure card simulator screen was exactly the same size as a real card. Button presses were simulated with circular regions matching the exact dimensions of the buttons on the real card, Figure 3.2. The touch-screen device used in the CodeSure card simulator was an ARCHOS 7 media tablet. The simulator was written in a combination of **ASP** code running on a remote Windows Server and **JavaScript** rendering in the Opera web browser running on the tablet.

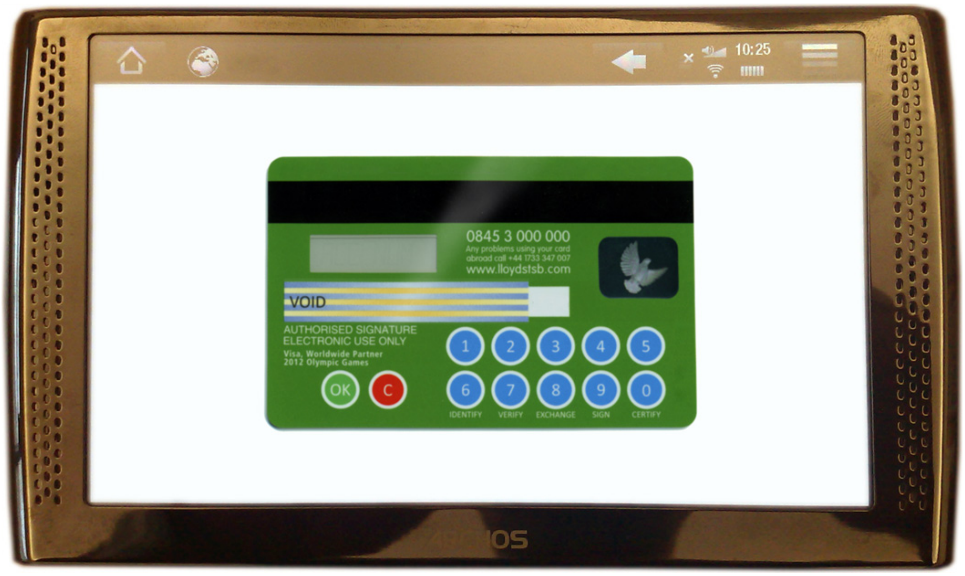


Figure 3.2: The CodeSure Card Simulator

3.2.3. Results

3.2.3.1. Demographic and Technographic Results

A good balance for gender (50% male and 50% female) and for age groups (49% aged 18 to 45 and 51% aged 46 and over) was achieved in the final sample of 70 participants.

Nearly all of the participants (99%) owned a mobile phone and of these, some 90% sent and received text messages, sending an average of 45 text messages a week, an important skill for an experiment involving usage of devices with small buttons and screens.

Finally, 16% of participants were noted as using their left hand to press the touch screen during the experiment.

During the experiment, data logs from the simulator were found to be incomplete for eleven participants. The cause of this was not known but the simulator did continue to function correctly. Therefore, task completion and timing data are only available for 59 participants, but usability and preference data are available for all 70 participants. Of the 11 participants with missing or incomplete task data, 10 of them experienced the enhanced design first, perhaps indicating that the cause of logging failure was related to a software problem with running the enhanced design first.

Even with this small imbalance of design order allocations there were enough samples in each of the subgroups to perform split analyses with the data.

3.2.3.2. Usability Results

The enhanced design scored higher for usability with an overall mean score of 5.01 than did the baseline design with a mean score of 4.83, Table 3.10. Usability scores in Sign mode were lower than in Identify mode, Table 3.10.

	Enhanced Design	Baseline Design
Overall	Mean=5.0073, SD=0.95983	Mean=4.8339, SD=1.08171
Identify mode	Mean=5.1563, SD=0.99044	Mean=4.9957, SD=1.07509
Sign mode	Mean=4.8593, SD=1.00752	Mean=4.6720, SD=1.16992

Table 3.10: Mean Usability Scores for the Two Designs (N=70)

A repeated-measures analysis of variance (ANOVA) was carried out for the overall mean usability scores (Table 3.11) taking design (baseline vs. enhanced) as the within-subjects factor and gender, age group and order of design experienced as between-subjects factors. An exploratory analysis beforehand was able to discount task (mode) order as not being a significant between-subjects factor so this was removed from any subsequent analysis. For this analysis the mean of the participant's usability scores of first and fourth use of each design/mode were used and only main effects and two-way interactions were explored.

The overall difference in mean usability scores between the enhanced design (mean=5.01) and the baseline design (mean=4.83) was approaching significance, $F(1,62)=3.806$, $p=0.056$, Table 3.11. There were no significant main effects on the mean attitude scores due to gender or age group, and also no significant between-subjects interactions, Table 3.11. There was a significant effect due to design order, $F(1,62)=25.141$, $p<0.001$, Table 3.11, but the order of experience was equally balanced across the cohort.

Source	Sum of Squares	*df*	Mean Square	*F*	Sig.
Within-Subjects Effects					
Design	0.745	1	0.745	3.806	0.056
Design × Gender	0.390	1	0.390	1.993	0.163
Design × Age	0.151	1	0.151	0.774	0.382
Design × Order	4.920	1	4.920	25.141	0.000
Error (Design)	12.132	62	0.196		
Between-Subjects Effects					
Gender	1.734	1	1.734	0.897	0.347
Age	1.275	1	1.275	0.660	0.420
Order	0.029	1	0.029	0.015	0.902
Gender × Age	1.317	1	1.317	0.682	0.412
Gender × Order	0.004	1	0.004	0.002	0.965
Age × Order	2.298	1	2.298	1.189	0.280
Error	119.795	62	1.932		

Table 3.11: ANOVA Results for Overall Mean Usability

Similar ANOVA tests on each of the individual usability attributes show that there were significant main effects due to design, gender or age group in 7 of the 25 attributes, Table 3.12.

Attribute	Significant Effects	Details
Flustered	Age (*df*=1; *F*=7.250; *p*=0.009)	Age 18 to 44 > Age 45 and over
Stressed	Age (*df*=1; *F*=4.592; *p*=0.036)	Age 18 to 44 > Age 45 and over
Frustration	Gender (*df*=1; *F*=4.212; *p*=0.044)	Females > Males
Screen response	Design (*df*=1; *F*=6.199; *p*=0.016)	Enhanced > Baseline design
	Gender (*df*=1; *F*=5.402; *p*=0.023)	Females > Males
Easy to use buttons	Design (*df*=1; *F*=7.518; *p*=0.008)	Enhanced > Baseline design
Use again	Design (*df*=1; *F*=4.385; *p*=0.040)	Enhanced > Baseline design
User friendly	Design (*df*=1; *F*=4.451; *p*=0.039)	Enhanced > Baseline design

Table 3.12: Summary of Significant Differences in Usability Attributes

ANOVA tests for the individual attributes show that the enhanced design scored significantly higher for overall usability than the baseline design for ‘screen response’, ‘easy to use buttons’, ‘use again’ and ‘user friendly’. In Table 3.12, high scores for ‘flustered’, ‘stressed’ and ‘frustration’ imply that participants were actually less flustered, less stressed and less frustrated as these were presented as negative polarity statements. Females found both designs less frustrating than did males and also scored both designs higher for ‘screen response’. Younger participants found that both designs left them less flustered and less stressed than did older participants, Table 3.12.

For the eight sets of usability questionnaire data, reliability analysis on the scale of each (based on Cronbach’s α) reported a minimum value of α=0.959 (for the first questionnaire for the enhanced design in Sign mode), indicating that the questionnaire reliability was high.

3.2.3.3. Timings

Card usage times, taken as being the time from the first button press to the passcode being displayed, were measured with the CodeSure card simulator for each task, Table 3.13. For Identify mode, mean card usage times for the enhanced design and baseline design were similar, Figure 3.3, but mean card usage times in Sign mode appear faster by 9 seconds with the enhanced design, Figure 3.4.

	Enhanced Design (s)	Baseline Design (s)
Identify mode	Mean=25.3153, SD=25.64611	Mean=24.9932, SD=14.60913
Sign mode	Mean=63.2481, SD=25.12291	Mean=72.4793, SD=46.75887

Table 3.13: Mean Device Timing Results for the Two Designs (N=59)

Statistical tests (ANOVA) on these data indicate that the difference in card usage times between the enhanced design and the baseline was not statistically significant for either mode. Removing the 3 outliers reduces the mean card usage time for Identify mode in the enhanced design to 21.41s (24.56s in the baseline design) and reduces the mean card usage time for Sign mode in the baseline design to 67.84s (62.13s in the enhanced design), and although this resulted in reduced card usage times the difference remains not statistically significant.

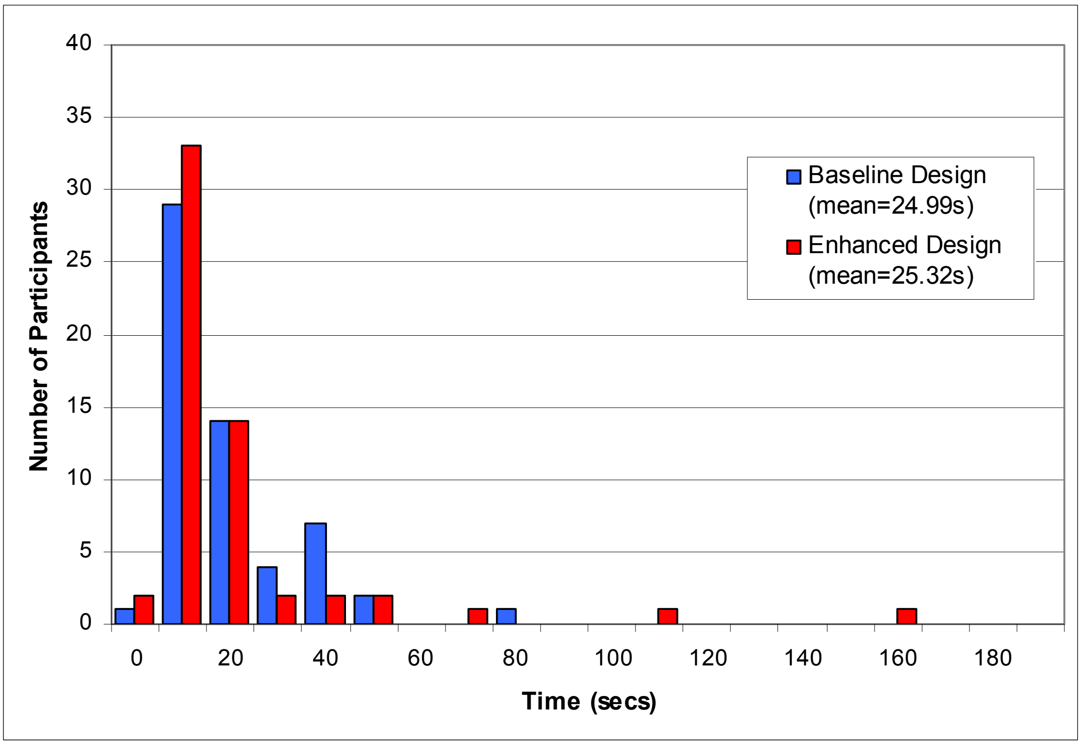


Figure 3.3: Card Usage Times – Identify Mode

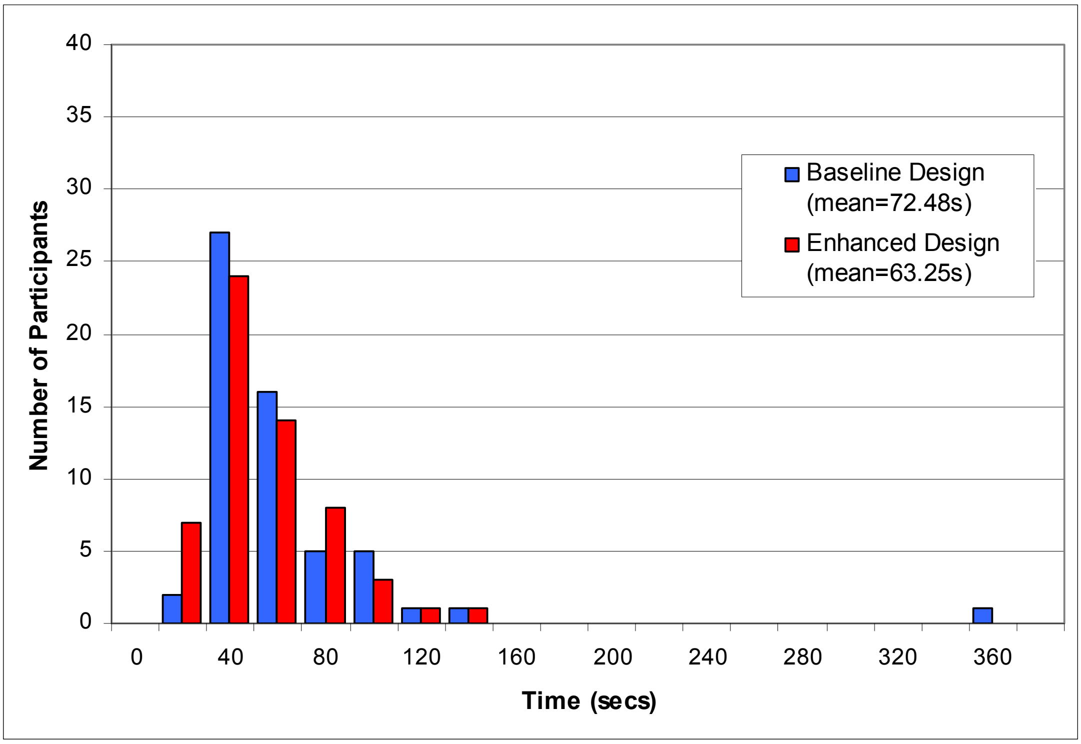


Figure 3.4: Card Usage Times – Sign Mode

The overall task times, taken as being the time from the task web page being displayed until the time of hitting submit were measured, Table 3.14. On average overall task times for the enhanced design were faster than with the baseline design by 7 seconds in Identify mode, Figure 3.5, and by 13 seconds in Sign mode, Figure 3.6.

	Enhanced Design (s)	Baseline Design (s)
Identify mode	Mean=63.1907, SD=40.21658	Mean=70.2119, SD=31.77859
Sign mode	Mean=103.6890, SD=39.73001	Mean=116.9534, SD=60.19948

Table 3.14: Mean Task Timing Results for the Two Designs (N=59)

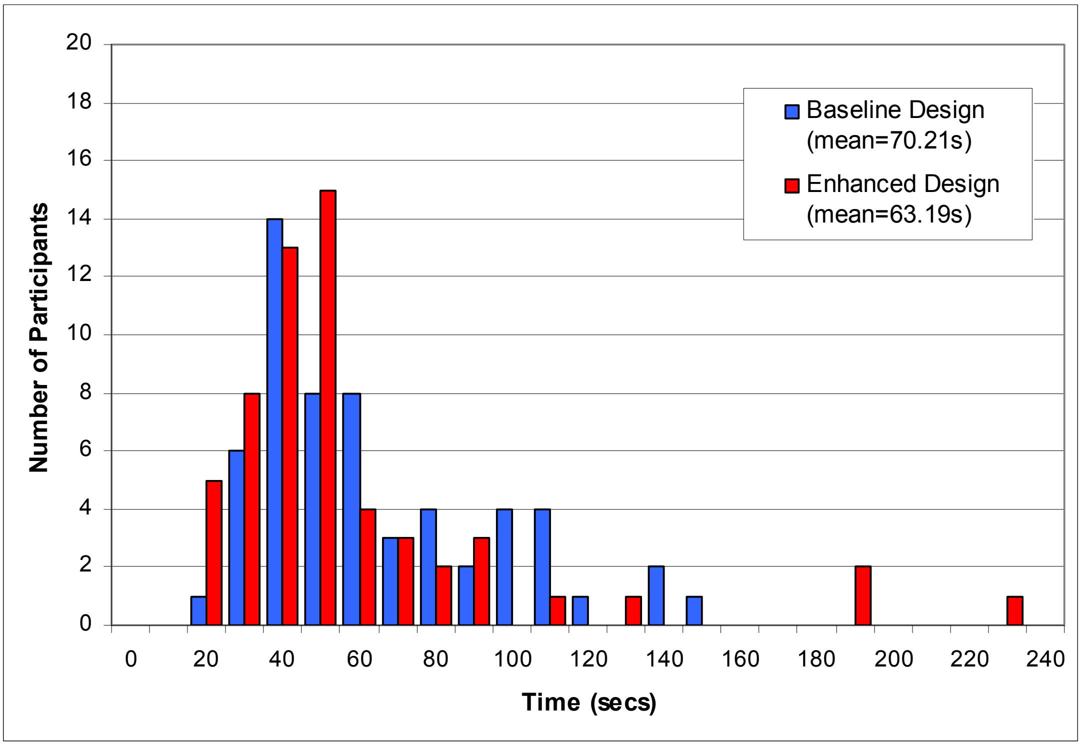


Figure 3.5: Overall Task Times – Identify Mode

Again, statistical tests (ANOVA) on these data show that the difference in task times between the enhanced design and the baseline was not statistically significant for either mode. Removing the 4 outliers reduces the mean for Identify mode in the enhanced design to 55.55s (68.83s in the baseline design) and reduces the mean for Sign mode in the baseline design to 110.65s (99.27s in the enhanced design), resulting in a statistically significant difference for task times in Identify mode, *F*(1,51)=11.845, *p*=0.001, although the difference for Sign mode remains not statistically significant.

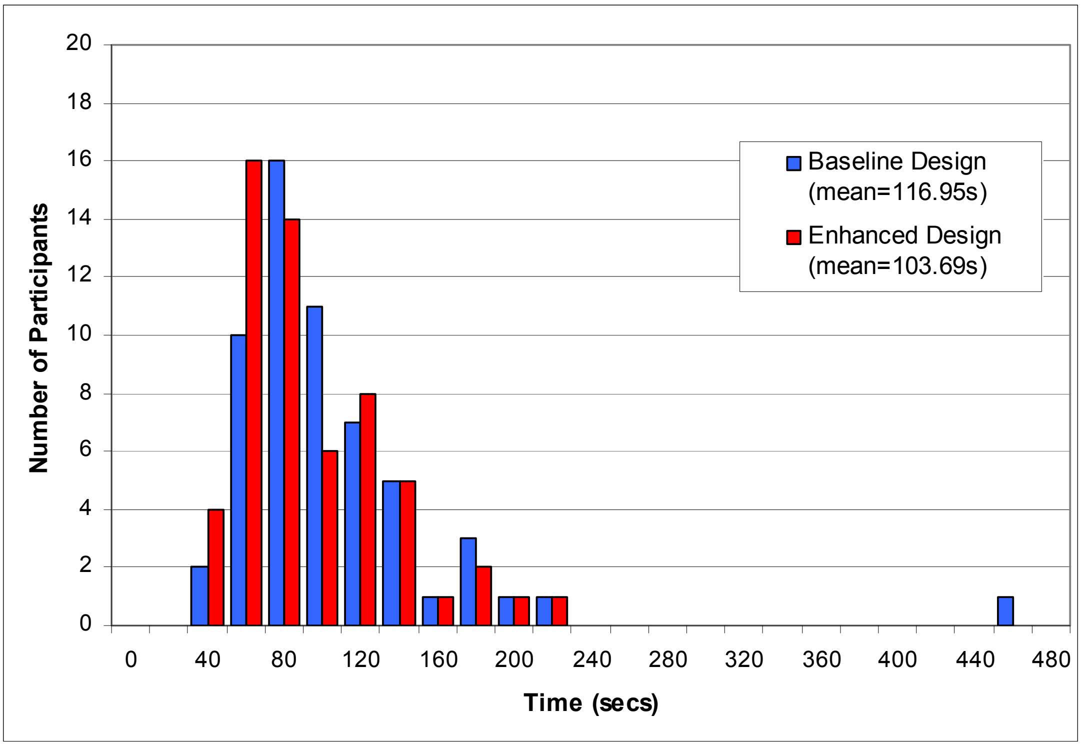


Figure 3.6: Overall Task Times – Sign Mode

3.2.3.4. Success Rates

In both modes of operation the PIN that was entered into the CodeSure simulator was recorded for later analysis, Table 3.15. Success rates with the enhanced design were slightly higher than with the baseline design. During mode selection, only two participants selected the wrong mode for some of their tasks, and although this was recorded the CodeSure card simulator still generated a Passcode so that they could continue their task.

Mode	Baseline Design (%)	Enhanced Design (%)
Identify	97%	99%
Sign	97%	98%

Table 3.15: PIN Entry Success Rates – both Modes

With Sign mode, both the account number and amount (in pence) entered were recorded by the CodeSure card simulator, with results summarised in Table 3.16. A passcode was generated regardless of whether or not the account number and amount were correct for the task.

	Baseline Design (%)	Enhanced Design (%)
Correct account number	89%	94%
Correct amount (but specified in pounds)	35%	34%
Correct amount	56%	59%

Table 3.16: Data Entry Success Rates – Sign Mode

Again, success rates for entry of account number and amount in Sign mode were higher for the enhanced design than for the baseline design: evidence of the success of the key-ahead functionality of the enhanced design. Although participants were asked to enter the amount in pence, just over one third of amounts were entered in pounds.

3.2.3.5. Impact of Display Format

Changing the format of the 8 digit passcode display into two blocks of 4 digits resulted in a slightly higher success rate when transcribing the passcode from the CodeSure card display to a web page, Table 3.17. Errors reading and then entering the 8-digit passcode were reduced from 3.4% to 1.9% by use of two blocks of 4 digits in the passcode display.

Result	Block of 8 Digits (%)	Two Blocks of 4 Digits (%)
Passcode entry success	95.5%	97.3%
Wrong (PIN entered)	1.1%	0.6%
Wrong (User ID entered)	0.0%	0.2%
Wrong (other)	3.4%	1.9%

Table 3.17: Error Rates for Passcode Entry – by Display Format

3.2.3.6. Key-ahead Frequency

In the baseline design, the number of key-ahead attempts that were made was recorded for each use, Table 3.18.

	Identify	Sign	Overall
% of participants	83%	78%	81%
Mean attempts per task	0.55	0.68	0.61

Table 3.18: Key-ahead Attempts with the Baseline Design

Overall, 81% of participants attempted to key-ahead at least once during their eight uses of the baseline design, with an average of 0.61 attempts per use.

More participants pressed keys ahead of time in a task using Identify mode compared to Sign mode, although there were more attempts per use when using Sign mode, which is understandable given the increased number of button presses required.

Interestingly, participants using Identify mode for the first time in the baseline design made an average of 0.81 key-ahead attempts which decreased to less than 0.6 attempts for the fourth use, whereas the average number of key-ahead attempts for the first use of Sign mode was 0.64 which did not decrease on subsequent uses. This is further evidence of the value of the keypad buffer in the enhanced design.

3.2.3.7. PIN Masking Preferences

After performing the two tasks for both designs, participants were asked to comment on an alternative PIN masking design on the CodeSure card simulator where the digits were immediately masked with a star rather than initially showing the PIN digit and then masking it. They were then asked which design they preferred and why.

The majority of those that preferred the digit-to-star design, Table 3.19, indicated that it was because they could "*Check they had entered the correct number*", whilst those that preferred the straight-to-star design said that it "*Would prevent others seeing the PIN*".

Preference	%
Digit to star	61%
Straight to star	33%
No preference	6%

Table 3.19: PIN Masking Preferences

3.2.4. Discussion on Experiment Results

The null hypothesis (H_0) for this experiment was:

> *"There will be no differences in metrics of participant attitude to usability, task timings or error rates between different designs for the CodeSure card user interface."*

The evidence presented here is sufficient to refute this null hypothesis since usability scores with the enhanced design of the card were superior to those with the baseline card design. There is also evidence to suggest that task timings and success rates were also improved. Statistical tests show the difference in usability scores between the enhanced card design and the baseline card design was statistically significant. The significant differences (improvements) reported in scores for 'screen response' and 'easy to use buttons' can be attributed to inclusion of a keypad buffer in the enhanced design and the significant differences (improvements) reported for 'use again' and user friendly' can be attributed to the improvement in navigation cues for mode selection and data entry in the enhanced card design.

Usage times and task times decreased with the enhanced design but the differences were not statistically significant. However, removing the 4 outliers from the task timing results resulted in a significant reduction in task times in the enhanced design when using Identify mode.

Success rates for data entry were slightly higher with the improved navigation cues and keypad buffer in the enhanced design. Errors in reading and then entering the 8-digit passcode were reduced with the 2 blocks of 4 digits format for passcode display, and users expressed a preference for the digit-to-star mode of PIN masking. The fact that 82% of participants attempted key-ahead button presses with the baseline design is evidence of the need for a keypad buffer as used in the enhanced design.

3.3. The CodeSure Keypad and Display Synchrony Experiment

The previous experiment identified that the usability attributes ‘easy to use buttons’ and ‘screen response’ were higher for the enhanced user design for the CodeSure card in comparison to the baseline design and the difference was statistically significant. However, neither the buttons themselves, nor the screen response times, were actually different between the two designs, implying that the keypad buffer in the enhanced design was the key contributor to this improvement.

To further investigate this design aspect, a touch screen simulation of the CodeSure card (enhanced user interface design as discussed above) was prepared on a desktop PC (touch) screen connected to a gaze tracking camera making it possible to determine if participants gazed at the display of the CodeSure card after each button press[6]. If they did not gaze at the display after button presses then they would be relying on the button presses to be registered on the display, a false assumption if they typed too quickly (key-ahead) without a keypad buffer and this would contribute to their perceptions of slower screen response and buttons being harder to use buttons.

3.3.1. Experiment Design

The experiment used a touch screen display simulation of the CodeSure card, Figure 3.7. The ELO Intellitouch touch screen technology used was built in to a 17 inch Iiyama LCD monitor with Surface Acoustic Wave (SAW) technology.

[6] The gaze tracking camera available for this research is not able to track gaze for a physical CodeSure card because the target points for the gaze tracker need to be fixed relative to the camera position.

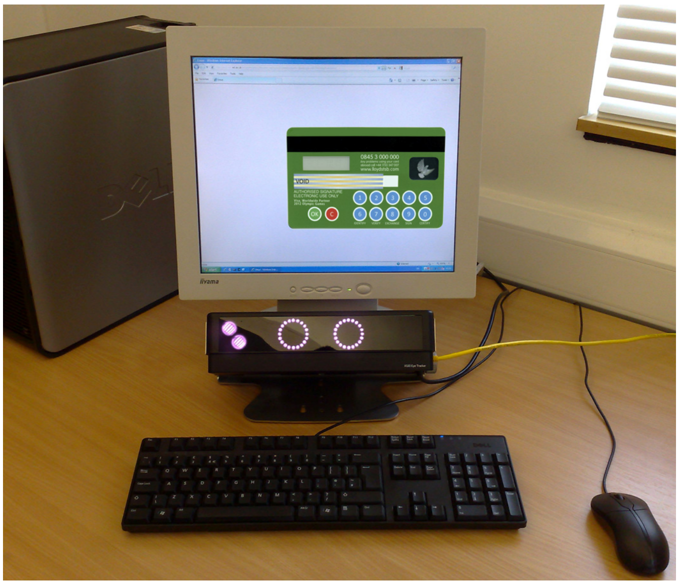


Figure 3.7: Experiment Setup

The dimensions of the simulated card on the screen (exactly 2.5 times the original size) was designed to be of a similar perspective size, for a participant sitting approximately 75cm from the touch screen, as a physical CodeSure card when held at some 30cm from the eyes.

The simulator was written in a combination of **ASP** code running on a remote Windows Server and **JavaScript** rendering in the Internet Explorer web browser running on the desktop PC.

The gaze tracking camera used was the Tobii X120 which uses infrared beams to detect reflections (Purkinje images) from the corneas of participants' eyes that can be used in conjunction with the pupil position to calculate gaze direction (the human cornea is non-spherical). A calibration stage is necessary for each participant before using the eye tracker so that gaze direction can be mapped directly on to screen co-ordinates. This is especially important with the X120 as it is not fixed rigidly to the display.

The human eye is capable of focusing on stationary objects for small periods of time (fixations) separated by periods of jitter (saccades). By applying a fixation filter to filter out saccades from an eye tracking log, the fixations of interest can be identified.

The fixation filter used in this experiment was a dispersion-threshold identification (I-DT) algorithm, Figure 3.8 (based on Salvucci and Goldberg, 2000). For this experiment the minimum duration of a fixation was set to 50ms and the maximum radius of a fixation was 50 pixels.

```
$d \leftarrow$ minimum duration of a fixation (ms)
$r \leftarrow$ maximum radius of a fixation (pixels)
while $p \leftarrow$ next gaze point do
  $F = \{p\}$ // start a new fixation
  $c \leftarrow p$ // initialise the fixation centroid
  while $p \leftarrow$ next gaze point and $p \leq r$ pixels away from $c$ do
    $F = F \cup \{p\}$
    $c \leftarrow$ centroid of all points in $F$
  end while
  if overall duration of $F \geq d$ then
    replace all points in $F$ with fixation, $c$
  else
    mark all points in $F$ as saccades
  end if
end while
```

Figure 3.8: Dispersion Threshold Identification Fixation Filter Code

It is important to note that not all of the gazes to the screen could be captured in the simulator, particularly during button presses where an arm might have obscured the gaze tracking camera. However, even having only one detected fixation to the appropriate area of the screen was judged enough to be included in the analysis.

At the start of their session, the gaze tracking camera was calibrated to suit the participant's eyes / gaze by asking the participant to follow and focus on a set of 5 target points which appeared in sequence at a set of fixed positions over the screen area.

Participants were asked to generate one-time passcodes using the Identify mode of the CodeSure card simulator and also using a real CodeSure card. As with the previous experiments participants were provided with a paper User Guide for the CodeSure card (covering both the simulated and real form), as shown in Figure B.5 in Appendix B. None of the participants had taken part in either of the two experiments reported earlier in this chapter.

The initial batch of five passcode generation tasks used the CodeSure simulator on the touch screen display with the gaze tracking camera. This first batch of five uses with the simulator / gaze tracking camera was then followed by a batch of five passcode generation tasks with a physical CodeSure card. A further batch of five passcode generation tasks used the CodeSure simulator and gaze tracking camera, to investigate if gaze behaviour differed after use of the physical CodeSure card.

For each task, participants generated the one-time passcode and were asked to write down the passcode for subsequent checking.

Participants were thanked with an honorarium cheque for £30. The details of the experiment are summarised in Table 3.20.

The CodeSure Keypad and Display Synchrony Experiment	
Experiment purpose	Exploration of gaze behaviour when using a CodeSure card.
Null hypothesis (H_0)	There will be no difference in participant gaze behaviour between the first five uses of the CodeSure card simulator and the second five uses.
Experiment design	Participants use the CodeSure card simulator / gaze tracking camera 5 times to generate one-time passcodes, followed by 5 similar uses of a physical CodeSure card, followed by another 5 uses of the CodeSure card simulator / gaze tracking camera. Repeated measures (balanced order), within subjects.
Dependent variables	Gaze tracking data.
Other data	Demographic and technographic data. Performance data: simulator error rates.
Independent variables	Participant: gender (2 genders, balanced), age group (2 groups, balanced, split ages 18 to 44 / 45 and over).
Cohort	40 (2 age groups × 2 genders × over-sampling ratio 10:1) = 40.
Honorarium	Personal cheque for £30.
Session time	60 minutes.

Table 3.20: Experiment Design Details

3.3.2. Results

3.3.2.1. Demographic and Technographic Results

A cohort of 40 participants took part in the experiment. Data from 2 participants were removed from the analysis because they were unable to complete the practice gaze tracking calibration due to eye problems (one who reported suffering from a 'lazy eye' and one who was about to have an operation to remove cataracts).

Of the remaining 38 participants, 19 were male and 19 were female: 20 were aged 18 to 44 and 18 were aged 45 and over. There were 14 participants who wore glasses during the experiment and 5 participants wore contact lenses although none of these cases caused any problems for the gaze tracking camera.

Only 4 participants were left-handed and the remainder were right-handed. The simulated card was positioned on the touch screen display according to the participant's handedness during the experiment in order to minimise the likelihood of an arm blocking the line of sight for the gaze tracker when pressing the buttons on the screen.

3.3.2.2. Task Results

For the simulator tasks only 2 tasks had the PIN submitted wrongly (the participants were not notified in these cases since a passcode was generated regardless). There were many more tasks where the PIN was initially entered wrongly but corrected by the participant before submission. There were 18 tasks where the OK button was pressed to confirm the mode selection rather than just waiting for the mode selection to automatically proceed. This issue only occurred for 4 participants and did not affect their ability to generate a passcode.

Participants were also asked to write down each passcode as it was generated in the simulator tasks and by the real CodeSure card. For the simulator tasks, 100% of passcodes were recorded correctly. The passcodes generated by the physical CodeSure card were not checked during the experiment.

3.3.2.3. Gaze Tracking Results

Each participant provided gaze tracking data from the first batch of 5 simulator tasks before using the physical CodeSure card and from the second batch of 5 simulator tasks. There were data for 379 simulator tasks in total. There was a technical issue with the eye tracking hardware with one participant where the first simulator task was not run.

Boolean values recording if a participant gazed at least once at the simulator display after each button press for each task were calculated from the fixation-filtered gaze tracking data and the results summed for each participant, per button, per task and per batch of 5 tasks.

Table 3.21 shows the percentage of participants who gazed at the display after each button press in the sequence <button OK then button 6> for mode selection (Identify mode), followed by the four buttons for PIN digit entry. These results, Table 3.21, show that on some 40% of occasions participants did not look at the display after entering a PIN digit. Interestingly, Table 3.21, on some 20% of cases they did not even look at the display after mode selection. Only 13% of participants looked at the display after every single PIN digit entry.

	% of Participants who Gazed at Display		
State in Sequence	**First Batch**	**Second Batch**	**Overall**
After OK button	73.54%	59.47%	66.49%
After mode select digit button	85.19%	75.26%	80.21%
After first PIN digit	56.61%	61.05%	58.84%
After second PIN digit	62.96%	65.26%	64.12%
After third PIN digit	66.14%	61.05%	63.59%
After fourth PIN digit	100.00%	100.00%	100.00%
Overall (all 6 buttons)	74.07%	70.35%	72.21%

Table 3.21: Display Gazes After Button Presses for <OK [Mode 6] DDDD>

Between the first batch and second batch of simulator tasks there was a 14% decrease in gazes to the display after pressing the OK button, and a 10% decrease after mode selection. This shows an increased familiarity with the process of mode selection, resulting in less need to check that the mode was correctly selected on the display. There were only slight differences in gaze behaviour after entering the first, second and third PIN digits between the first and second batch of simulator tasks.

The data, Table 3.21, confirm however that for 100% of uses participants gazed at the display after the fourth PIN digit – since of course they had then to read the passcode from the display.

A repeated-measures analysis of variance (ANOVA) was carried out to investigate overall gaze behaviour, Table 3.22, taking use (first batch of 5 tasks with the simulator vs. second batch of 5 tasks with the simulator) as the within-subjects factor and gender and age group as the between-subjects factors. The within-subjects factor (use) was measured on an integer scale of 0 to 5 representing the number of tasks in which the participant gazed at least once at the display. Only main effects and two-way interactions were explored.

Source	Sum of Squares	*df*	Mean Square	*F*	Sig.
Within-Subjects Effects					
Use	19.799	1	19.799	2.202	0.147
Use × Gender	2.471	1	2.471	0.275	0.604
Use × Age	3.378	1	3.378	0.376	0.544
Error (Use)	305.678	34	8.991		
Between-Subjects Effects					
Gender	216.178	1	216.178	2.490	0.124
Age	732.258	1	732.258	8.435	0.006
Gender × Age	150.915	1	150.915	1.738	0.196
Error	2951.722	34	86.815		

Table 3.22: ANOVA Results for Gaze Behaviour between First and Second Batches of Simulator Tasks

There was no significant effect on overall gaze behaviour between the first and second batches of simulator tasks, although there was a significant effect of age group on gaze behaviour with younger participants gazing at the display significantly more, *F*(1,34)=8.435, *p*=0.006, than did older participants, Table 3.22.

Further ANOVAs were run to investigate gazes to the display after each button press, Table 3.23, between the first batch and second batch of simulator tasks. The analyses confirm that the decrease in rate of gazes to the display when selecting the operating mode between the first and second batch of simulator tasks was statistically significant, *F*(1,34)=11.791, *p*=0.002. Similarly for the decrease in rate of gazes to the display for the mode select digit, *F*(1,34)=5.003, *p*=0.032. There were no effects due to use for the PIN digits, although there was between-subject effect of age group with younger participants looking at the display more than did older participants – significant differences for first PIN digit, *F*(1,34)=8.276, *p*=0.007; for second PIN digit, *F*(1,34)=11.105, *p*=0.002; and for third PIN digit, *F*(1,34)=6.670, *p*=0.014. The analyses did not include gaze behaviour after the fourth PIN digit since all participants gazed at the display after that event.

After Button Press	Significant Effects	Details
OK	Use (*df*=1, *F*=11.791, *p*=0.002)	First 5 uses > Second 5 uses
	Age (*df*=1, *F*=5.031, *p*=0.032)	Age 18 to 44 > Age 45 and over
Mode select digit	Use (*df*=1, *F*=5.003, *p*=0.032)	First 5 uses > Second 5 uses
First PIN digit	Age (*df*=1, *F*=8.276, *p*=0.007)	Age 18 to 44 > Age 45 and over
Second PIN digit	Age (*df*=1, *F*=11.105, *p*=0.002)	Age 18 to 44 > Age 45 and over
Third PIN digit	Age (*df*=1, *F*=6.670, *p*=0.014)	Age 18 to 44 > Age 45 and over

Table 3.23: Significant Effects in Gaze Behaviour between First and Second Batches of Simulator Tasks

Men tended to look at the display more frequently than did women, Figure 3.9, which shows that 14 men gazed at the display more than 50% of the time after entering a PIN digit, compared to 11 women.

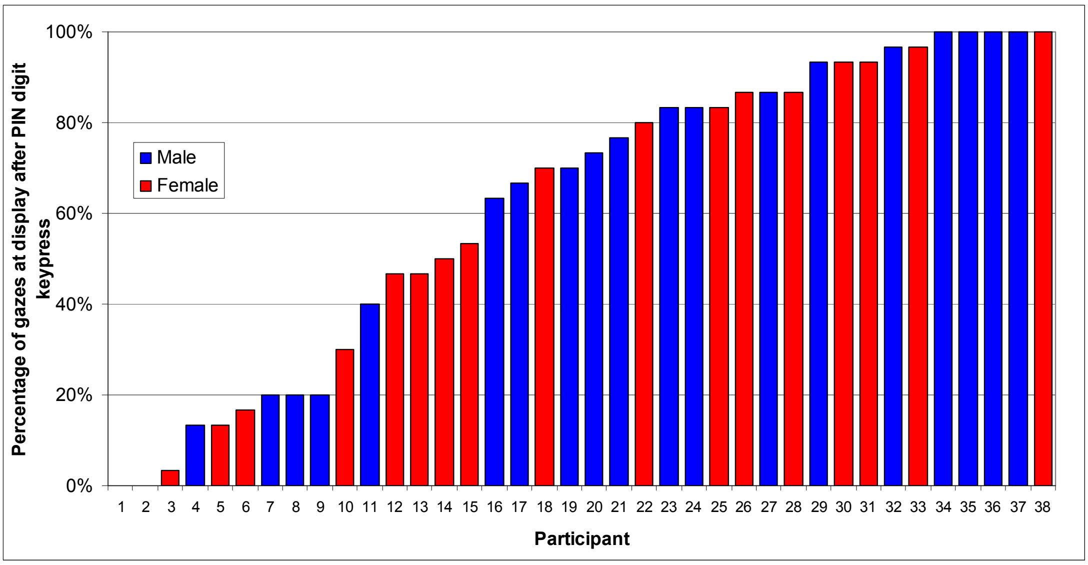


Figure 3.9: Display Gazes after Button Press – By Gender for PIN Entry

Younger participants tended to look at the display more than did older participants, Figure 3.10, which shows that 16 younger participants gazed at the display more than 50% of the time after entering a PIN digit, compared to 8 older participants.

Further repeated-measures ANOVAs were run to investigate gaze behaviour between successive uses of the first batch of 5 simulator tasks, Table 3.24.

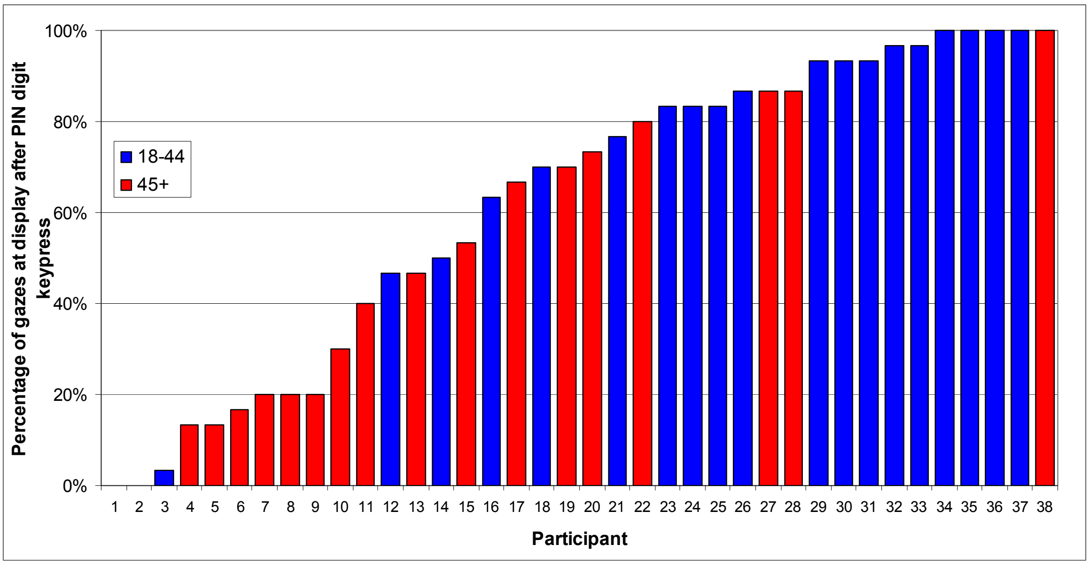


Figure 3.10: Display Gazes after Button Press – By Age for PIN Entry

The only significant effect of use on gaze behaviour was between the second and third uses, *F*(1,34)=4.744, *p*=0.036, Table 3.24, where there were more gazes to the display in the third use compared to the second use. There were some between-subjects effects of age group on gaze behaviour, confirming the previous findings in Table 3.22.

Use	Significant Effects	Details
Use 1 vs. Use 2	Age (*df*=1, *F*=5.791, *p*=0.022)	Age 18 to 44 > Age 45 and over
Use 2 vs. Use 3	Use (*df*=1, *F*=4.744, *p*=0.036)	Use 3 > Use 2
	Age (*df*=1, *F*=6.338, *p*=0.017)	Age 18 to 44 > Age 45 and over
Use 3 vs. Use 4	Age (*df*=1, *F*=5.163, *p*=0.030)	Age 18 to 44 > Age 45 and over
Use 4 vs. Use 5	Age (*df*=1, *F*=5.665, *p*=0.023)	Age 18 to 44 > Age 45 and over

Table 3.24: Significant Effects in Gaze Behaviour between Successive First 5 Uses of Simulator

3.3.3. Discussion on Experiment Results

The null hypothesis (H_0) for this experiment was:

> *"There will be no difference in participant gaze behaviour between the first five uses of the CodeSure card simulator and the second five uses."*

The evidence presented here supports this null hypothesis, as far as behaviour with PIN digits goes where gaze behaviour only differed slightly overall with repeated use after entering each PIN digit. However, there was a statistically significant decrease in gazes to the display after pressing the OK button and the mode selection button, between the first batch of five uses and the second batch of five uses.

Only 13% of participants looked at the display after every single PIN digit entry. Overall some 40% of button presses were not associated with gazes at the display by participants, confirming the importance of key-ahead functionality with a keypad buffer in the design of the CodeSure card.

Interestingly, younger participants tended to look at the card display more than did older participants after each button press during PIN entry. Similarly men tended to look at the display more than did women.

Taken together these data confirm the important role being played by the keypad buffer in the enhanced CodeSure card design.

3.4. Summary

Usability scores for the CodeSure card at 4.49 (on a 7-point scale) for first use rising to 4.66 on tenth use confirm the device usability as suited for use in a usable security procedure.

The enhanced card design, featuring improved mode selection and navigation cues and importantly, a keypad buffer to allow key-ahead served to further boost usability scores for the CodeSure card to 5.01 with improved usage times and improved task times and success rates.

Participants recognised that the CodeSure card was convenient to use as reflected in the quality ratings and comments; and appreciated the role of the card as a multi-channel, usable security solution where 82% expected to be able to use it to access their Bank's Internet banking service, 73% to use if for Telephone banking and 88% to use it for online shopping. Some participants would even expect to use direct channels more often when using the CodeSure card.

This use of the CodeSure card as a common solution for usable security in multi-channel two-factor authentication procedures is investigated in the next chapter.

Chapter 4. Common Multi-Channel Authentication Based on the CodeSure Card

One of the advantages of the CodeSure card expounded in the previous chapter is that it offers acceptable levels of usability, rendering it suitable for consideration as a common form factor component in usable security procedures applicable over all banking and eCommerce channels in the physical world and in the digital world. This chapter reports results from an experiment to investigate such a usable security role in multi-channel security (specifically Internet shopping, Internet banking and Telephone banking). Results are presented for a comparison of user attitudes to usability and perceived security from their experience with a unified security procedure based on the CodeSure card with their experience with disparate, channel-specific security procedures.

4.1. The CodeSure Multi-Channel Usability Experiment

The experiment sought to engage participants in scenarios simulating a series of typical financial transactions – making an Internet purchase, then logging in to Internet banking to check their purchase and then phoning Telephone banking to check their balance. Participants experienced all three transactions using disparate channel-specific procedures in the one case: and experienced all three transactions using a common security procedure based on the CodeSure card in the other case.

4.1.1. Experiment Design

For this experiment, participants were purposely recruited on the basis that they were familiar with the use of their bank's Internet banking and Telephone banking services ('power users'), and that they were comfortable and familiar with Internet / online shopping. Volunteer participants were provided with a fictitious persona and fictitious account numbers and passwords. The procedure of using a persona allows participants to feel secure that their personal details remain confidential, and it allows control of what each participant encounters in the experiment, ensuring a standardised experience that can be compared across participants and groups.

The experiment scenarios were based around participants using their (fictitious) customer details, account numbers and security details to experience three disparate channel-specific security procedures to "make an online purchase of a DVD", experiencing a typical one-factor security procedure based on a password in Visa's 3-D Secure[1] Verified by Visa challenge: to then log in to their bank's Internet banking service to "check on their purchase", experiencing a second, different one-factor security procedure for Internet banking involving (different) password and memorable information (password) challenges: and to then "imagine a few days had passed" and phone their bank's Telephone banking service to "check their balance", experiencing a third, different, one-factor security procedure involving yet another (different) Telephone banking security number (password) challenge. In this scenario, participants were thus being asked to perform tasks involving a total of four different security passwords.

Participants also completed a similar scenario and set of financial transactions (in the same sequence: Internet purchase, Internet banking, Telephone banking) using a common security procedure based on OTP generation using the Identify mode of the

[1] 3-D Secure is an industry standard, extra security layer for online (card not present) transactions, branded as Verified by Visa and MasterCard SecureCode.

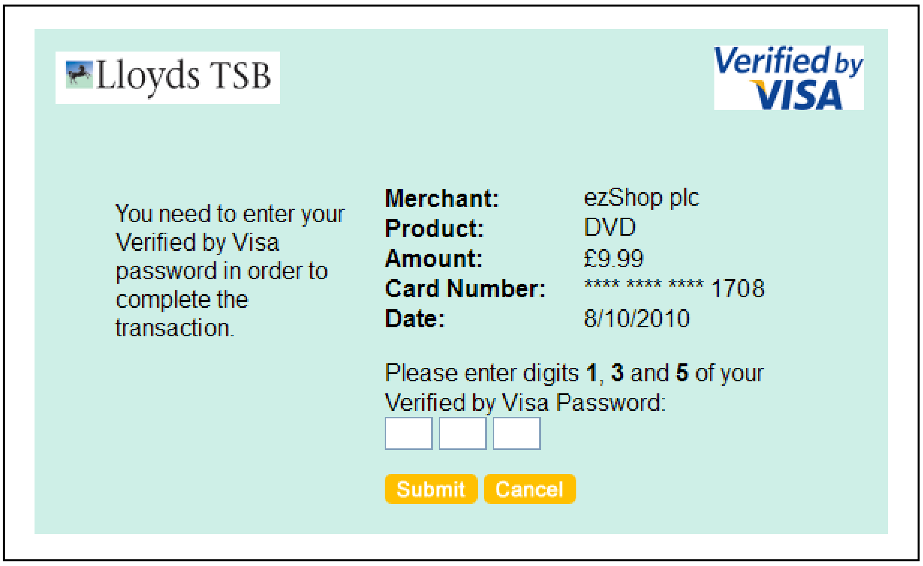
CodeSure card. In the experiment, the order of presentation of the two scenarios was controlled and balanced across the cohort of volunteer participants.

For the experiment, a music store web site modelled on a popular commercial web site for which the participant's persona "had an existing account" was created. This allowed the convenience of pre-populating fields for the customer account information in the experiment, thereby minimising any unnecessary typing required by the participant to enter only the 3-digit security code (shown on the reverse of their credit card) in the CVV[2] field, followed by three random digits from their Verified by Visa password, Figure 4.1, to complete the transaction. For the security procedure based on the CodeSure card, the Verified by Visa password screen was replaced by a screen asking the participant to generate an OTP with their CodeSure card (Identify mode), Figure 4.2.

The Verified by Visa challenge, although adopting an industry standard procedure, has encountered consumer resistance because the password challenge only ever appears after a purchase has been made on a web site and users tend to assume the password being asked for is their account password on the site, causing input errors, user frustration and abandoned sales. Also, if the customer fails the password challenge they are presented with a 'password reset' screen which asks them to enter their card CVV, expiry date, the name on the card, their date of birth and email address; and then create a new password, none of which seems relevant to the transaction itself and none of which suggests to the customer that they should memorise this Verified by Visa password as being associated with their card for use with future purchases. Use of the (common) CodeSure OTP challenge removes the problems associated with generating and remembering the Verified by Visa password. Finally, the Verified by Visa password is by design complex. It has to be generated by the customer; it has to be between 8 and 15 characters; it has to include at least two letters and two numbers; and at least one of the letters has to be lower

[2] Card Verification Value (also known as Card Security Code – CSC) consists of 3 digits printed on the signature strip of a debit or credit card.

case and at least one has to be upper case. The security procedure with the CodeSure card only requires the user to remember the existing everyday PIN for their card.



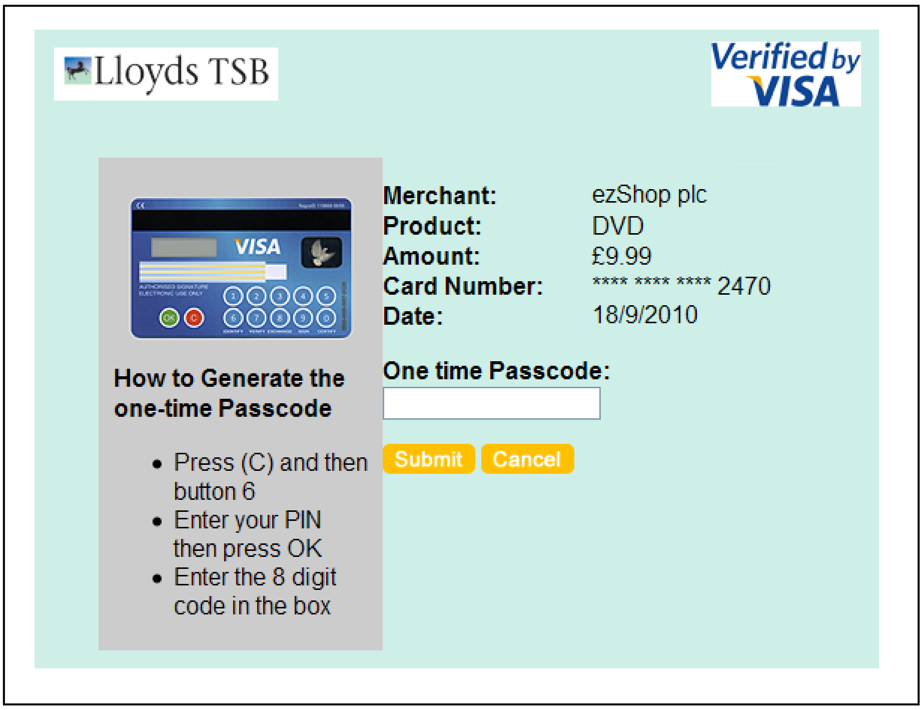
You need to enter your Verified by Visa password in order to complete the transaction.

Merchant: ezShop plc
Product: DVD
Amount: £9.99
Card Number: **** **** **** 1708
Date: 8/10/2010

Please enter digits **1**, **3** and **5** of your Verified by Visa Password:

□ □ □

Figure 4.1: Internet Shopping: Verified by Visa Password Challenge



How to Generate the one-time Passcode

- Press (C) and then button 6
- Enter your PIN then press OK
- Enter the 8 digit code in the box

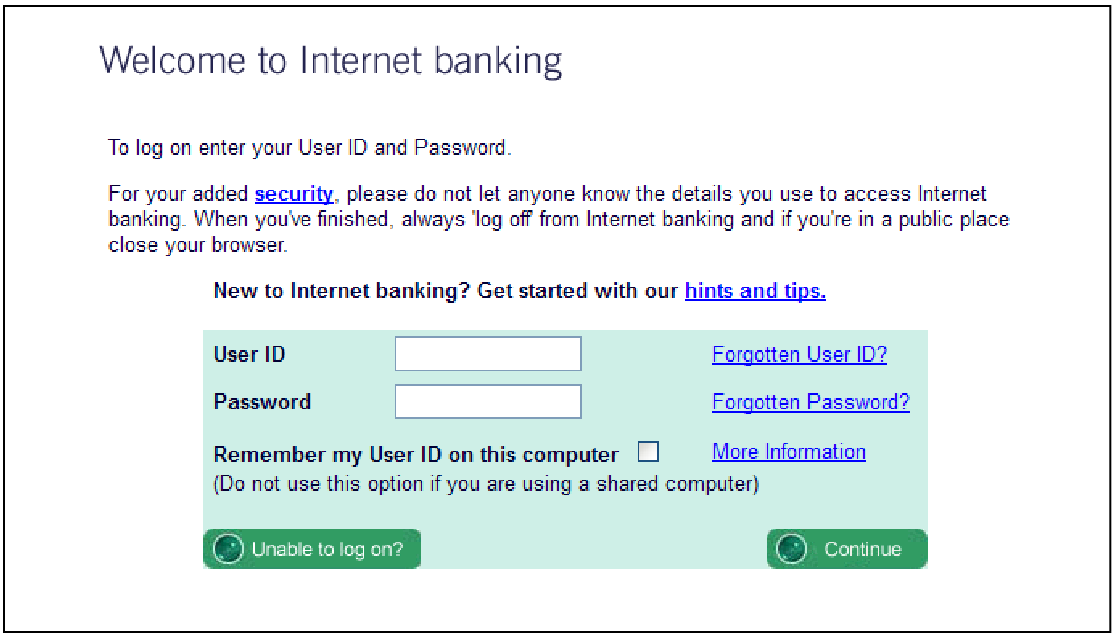
Merchant: ezShop plc
Product: DVD
Amount: £9.99
Card Number: **** **** **** 2470
Date: 18/9/2010

One time Passcode:

□

Figure 4.2: Internet Shopping: CodeSure OTP Challenge

The Internet banking site used in the experiment was a replica of the customer's bank's web site written in **ASP** running on a Microsoft IIS web server. To log in to the service the participant completed an authentication stage where they were asked to enter their customer User ID and password, Figure 4.3, followed by 3 random characters from previously registered memorable information (basically another type of password) using a pull-down list to avoid keyboard logging threats, Figure 4.4. For Internet banking both password and memorable information tokens have to be generated by the customer; between 6 and 15 characters including both letters and characters. With the security procedure based on the CodeSure card, the memorable information entry screen was replaced with a CodeSure OTP generation screen, Figure 4.5, asking the participant to use Identify mode on the card to generate an OTP.



Welcome to Internet banking

To log on enter your User ID and Password.

For your added security, please do not let anyone know the details you use to access Internet banking. When you've finished, always 'log off' from Internet banking and if you're in a public place close your browser.

New to Internet banking? Get started with our hints and tips.

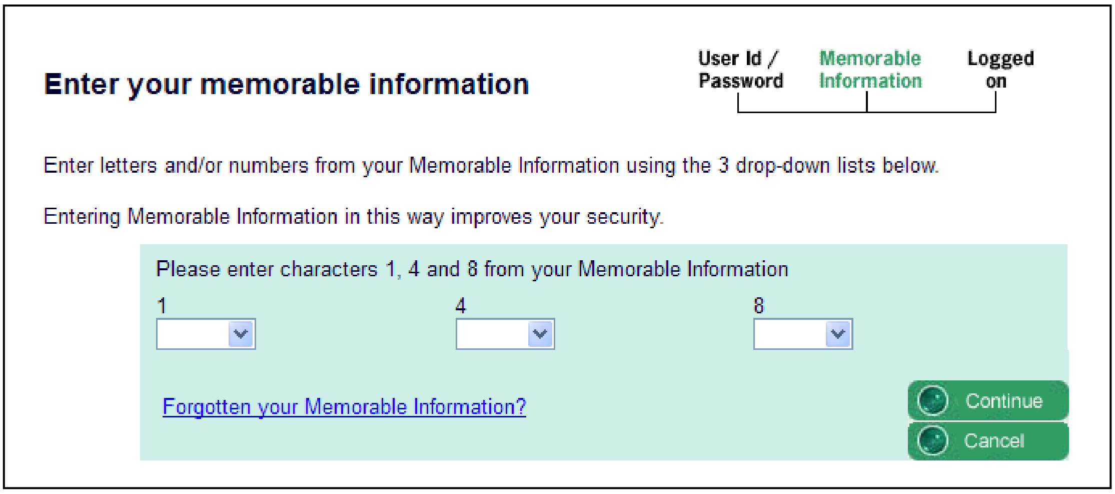
User ID Forgotten User ID?

Password Forgotten Password?

Remember my User ID on this computer □ More Information

(Do not use this option if you are using a shared computer)

Figure 4.3: Internet Banking Login: User ID and Password Challenge



Enter your memorable information

Enter letters and/or numbers from your Memorable Information using the 3 drop-down lists below.

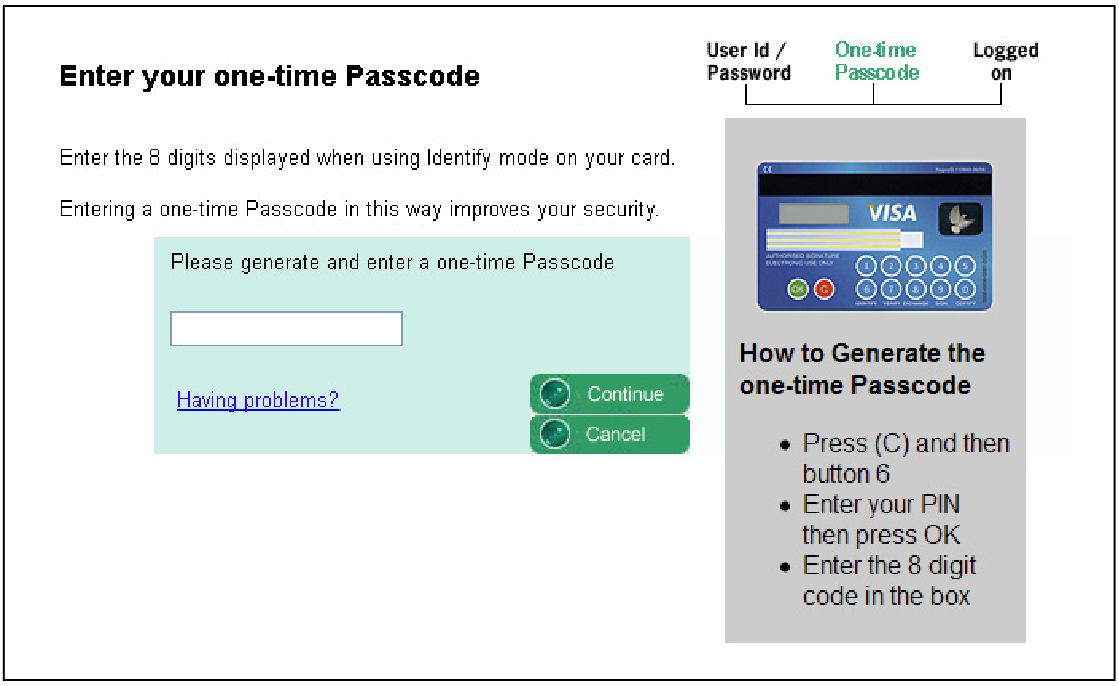
Entering Memorable Information in this way improves your security.

Please enter characters 1, 4 and 8 from your Memorable Information

1 4 8

Forgotten your Memorable Information?

Figure 4.4: Internet Banking Login: Memorable Information Challenge



Enter your one-time Passcode

Enter the 8 digits displayed when using Identify mode on your card.

Entering a one-time Passcode in this way improves your security.

Please generate and enter a one-time Passcode

How to Generate the one-time Passcode

- Press (C) and then button 6
- Enter your PIN then press OK
- Enter the 8 digit code in the box

Figure 4.5: Internet Banking Login: CodeSure OTP Challenge

In addition to the cognitive effort for the user to remember and input two unique passwords / memorable information secrets, the use of memorable information has a security weakness because most common choices of memorable information are easily obtainable by friends or family, or in identity theft. The increased use of memorable information across different online services also means that users are less likely to remember which memorable information is used by which service, causing errors, user frustration and limiting the usability and effectiveness of security procedures. The security procedure with the CodeSure card only requires the user to remember the existing everyday PIN for their card.

The IVR[3] telephone banking service used in the experiment was a replica of the customer's bank's automated interactive voice response (IVR) telephone banking service (Wilkie et al., 2005). This was written in **C** using Nuance speech recognition technology and the Nuance Dialog Builder API. To access the service the participants completed an authentication stage where they were asked to say or key in their account details (account number and sort code). This was followed by a

[3] Interactive Voice Response – utilises speech recognition technology to provide an automated telephone service.

security procedure based on saying or keying two random digits from a 'security number' (a type of password), Figure 4.6. The telephone banking security number has to be a fixed length of 6 digits. The security procedure based on the CodeSure card replaces the security number stage with a request for an 8 digit OTP generated with the Identify mode on the CodeSure card, Figure 4.7.

> **"Please say or key in your eight digit account number …**
>
> **and your sort code …**
>
> **Thanks …**
>
> **Now give the *m*th digit of your security number …**
>
> **and the *n*th digit."**

Figure 4.6: Telephone Banking Login: Security Number (Password) Challenge

> **"Please say or key in your eight digit account number …**
>
> **and your sort code …**
>
> **Thanks …**
>
> **Now I need information from your debit card. Do you have it with you? …**
>
> **Thank you …**
>
> **Now please use the identify mode on the reverse of your debit card and using your telephone keypad please key in the eight digit code displayed. For help on how to get your code, press hash."**

Figure 4.7: Telephone Banking Login: CodeSure OTP Challenge

The 6 digit security number used in telephone banking is initially assigned to users by the bank, sent through the post to their home address for security, and users are required to change that number to one of their choosing during their first use of the service. This introduces a certain amount of confusion in that it is an additional security number / password that users must remember; adding further to the confusion, the sort code for their account is also a 6 digit number; and the initial assignment of security number is reliant on being delivered through the post, which could be intercepted, representing a security weakness. The security procedure with the CodeSure card only requires the user to remember the existing everyday PIN for their card.

Participants were all already familiar with the set of channel-specific security procedures involved for each task. In the scenario with the common security procedure based on the CodeSure card, participants were presented with a User Guide for the CodeSure card, Figure B.6 in Appendix B.

After each of their six tasks, participants completed a usability questionnaire.

Participants were balanced by gender and age group: 'Generation Y' (ages 18 to 35), 'Generation X' (ages 36 to 55) and the 'Baby Boomers' (56 and over). These groups were balanced in the experiment design so that robust statistical analysis could be performed on the data collected to explore differences in attitudes and reactions between various groups.

Experiment sessions took approximately 90 minutes and were conducted in the experiment research suite in Edinburgh. Participation was rewarded with a £50 cheque. The details of the experiment are summarised in Table 4.1.

The CodeSure Multi-Channel Usability Experiment	
Experiment purpose	Exploration of customer attitude to the usability of a common security procedure based on the CodeSure card in multi-channel banking.
Null hypothesis (H_0)	There will be no differences in metrics of attitude to usability and preference between the two scenarios.
Experiment design	Participants experience a scenario with multiple disparate security procedures for different banking channels, and a scenario with common authentication based on the CodeSure card. Repeated measures (balanced order), within subjects.
Dependent variables	Usability questionnaire (7-point Likert scale).
Other data	Demographic and technographic data. Exit interview data.
Independent variables	Experiment: treatment order (2 orders, disparate security procedures or CodeSure, balanced). Participant: gender (2 genders, balanced), age group (3 groups, balanced, ages 18 to 35, 36 to 55, 56 and over).
Cohort	144 (3 age groups × 2 genders × 2 treatment orders × over-sampling ratio 12:1) = 144.
Honorarium	Personal cheque for £50.
Session time	90 minutes.

Table 4.1: Experiment Design Details

4.1.2. Summary of Experiment Metrics

Measurements collected included demographic and technographic characteristics of participants (e.g. age, gender); attitude toward usability, per-channel, per scenario; task completion data, error rates; interview comments on specific issues.

The design of the usability questionnaire for this experiment closely followed that of the previous CodeSure experiments, modified to reflect the focus on the overall process rather than just being about the device itself. Statements about the card display and buttons were replaced with statements covering confusion and the need for more instructions, the ease of use of the process and its perceived efficiency, and the volume of (security) details involved, Table 4.2.

As previously, the questionnaire was administered on a laptop with all statements presented in randomised order.

Questionnaire Statement (polarity)	Abbreviated Form
I had to concentrate hard to use this process (-)	Concentration
I felt flustered when using this process (-)	Flustered
I felt under stress when using this process (-)	Stressed
Using this process was very frustrating (-)	Frustration
Using this process was too complicated (-)	Complication
When using this process I didn't always know what to do next (-)	Knew what to do next
I felt in control when using this process (+)	In control
Using this process was quick (+)	Quick
I would be happy to use this process again (+)	Use again
I felt this process was reliable (+)	Reliable
I felt this process needs a lot of improvement (-)	Needs improvement
The operation of this process didn't match my expectations (-)	Matched expectations
I liked using this process (+)	Liked
I enjoyed using this process (+)	Enjoyment
Using this process felt trustworthy (+)	Trustworthy
Using this process felt secure (+)	Security
I found this process confusing to use (-)	Confusion
I would have liked more instructions on how to use this process (-)	Instructions
I felt this process was easy to use (+)	Ease of use
I thought this process was efficient (+)	Efficiency
I had to enter too many details during this process (-)	Too many details

Table 4.2: Usability Questionnaire Statements

4.1.3. Results

4.1.3.1. Demographic and Technographic Results

The final customer sample consisted of 165 Internet banking and Telephone banking customers in Edinburgh. A good balance for gender (47% male and 53% female) and for age groups (30% aged 18 to 35, 35% aged 36 to 55 and 35% aged 56 and over) was achieved in the final sample.

Some 66% of participants reported that they buy something online at least once a month; 60% reported that they log in to their bank's Internet banking site at least once a month; 36% reported that they phone their bank's automated telephone banking service at least once a month.

The socio-economic profile (based on occupation) of the sample was A(6%), B(41%), C1(35%), C2(15%), D(3%), E(0%), which can be judged as representing a reasonable cross-section of economic activity for the general population.

4.1.3.2. Usability Results

Online Shopping Task

The online shopping experience incorporating the CodeSure security challenge (on first time use) scored an acceptable 5.11 for overall usability. The more familiar procedure, based on a Verified by Visa password scored higher for usability at 5.61, Table 4.3. A repeated-measures analysis of variance (ANOVA) was carried out for the online shopping task, Table 4.4, with design (Verified by Visa vs. CodeSure card) as the within-subjects factor: and gender (78 male; 87 female), age group (50 aged 18 to 35; 57 aged 36 to 55; 58 aged 56 and over) and order of experience (83 Verified by Visa first; 82 CodeSure card first) as the between-subjects factors. Only main effects were explored.

Design	Mean Score	Std. Dev.	N
CodeSure Card	5.1100	1.12097	165
Verified by Visa	5.6072	0.80720	165

Table 4.3: Usability Questionnaire Means: Online Shopping

There was a very highly significant main effect for the design used, $F(1,153)=30.247$, $p<0.001$, confirming that the difference in usability scores between the online shopping experience based on the CodeSure card and the online shopping experience based on the Verified by Visa password was very highly significant.

Source	Sum of Squares	*df*	Mean Square	*F*	Sig.
Within-Subjects Effects					
Design	19.129	1	19.129	30.247	0.000
Design × Gender	0.933	1	0.933	1.475	0.226
Design × Age	3.127	2	1.564	2.473	0.088
Design × Order	0.587	1	0.587	0.928	0.337
Error (Design)	96.763	153	0.632		
Between-Subjects Effects					
Gender	0.005	1	0.005	0.004	0.947
Age	14.085	2	7.043	5.887	0.003
Order	1.027	1	1.027	0.858	0.356
Gender × Age	0.027	2	0.013	0.011	0.989
Gender × Order	0.047	1	0.047	0.040	0.843
Age × Order	4.109	2	2.054	1.717	0.183
Error	183.036	153	1.196		

Table 4.4: ANOVA Results for Usability: Online Shopping

There was a very highly significant main effect due to age group, *F*(2,153)=5.887, *p*=0.003, but no significant main effects due to gender or order of experience, and also no significant between-subjects interactions.

The online shopping experience based on the CodeSure card was scored significantly lower than the online shopping experience based on the Verified by Visa password for 18 of the 21 usability attributes, Table 4.5. However, the online shopping experience based on the CodeSure card was rated significantly higher for security, *F*(1,153)=5.836, *p*=0.017, than the online shopping experience based on the Verified by Visa password.

There were 11 attributes where age group had a significant effect: those aged 18 to 35 gave significantly higher usability scores overall than those aged 36 to 55, and this was the same for those aged 36 to 55 compared to those who were over 55.

Attribute	Significant Effects	Details
Confusion	Design (*df*=1; *F*=16.923; *p*<0.001)	CodeSure < Verified by Visa
	Age (*df*=2; *F*=8.552; *p*<0.001)	Age 18 to 35 > 36 to 55 > 56+
Concentration	Design (*df*=1; *F*=47.991; *p*<0.001)	CodeSure < Verified by Visa
	Age (*df*=2; *F*=12.073; *p*<0.001)	Age 18 to 35 > 36 to 55 > 56+
Flustered	Design (*df*=1; *F*=23.344; *p*<0.001)	CodeSure < Verified by Visa
	Age (*df*=2; *F*=7.746; *p*=0.001)	Age 18 to 35 > 36 to 55 > 56+
Stressed	Design (*df*=1; *F*=22.739; *p*<0.001)	CodeSure < Verified by Visa
	Age (*df*=2; *F*=7.149; *p*=0.001)	Age 18 to 35 > 36 to 55 > 56+
Frustration	Design (*df*=1; *F*=40.371; *p*<0.001)	CodeSure < Verified by Visa
	Age (*df*=2; *F*=3.625; *p*=0.029)	Age 18 to 35 > 36 to 55 > 56+
Complication	Design (*df*=1; *F*=30.336; *p*<0.001)	CodeSure < Verified by Visa
	Age (*df*=2; *F*=6.731; *p*=0.002)	Age 18 to 35 > 36 to 55 > 56+
Knew what to do next	Design (*df*=1; *F*=12.897; *p*<0.001)	CodeSure < Verified by Visa
	Age (*df*=2; *F*=6.354; *p*=0.002)	Age 18 to 35 > 36 to 55 > 56+
	Order (*df*=1; *F*=4.175; *p*=0.043)	CodeSure first, both higher
Instructions	Design (*df*=1; *F*=8.218; *p*=0.005)	CodeSure < Verified by Visa
	Age (*df*=2; *F*=8.953; *p*<0.001)	Age 18 to 35 > 36 to 55 > 56+
In control	Design (*df*=1; *F*=8.927; *p*=0.003)	CodeSure < Verified by Visa
	Age (*df*=2; *F*=4.402; *p*=0.014)	Age 18 to 35 > 36 to 55 > 56+
Quick	Design (*df*=1; *F*=64.472; *p*<0.001)	CodeSure < Verified by Visa
Ease of use	Design (*df*=1; *F*=36.857; *p*<0.001)	CodeSure < Verified by Visa
	Age (*df*=2; *F*=5.915; *p*=0.003)	Age 18 to 35 > 36 to 55 > 56+
Too many details	Design (*df*=1; *F*=17.312; *p*<0.001)	CodeSure < Verified by Visa
	Age (*df*=2; *F*=4.075; *p*=0.019)	Age 18 to 35 > 36 to 55 > 56+
Use again	Design (*df*=1; *F*=5.273; *p*=0.023)	CodeSure < Verified by Visa
Efficiency	Design (*df*=1; *F*=20.035; *p*<0.001)	CodeSure < Verified by Visa
Needs improvement	Design (*df*=1; *F*=40.352; *p*<0.001)	CodeSure < Verified by Visa
Matched expectations	Design (*df*=1; *F*=6.709; *p*=0.011)	CodeSure < Verified by Visa
Liked	Design (*df*=1; *F*=7.585; *p*=0.007)	CodeSure < Verified by Visa
Enjoyment	Design (*df*=1; *F*=6.133; *p*=0.014)	CodeSure < Verified by Visa
Trustworthy	Gender (*df*=1; *F*=4.775; *p*=0.030)	Males > Females
Security	Design (*df*=1; *F*=5.836; *p*=0.017)	CodeSure > Verified by Visa

Table 4.5: Significant Differences in Usability Attributes: Online Shopping

Internet Banking Task

The Internet banking experience incorporating the CodeSure security challenge (on first time use) scored an acceptable 5.29 for overall usability. The more familiar experience, based on Internet banking password and memorable information scored higher for usability at 5.47, Table 4.6. A repeated-measures analysis of variance (ANOVA) was carried out on the data for the Internet banking task, Table 4.7, with design (Memorable Information design vs. CodeSure card challenge) as the within-subjects factor and gender (78 male; 87 female), age group (50 aged 18 to 35; 57 aged 36 to 55; 58 aged 56 and over) and order of experience (83 experienced the scenario with memorable information challenge first; 82 experienced the scenario with CodeSure card first) as the between-subjects factors. Only main effects were explored.

Design	Mean Score	Std. Dev.	N
CodeSure Card	5.2903	1.07619	165
Memorable Information	5.4678	1.00716	165

Table 4.6: Usability Questionnaire Means: Internet Banking

There was a significant main effect for the design used, *F*(1,153)=4.193, *p*=0.042. The mean attitude score for usability for the overall experience of Internet banking with the CodeSure card challenge at 5.29, whilst acceptable high in its own right, was lower than the mean attitude score for usability for the overall experience of Internet banking using the memorable information challenge at 5.47. The difference was statistically significant, *F*(1,153)=4.193, *p*=0.042.

Source	Sum of Squares	*df*	Mean Square	*F*	Sig.
Within-Subjects Effects					
Design	2.529	1	2.529	4.193	0.042
Design × Gender	0.181	1	0.181	0.300	0.585
Design × Age	0.421	2	0.211	0.349	0.706
Design × Order	0.008	1	0.008	0.014	0.907
Error (Design)	92.264	153	0.603		
Between-Subjects Effects					
Gender	1.848	1	1.848	1.339	0.249
Age	31.313	2	15.657	11.348	0.000
Order	0.001	1	0.001	0.000	0.983
Gender × Age	0.369	2	0.184	0.134	0.875
Gender × Order	0.802	1	0.802	0.581	0.447
Age × Order	5.055	2	2.528	1.832	0.164
Error	211.082	153	1.380		

Table 4.7: ANOVA Results for Usability: Internet Banking

There was a very highly significant main effect due to age group, $F(2,153)=11.348$, $p<0.001$, but no significant main effects due to gender or order of experience, and also no significant between-subjects interactions.

The overall Internet banking experience based on the CodeSure card was rated significantly lower for usability than the overall Internet banking experience using memorable information for 11 of the 21 usability attributes, Table 4.8. There were 12 attributes where age group had a significant effect: those aged 18 to 35 gave significantly higher usability scores overall than those aged 36 to 55, and this was the same for those aged 36 to 55 compared to those who were aged 56 and over.

Attribute	Significant Effects	Details
Confusion	Age (*df*=2; *F*=16.179; *p*<0.001)	Age 18 to 35 > 36 to 55 > 56+
Concentration	Design (*df*=1; *F*=6.397; *p*=0.012)	CodeSure < Memorable Info
	Age (*df*=2; *F*=22.252; *p*<0.001)	Age 18 to 35 > 36 to 55 > 56+
Flustered	Design (*df*=1; *F*=11.089; *p*=0.001)	CodeSure < Memorable Info
	Age (*df*=2; *F*=13.082; *p*<0.001)	Age 18 to 35 > 36 to 55 > 56+
Stressed	Design (*df*=1; *F*=10.123; *p*=0.002)	CodeSure < Memorable Info
	Age (*df*=2; *F*=14.312; *p*<0.001)	Age 18 to 35 > 36 to 55 > 56+
Frustration	Design (*df*=1; *F*=7.430; *p*=0.007)	CodeSure < Memorable Info
	Age (*df*=2; *F*=7.674; *p*=0.001)	Age 18 to 35 > 36 to 55 > 56+
Complication	Age (*df*=2; *F*=8.535; *p*<0.001)	Age 18 to 35 > 36 to 55 > 56+
Knew what to do next	Age (*df*=2; *F*=19.903; *p*<0.001)	Age 18 to 35 > 36 to 55 > 56+
Instructions	Design (*df*=1; *F*=7.120; *p*=0.008)	CodeSure < Memorable Info
	Age (*df*=2; *F*=18.933; *p*<0.001)	Age 18 to 35 > 36 to 55 > 56+
In control	Design (*df*=1; *F*=5.943; *p*=0.016)	CodeSure < Memorable Info
	Age (*df*=2; *F*=11.178; *p*<0.001)	Age 18 to 35 > 36 to 55 > 56+
Quick	Design (*df*=1; *F*=3.985; *p*=0.048)	CodeSure < Memorable Info
Ease of use	Design (*df*=1; *F*=3.915; *p*=0.050)	CodeSure < Memorable Info
	Age (*df*=2; *F*=8.022; *p*<0.001)	Age 18 to 35 > 36 to 55 > 56+
Too many details	Age (*df*=2; *F*=5.959; *p*=0.003)	Age 18 to 35 > 36 to 55 > 56+
	Order (*df*=1; *F*=4.389; *p*=0.038)	CodeSure first, both lower
Use again	Design (*df*=1; *F*=5.727; *p*=0.018)	CodeSure < Memorable Info
Matched expectations	Design (*df*=1; *F*=5.726; *p*=0.018)	CodeSure < Memorable Info
	Age (*df*=2; *F*=4.421; *p*=0.014)	Age 18 to 35 > 36 to 55 > 56+
Trustworthy	Design (*df*=1; *F*=4.194; *p*=0.042)	CodeSure < Memorable Info

Table 4.8: Significant Differences in Usability Attributes: Internet Banking

Telephone Banking Task

The Telephone banking experience using the CodeSure card (on first time use) scored an acceptable 5.01 for overall usability. The more familiar Telephone banking experience based on security number password scored higher for usability at 5.51, Table 4.9. A repeated-measures analysis of variance (ANOVA) was carried out on the scores for the Telephone banking task, Table 4.10, with design (security number vs. CodeSure card) as the within-subjects factor and gender (78 male; 87 female), age group (50 aged 18 to 35; 57 aged 36 to 55; 58 aged 56 and over) and order of experience (83 security number experience first; 82 CodeSure card experience first) as the between-subjects factors. Only main effects were explored.

Design	Mean Score	Std. Dev.	N
CodeSure Card	5.0139	1.13579	165
Security Number	5.5102	0.97030	165

Table 4.9: Usability Questionnaire Means: Telephone Banking

Source	Sum of Squares	*df*	Mean Square	*F*	Sig.
Within-Subjects Effects					
Design	20.615	1	20.615	29.093	0.000
Design × Gender	1.012	1	1.012	1.428	0.234
Design × Age	2.080	2	1.040	1.468	0.234
Design × Order	3.849	1	3.849	5.431	0.021
Error (Design)	108.413	153	0.709		
Between-Subjects Effects					
Gender	0.256	1	0.256	0.169	0.682
Age	1.978	2	0.989	0.652	0.523
Order	0.770	1	0.770	0.508	0.477
Gender × Age	1.385	2	0.692	0.456	0.635
Gender × Order	2.407	1	2.407	1.586	0.210
Age × Order	1.715	2	0.857	0.565	0.570
Error	232.216	153	1.518		

Table 4.10: ANOVA Results for Usability: Telephone Banking

There was a very highly significant main effect for the design used, $F(1,153)=29.093$, $p<0.001$. The mean attitude score for usability for the overall experience of Telephone banking with the CodeSure card challenge at 5.01, whilst acceptably high in its own right, was lower than the mean attitude score for usability for the overall experience of Telephone banking using the security number challenge at 5.51.

There were no significant main effects due to gender, age group or order of experience.

The usability of the Telephone banking experience using the CodeSure card was rated significantly lower than the usability of the Telephone banking experience using security number for 18 of the 21 usability attributes, Table 4.11. The three attributes where this was not the case were 'reliable', 'trustworthy' and 'security'.

Attribute	Significant Effects	Details
Confusion	Design (*df*=1; *F*=17.479; *p*<0.001)	CodeSure < Security No.
Concentration	Design (*df*=1; *F*=25.066; *p*<0.001)	CodeSure < Security No.
Flustered	Design (*df*=1; *F*=31.578; *p*<0.001)	CodeSure < Security No.
Stressed	Design (*df*=1; *F*=31.360; *p*<0.001)	CodeSure < Security No.
Frustration	Design (*df*=1; *F*=20.034; *p*<0.001)	CodeSure < Security No.
Complication	Design (*df*=1; *F*=23.142; *p*<0.001)	CodeSure < Security No.
Knew what to do next	Design (*df*=1; *F*=19.443; *p*<0.001)	CodeSure < Security No.
Instructions	Design (*df*=1; *F*=25.735; *p*<0.001)	CodeSure < Security No.
In control	Design (*df*=1; *F*=27.729; *p*<0.001)	CodeSure < Security No.
Quick	Design (*df*=1; *F*=19.381; *p*<0.001)	CodeSure < Security No.
Ease of use	Design (*df*=1; *F*=28.439; *p*<0.001)	CodeSure < Security No.
Too many details	Design (*df*=1; *F*=33.019; *p*<0.001)	CodeSure < Security No.
Use again	Design (*df*=1; *F*=6.740; *p*=0.010)	CodeSure < Security No.
Efficiency	Design (*df*=1; *F*=4.477; *p*=0.036)	CodeSure < Security No.
Needs improvement	Design (*df*=1; *F*=17.648; *p*<0.001)	CodeSure < Security No.
Matched expectations	Design (*df*=1; *F*=16.470; *p*<0.001)	CodeSure < Security No.
Liked	Design (*df*=1; *F*=8.460; *p*=0.004)	CodeSure < Security No.
Enjoyment	Design (*df*=1; *F*=8.313; *p*=0.005)	CodeSure < Security No.

Table 4.11: Significant Differences in Usability Attributes: Telephone Banking

In summary, usability scores with the experiences based on disparate, channel-specific security procedures were higher than the experiences based on a common, multi-channel security procedure based on the CodeSure card. This was the case for all three tasks examined. The differences in usability scores overall were statistically significant.

For the six usability questionnaires posed, reliability analysis on the scale of each (based on Cronbach's α) reported a minimum of α=0.937, indicating once again that the questionnaire reliability was good.

4.1.3.3. Quality Rating Results

Participants were asked to rate use of the common security procedure based on the CodeSure card with use of the disparate channel-specific security procedures, between best and worst on a scale of 0 to 30, both overall and for the three attributes, ‘convenience’, ‘security’ and ‘ease of use’, Table 4.12. This was done after all tasks had been completed.

Attribute	Disparate	CodeSure
Overall	21.70	20.35
Convenience	22.95	18.75
Security	20.55	24.18
Ease of use	24.27	18.42

Table 4.12: Mean Quality Rating Results

The CodeSure card was rated higher for ‘security’ but lower for ‘convenience’ and ‘ease of use’.

Participants gave several reasons for their overall ratings: “*The CodeSure card was a bit confusing to use. I feel that I wasn’t as familiar with it*”, “*The debit card was easiest to use, but once you know what you’re doing with the CodeSure card you will have more security*.”

In terms of convenience, most participants recognised the benefits of the CodeSure card but were concerned with the usability of the buttons, and therefore confused convenience with ease of use: “*The current system is so straightforward and the card is a bit harder with the buttons*.”

In terms of security, participants appreciated the one-time passcode generation feature of the CodeSure card: “*The CodeSure card seems a lot stronger for security, randomly generated code more secure than password as it keeps changing*”, “*CodeSure card is better due to random numbers which keep changing*.”

For ease of use, the focus was again on the buttons of the CodeSure card: “*The current systems were easier and quicker*”, “*The existing approaches were familiar. I know what I’m doing. The new card’s buttons are unresponsive*.”

A repeated-measures analysis of variance (ANOVA) was carried out on these data, Table 4.13, with design (disparate vs. CodeSure card) as the within-subjects factor and gender, age group and order of experience as the between-subjects factors. As before only main effects were explored.

Attribute	Significant Effects	Details
Overall	Age (*df*=2; *F*=4.847; *p*=0.009)	Age 18 to 35 > 36 to 55 > 56+
Convenience	Design (*df*=1; *F*=29.021; *p*<0.001)	Disparate > CodeSure
Security	Design (*df*=1; *F*=29.874; *p*<0.001)	CodeSure > Disparate
Ease of Use	Design (*df*=1; *F*=60.018; *p*<0.001)	Disparate > CodeSure
	Age (*df*=2; *F*=3.741; *p*=0.026)	Age 18 to 35 > 36 to 55 > 56+

Table 4.13: Summary of Significant Differences in Quality Attributes

The difference in scores between the overall experience with channel-specific security procedures and the overall experience with a common CodeSure security procedure was not statistically significant. The score for security was higher with the common CodeSure card procedure than with the disparate security procedures and the difference was statistically significant, *F*(1,153)=29.874, *p*<0.001. The scores for convenience and ease of use were lower with the common CodeSure procedure than with the disparate security procedures and the differences were statistically significant for convenience, *F*(1,153)=29.021, *p*<0.001, and for ease of use, *F*(1,153)=60.018, *p*<0.001. There was also a between-subjects effect of age group for ease of use, *F*(2,153)=3.741, *p*=0.026, and overall, *F*(2,153)=4.847, *p*=0.009.

Creating a rank ordered list from the overall quality ratings gives participants' overall preferences, Table 4.14. As relative score differences can vary greatly, this can give a better indication of what participants' true preferences were, and can even differ from the mean quality ratings if there are clear leaders in the rankings. The disparate security procedures and the common CodeSure procedure were ranked similar overall, and the rankings matched those of the quality ratings for convenience, security and ease of use.

	Disparate	CodeSure	No Preference
Overall	48.48%	51.52%	0.00%
Convenience	61.82%	28.48%	9.70%
Security	15.76%	69.09%	14.55%
Ease of Use	74.55%	15.15%	10.30%

Table 4.14: Preferences between Disparate and Common CodeSure Procedures

A non-parametric statistical test (chi-square test) was performed on these categorical preference data in Table 4.14. Participants who ranked both devices equally were marked as no preference and these cases were removed from the subsequent chi-square analysis.

The difference in rankings between the disparate procedures and the common CodeSure procedure was not statistically significant for the overall rating, but the differences were significant for convenience, $\chi^2(1)=20.302$, $p<0.001$, security, $\chi^2(1)=56.177$, $p<0.001$, and ease of use, $\chi^2(1)=64.892$, $p<0.001$.

4.1.3.4. Success Rates

For both the online shopping task and the Internet banking task, the common CodeSure procedure had a higher success rate than with the disparate security procedures, Table 4.15.

Online shopping		Internet banking		Telephone banking	
Disparate	**CodeSure**	**Disparate**	**CodeSure**	**Disparate**	**CodeSure**
57.8%	69.9%	69.3%	77.1%	82.5%	77.7%

Table 4.15: Success Rates: Access First Attempt

The higher success rate with the CodeSure card was statistically significant (*p*=0.010) for the online shopping task, but not for the Internet banking task (*p*=0.110) (McNemar's test).

In the Telephone banking task the CodeSure card had a lower success rate but the difference was not statistically significant (*p*=0.302). The extra failures here were typically due to users being unable to complete generation of the one-time passcode on the CodeSure card and keying in to the phone within the 30 second limit imposed by the IVR system; or users having problems pressing the buttons on the CodeSure card using only one hand.

4.1.3.5. Interview Comments

Asked what they liked about using the CodeSure card, participants responded: "*The display was very clear and easy to see. It is also a lot more secure*." Their dislikes were mainly about the buttons: "*Pressing the buttons can be difficult at the start*."

Asked about the benefits to them of using the CodeSure card, participants responded: "*It's fast and simple to use and it means I will not have to write down or memorise my numbers*", "*It is very innovative and modern, security conscious, good impression on the company.*"

Some 66% of participants stated they preferred use of a common security procedure for all channels while 28% said they preferred the disparate security procedures currently used by the bank.

Some 79% of participants would like to use the CodeSure card for online shopping, with 24% saying they would shop more often than at present; 65% of participants would like to use the CodeSure card to access their bank's Internet banking service, with 36% saying they would use Internet banking more often than at present; 57% of participants would like to use the CodeSure card to access their bank's telephone banking service, with 28% saying they would use telephone banking more often than at present.

4.1.4. Discussion on Experiment Results

The null hypothesis (H_0) for this experiment was:

> *"There will be no differences in metrics of attitude to usability and preference between the two scenarios."*

The evidence presented here is sufficient to refute this null hypothesis since usability scores for the online shopping, Internet banking and telephone banking tasks with the existing disparate channel-specific security procedures were higher than those with the common CodeSure card security procedure; and statistical tests confirm the differences to be significant. The difference in quality ratings and preferences for convenience, security and ease of use were also statistically significant with the CodeSure card procedures scoring higher in terms of security but scored lower for convenience and ease of use.

For online shopping, the mean usability score at 5.11 with the common CodeSure card procedure, whilst acceptably high in its own right, was significantly lower than with the existing Verified by Visa procedure at 5.61, and a similar finding was obtained for 18 of the 21 usability attributes examined. However, the CodeSure card procedure was rated significantly higher than the Verified by Visa procedure for the security attribute and had a significantly higher success rate (70%) than did the Verified by Visa procedure (58%) during the task. Many of the usability attributes that were rated significantly lower for the CodeSure card procedure were related to the lack of familiarity with the CodeSure procedure compared to the existing Verified by Visa procedure, which could explain why users felt so confused and flustered when first confronted with the new CodeSure-based procedure.

For Internet banking, the mean usability score at 5.29 with the CodeSure card for login was again acceptably high in its own right but was lower than with the existing memorable information security procedure at 5.47, and this difference was statistically significant. There were 11 out of 21 usability attributes where the CodeSure card procedure was rated significantly lower than the memorable information procedure, again many likely to be related to familiarity of the existing process. The CodeSure card had a higher success rate than the memorable information procedure but that difference was not statistically significant.

For telephone banking, the mean usability score at 5.01 with the CodeSure card for login was acceptably high in its own right, but was significantly lower than with the security number procedure at 5.51. The CodeSure card procedure was rated significantly lower than the security number procedure for 18 of the 21 usability attributes. The CodeSure card procedure also had a lower success rate than the security number procedure although this difference was not statistically significant. The CodeSure card proved to be difficult to use whilst holding a telephone handset from an ergonomics point of view – it was difficult to operate the card and hold the phone to the ear at the same time.

Although the disparate security procedures were scored significantly higher than the CodeSure card procedure for usability, the mean usability scores for the CodeSure card were still acceptably high, being scored over 5.0 for all 3 tasks. All of the participants were first time CodeSure card users and were therefore unfamiliar using the CodeSure card to perform the tasks they routinely do several times a month (66% of participants shopped online at least once a month, 60% logged in to Internet banking at least once a month and 36% phoned their automated telephone banking service at least once a month). Usability scores for the CodeSure card have been shown in Chapter 3 to increase with usage which would perhaps improve scores with the CodeSure card.

There appears to be some enthusiasm for the CodeSure card and the unified security procedure it offers: some 66% of participants would prefer a common CodeSure for all channels compared to the 28% who prefer disparate security procedures. Some 79% would like to use the CodeSure card for online shopping, 24% would shop more often; 65% would like to use the card for Internet banking, 36% would use Internet banking more often; and 57% would like to use the card for telephone banking, 28% would use it more often.

4.2. Summary

The experiment results presented here indicate that users are aware of a trade-off between the security of a procedure and its convenience. This is reflected especially in the usability results for the security attribute between Verified by Visa and the CodeSure card during the online shopping task.

Some 66% of participants would prefer a common security procedure across all banking and eCommerce channels, which is encouraging for the future of the CodeSure card. Only 28% would prefer the security to remain the same in the banking channels they use, but that is perhaps due to being familiar with existing procedures rather than thinking about the benefits a common security procedure would bring. Although the CodeSure card scored lower than today's security processes in terms of usability, the CodeSure card was still usable with mean usability scores above 5.0, and the main challenge for the bank would be to break users out of their comfort zone in using security processes that they are familiar with and move them over to the common security processes of the CodeSure card.

The results presented in this chapter confirm the role of the CodeSure card in a common security procedure across multiple channels, offering a usable solution in reducing fraud by means of improved customer authentication performance. The other element of fraud reduction, using the CodeSure card to combat identity theft, phishing attacks, malware and reverse (sender) authentication is addressed in the next chapter.

Chapter 5. Combating Identity Theft – the Role of the CodeSure Card in Sender Authentication

The search for a common, usable security procedure involves improved customer authentication on the one hand and improved protection against identify theft on the other. The results presented thus far have illustrated the role of the CodeSure card in a common security procedure offering strong two-factor customer authentication across multiple channels. The other element of fraud reduction, using the CodeSure card to combat identity theft from phishing attacks and malware and by use of reverse (sender) authentication is addressed here.

The increase in public awareness of online fraud has resulted in growing customer mistrust of online channels. As a result, banks have made a strategic decision to avoid use of the email channel with customers; instead using messages accessible only after the customer has logged in (securely) to access their accounts on the bank's Internet banking site. Indeed banks have been instructing their customers to ignore any other emails purporting to be from the bank as they are likely to be fraudulent. Customers are also at risk from malware threats that can infect their computer (or phone) and engage in identity theft without their knowledge. In an ideal situation, banks would like to ensure that all of their customers are sufficiently protected from such malware threats, and one solution is for the bank to provide malware detection software to their customers to protect against this threat.

Four experiments are reported in this chapter, three of which investigated the awareness of customers about identity theft of their personal details via phishing attacks and via malware; and one experiment which investigated the role of the CodeSure card in fraud prevention by allowing the bank as sender to authenticate themself in communications with the customer (emails, SMS text messages and phone calls) using reverse (sender) authentication.

The first experiment dealt with phishing emails, focusing on customers' understanding and awareness of the issues with phishing emails and the consequent fraud, characterising levels of customer awareness and the impact of social engineering schemes: and examined customers' abilities to actually notice when an email is a phishing email. A representative bundle of different types of phishing email attacks were explored in the experiment.

The second experiment explored the effectiveness of Internet Security tutorials (which in practice would be accessible to customers on the bank's web site) in educating customers on best practice in detecting and handling phishing emails.

The third experiment dealt with malware, investigating customer reactions to a possible malware fraud prevention strategy that involves the bank recommending that the customer download malware detection software (which is provided and certified by the bank) to be run on their computer. The idea with this strategy is that every time the customer logs on to the bank's Internet banking site, the malware detection software would send a report to the bank about any infection or malware presence on the customer's computer by scanning the applications and processes that are running on the customer's computer. This fraud prevention strategy would allow the bank to make a risk assessment for each transaction on each Internet banking session with the possibility of advising the customer of the existence of a malware threat, recommending that they disinfect their computer using anti-viral software, or requiring the customer to make an out-of-band contact with the bank to complete any specific (risky) transactions.

The fourth experiment investigated customer reactions to the reverse / sender authentication capabilities of the CodeSure card as used in a fraud prevention strategy by the bank to reassure customers that the communications (in all forms) that they receive from the bank are genuine.

5.1. The Phishing Emails Experiment

The key questions addressed by this experiment centre on the levels of customers' understanding about the substance of phishing emails and the financial consequences for them as Internet banking users; and on user behaviour on receipt of phishing emails in terms of their ability to detect phishing emails and their propensity to divulge security details. The experiment was designed to explore if different types of phishing email attacks would affect user behaviour.

5.1.1. Experiment Design

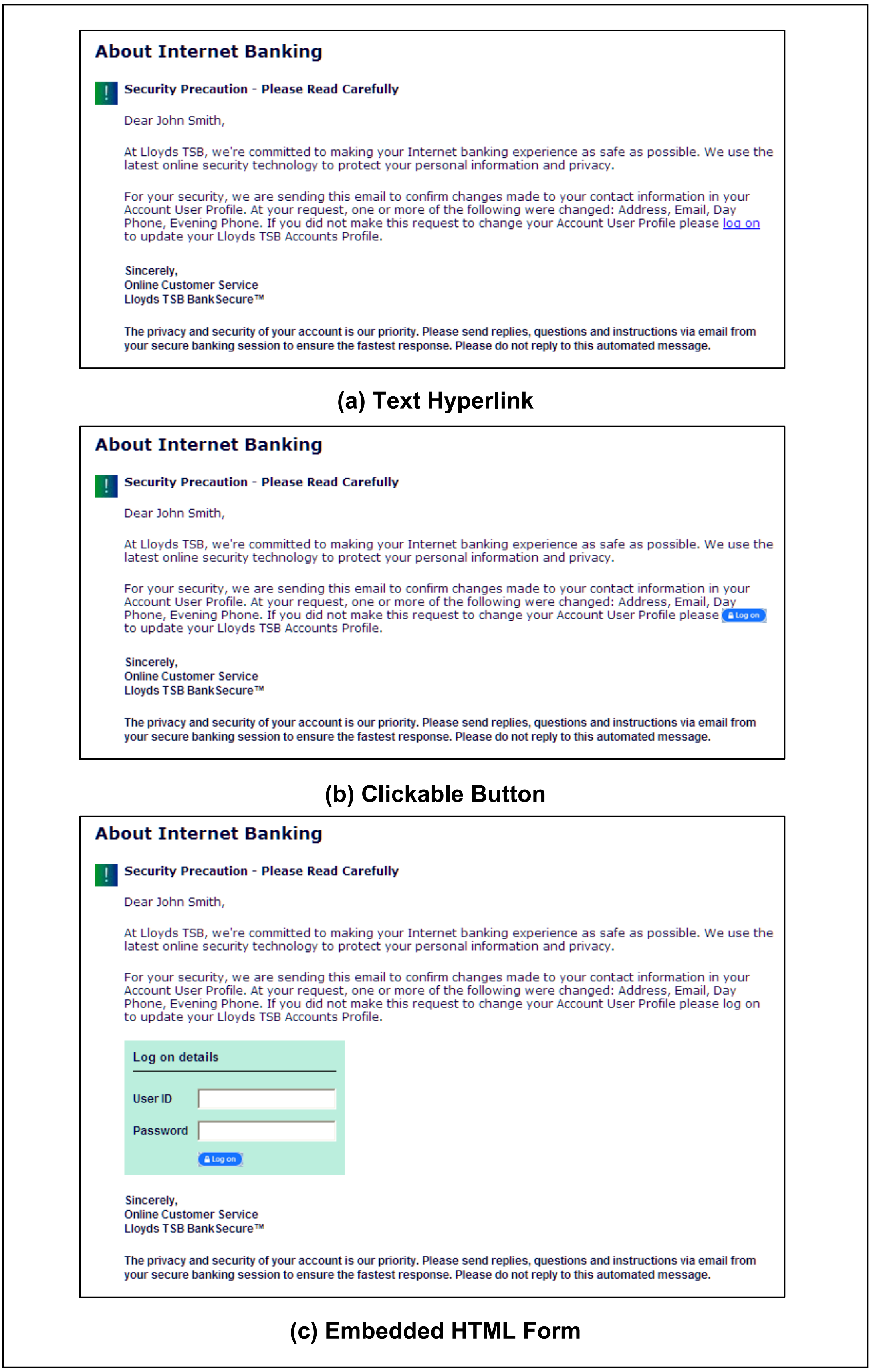
For this experiment participants were selected as having used their bank's Internet banking at some point during the 3 months prior to the experiment. In their experiment session, participants were asked to assume a persona and work through their email in-box list of 6 emails and to "*deal with the emails as you would in real life*". The 6 emails were pre-defined in a batch of 3 benign email messages and 3 phishing emails. The fraudulent, phishing emails presented to the participants were representative of the most sophisticated phishing email types reported to date, including addressing the participant by 'their' name and containing no spelling or grammatical mistakes.

The 3 phishing emails exhibited different styles in terms of modes of responding – asking the reader to click on a text hyperlink; to click on a clickable button; or to complete an embedded HTML form, Figure 5.1. With the text hyperlink style of phishing link, the links to the fraudulent Internet banking site were plain text hyperlinks such that moving the mouse over the hyperlinked word '`log on`' would display the target link, which was spoofed using the digit '`1`' as replacement for one of the lowercase '`l`'s in the target URL. With the clickable button style of phishing link, the clickable buttons acted as hyperlinks and as such represented a more determined phishing attack since moving the mouse over the button did not display the target link in this case. In all three cases, if the participant entered ‘their’ account details they were directed to a replica of their bank’s Internet banking site (at the fraudulent URL) which simulated a browser overlay frame spoof with fake padlock symbol.

The 3 phishing emails also exhibited different levels of apparent content urgency and pressure for the reader to attend and respond, ranging from confirming account changes with subject “Internet Banking: Security Precaution”; increasing urgency in the need to arrange payment of a forthcoming utility bill with subject “Bill Payment Reminder Service”; and further increasing urgency to a warning of an imminent overdraft charge with subject “Important Notice: Overdraft Charge”, Figure 5.2.

The 3 benign emails presented to the participant were a legitimate email from their bank suggesting that they might benefit from a new Internet Saver Account, using their (persona) name (with subject “New Internet Saver Account”); a ‘spam’ email which reports that a little-known company is worth investing in (with subject “Red Hot Stock Watch”); and a benign ‘newsletter’ style email from a supermarket store (with subject “Even More Inflation Busting Deals”), Figure 5.3.

When participants dealt with each of their 3 phishing emails and the one legitimate email from their bank, a realistic (mirror) copy version of the bank’s Internet banking web site was accessed such that participants were able to complete the task in hand (with the exception of the bill payment task), even when responding to the fraudulent emails, so as not to alert them and influence their behaviour with subsequent emails.



About Internet Banking

Security Precaution - Please Read Carefully

Dear John Smith,

At Lloyds TSB, we're committed to making your Internet banking experience as safe as possible. We use the latest online security technology to protect your personal information and privacy.

For your security, we are sending this email to confirm changes made to your contact information in your Account User Profile. At your request, one or more of the following were changed: Address, Email, Day Phone, Evening Phone. If you did not make this request to change your Account User Profile please log on to update your Lloyds TSB Accounts Profile.

Sincerely,
Online Customer Service
Lloyds TSB BankSecure™

The privacy and security of your account is our priority. Please send replies, questions and instructions via email from your secure banking session to ensure the fastest response. Please do not reply to this automated message.

(a) Text Hyperlink

About Internet Banking

Security Precaution - Please Read Carefully

Dear John Smith,

At Lloyds TSB, we're committed to making your Internet banking experience as safe as possible. We use the latest online security technology to protect your personal information and privacy.

For your security, we are sending this email to confirm changes made to your contact information in your Account User Profile. At your request, one or more of the following were changed: Address, Email, Day Phone, Evening Phone. If you did not make this request to change your Account User Profile please Log on to update your Lloyds TSB Accounts Profile.

Sincerely,
Online Customer Service
Lloyds TSB BankSecure™

The privacy and security of your account is our priority. Please send replies, questions and instructions via email from your secure banking session to ensure the fastest response. Please do not reply to this automated message.

(b) Clickable Button

About Internet Banking

Security Precaution - Please Read Carefully

Dear John Smith,

At Lloyds TSB, we're committed to making your Internet banking experience as safe as possible. We use the latest online security technology to protect your personal information and privacy.

For your security, we are sending this email to confirm changes made to your contact information in your Account User Profile. At your request, one or more of the following were changed: Address, Email, Day Phone, Evening Phone. If you did not make this request to change your Account User Profile please log on to update your Lloyds TSB Accounts Profile.

Log on details

User ID

Password

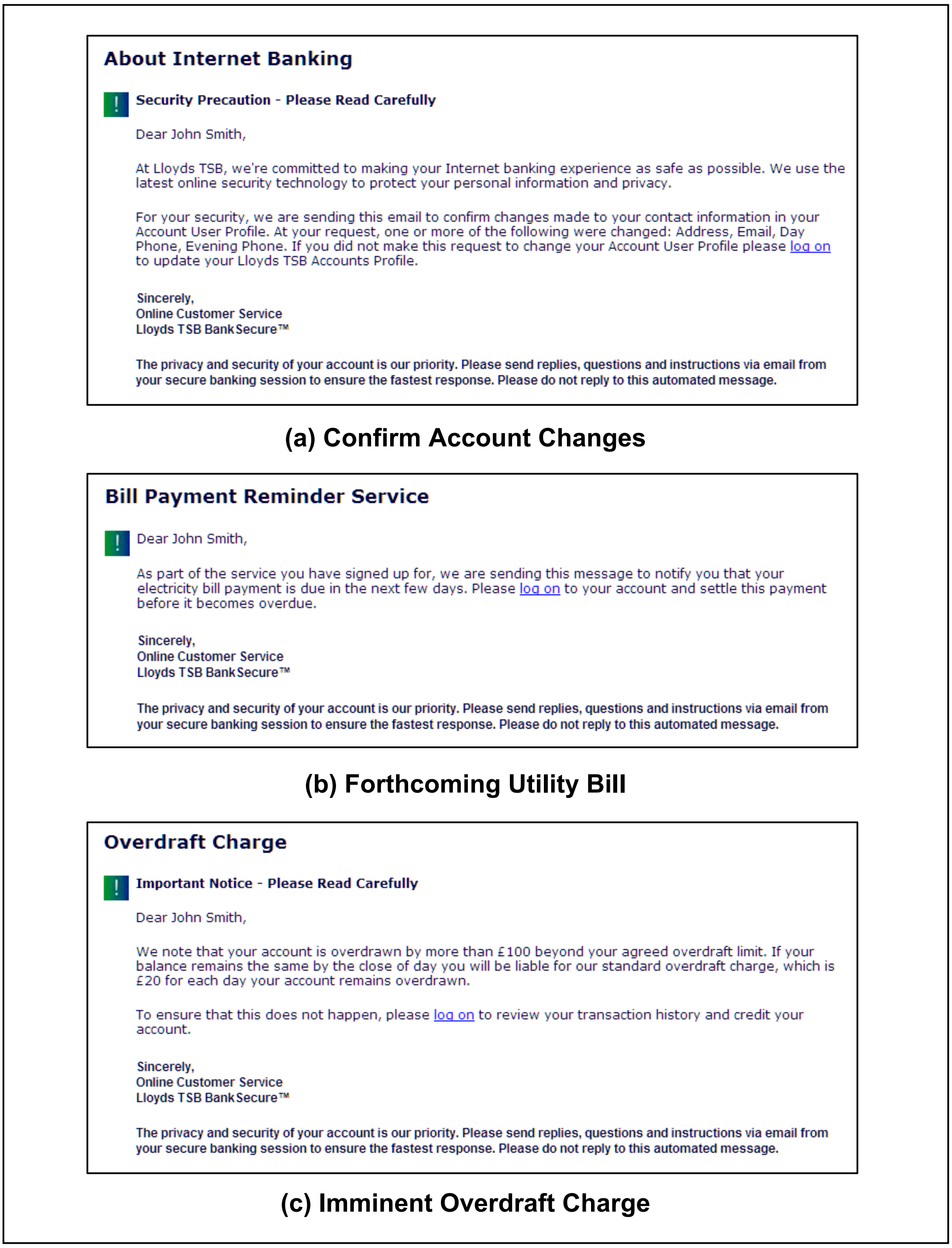
Log on

Sincerely,
Online Customer Service
Lloyds TSB BankSecure™

The privacy and security of your account is our priority. Please send replies, questions and instructions via email from your secure banking session to ensure the fastest response. Please do not reply to this automated message.

(c) Embedded HTML Form

Figure 5.1: Three Styles of Links in Phishing Emails



About Internet Banking

! **Security Precaution - Please Read Carefully**

Dear John Smith,

At Lloyds TSB, we're committed to making your Internet banking experience as safe as possible. We use the latest online security technology to protect your personal information and privacy.

For your security, we are sending this email to confirm changes made to your contact information in your Account User Profile. At your request, one or more of the following were changed: Address, Email, Day Phone, Evening Phone. If you did not make this request to change your Account User Profile please log on to update your Lloyds TSB Accounts Profile.

Sincerely,
Online Customer Service
Lloyds TSB BankSecure™

The privacy and security of your account is our priority. Please send replies, questions and instructions via email from your secure banking session to ensure the fastest response. Please do not reply to this automated message.

(a) Confirm Account Changes

Bill Payment Reminder Service

! Dear John Smith,

As part of the service you have signed up for, we are sending this message to notify you that your electricity bill payment is due in the next few days. Please log on to your account and settle this payment before it becomes overdue.

Sincerely,
Online Customer Service
Lloyds TSB BankSecure™

The privacy and security of your account is our priority. Please send replies, questions and instructions via email from your secure banking session to ensure the fastest response. Please do not reply to this automated message.

(b) Forthcoming Utility Bill

Overdraft Charge

! **Important Notice - Please Read Carefully**

Dear John Smith,

We note that your account is overdrawn by more than £100 beyond your agreed overdraft limit. If your balance remains the same by the close of day you will be liable for our standard overdraft charge, which is £20 for each day your account remains overdrawn.

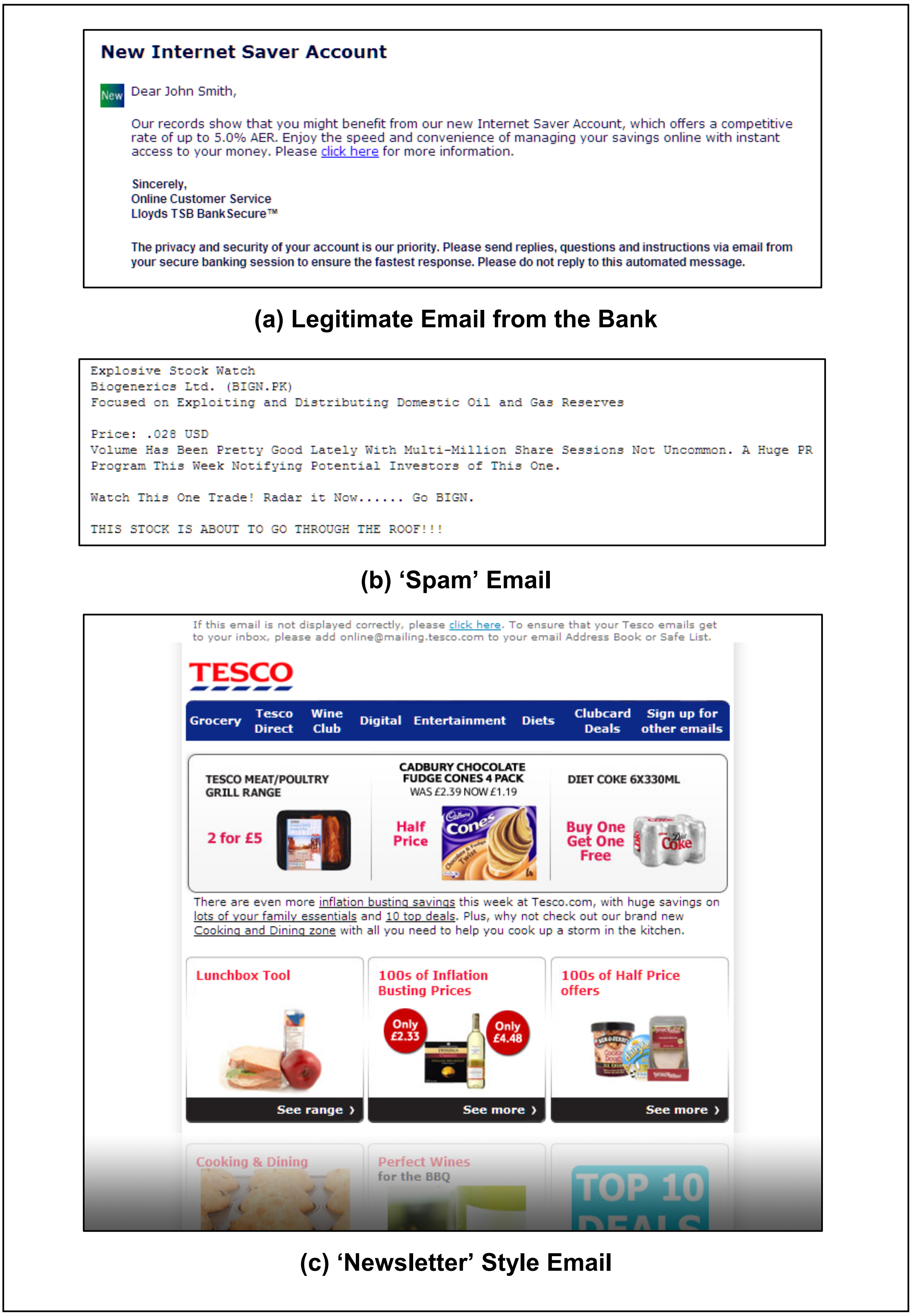
To ensure that this does not happen, please log on to review your transaction history and credit your account.

Sincerely,
Online Customer Service
Lloyds TSB BankSecure™

The privacy and security of your account is our priority. Please send replies, questions and instructions via email from your secure banking session to ensure the fastest response. Please do not reply to this automated message.

(c) Imminent Overdraft Charge

Figure 5.2: Three Different Levels of Content Urgency in Phishing Emails



New Internet Saver Account

Dear John Smith,

Our records show that you might benefit from our new Internet Saver Account, which offers a competitive rate of up to 5.0% AER. Enjoy the speed and convenience of managing your savings online with instant access to your money. Please click here for more information.

Sincerely,
Online Customer Service
Lloyds TSB BankSecure™

The privacy and security of your account is our priority. Please send replies, questions and instructions via email from your secure banking session to ensure the fastest response. Please do not reply to this automated message.

(a) Legitimate Email from the Bank

```
Explosive Stock Watch
Biogenerics Ltd. (BIGN.PK)
Focused on Exploiting and Distributing Domestic Oil and Gas Reserves

Price: .028 USD
Volume Has Been Pretty Good Lately With Multi-Million Share Sessions Not Uncommon. A Huge PR
Program This Week Notifying Potential Investors of This One.

Watch This One Trade! Radar it Now...... Go BIGN.

THIS STOCK IS ABOUT TO GO THROUGH THE ROOF!!!
```

(b) ‘Spam’ Email

If this email is not displayed correctly, please click here. To ensure that your Tesco emails get to your inbox, please add online@mailing.tesco.com to your email Address Book or Safe List.

There are even more inflation busting savings this week at Tesco.com, with huge savings on lots of your family essentials and 10 top deals. Plus, why not check out our brand new Cooking and Dining zone with all you need to help you cook up a storm in the kitchen.

(c) ‘Newsletter’ Style Email

Figure 5.3: Three Benign Emails

The email client used in the experiment was IE6 running a Webmail client application with a preview pane and a generic appearance and layout, modified to prevent easy deletion of emails without first reading them in the experiment. The Webmail client was connected to an hMailServer IMAP server running on a Windows 2003 server. The '**hosts**' file of computers used in the experiment was modified in order to redirect all clicks and web traffic to the same internal server.

An inventory of 12 emails was created, consisting of 3 fraudulent emails of increasing urgency, each with 3 styles of user response; plus 3 benign emails. To avoid experiment bias, the order of presentation of the emails in the list was randomised across participants. The order of the 4 'bank' emails (3 as phishing emails, 1 valid email) was randomised, a total of 4! = 24 orders. The two other (benign) emails were distributed amongst the other emails in a balanced, random ordering. Data from click logs were used to report when participants avoided links in phishing emails and to report when they were duped and responded to the different email types and different levels of message urgency.

Before starting the experiment, each participant was given a brief overview of the Webmail client being used in the experiment.

During their email reading sessions, all participants were monitored by Tobii 1750 gaze-tracking cameras (with their prior consent) to deliver data on the content in each of the emails that participants viewed; and to detect which security features they actually checked in the spoof 'bank' web sites.

After dealing with their batch of 6 emails, each participant was asked several questions to assess their knowledge of online security and phishing attacks and their consequences.

5.1.2. Summary of Experiment Metrics

The details of the experiment are summarised in Table 5.1.

The Phishing Emails Experiment	
Experiment purpose	Exploration of customer awareness of phishing attacks.
Null hypothesis (H_0)	The content urgency and the response style of phishing emails have no effect on the participants' propensity to click on phishing links within the email.
Experiment design	Participants were asked to 'respond appropriately' to 6 emails, 3 of which were phishing emails.
Dependent variables	Button click data. Gaze-tracking data.
Other data	Demographic and technographic data. Exit interview data.
Independent variables	Experiment: email order (24 orders, balanced). Participant: gender (2 genders, in proportions women:men 60:40), age group (2 groups, balanced, split aged 18 to 44 / 45 and over).
Cohort	96 (2 age groups × 2 genders × 24 email orders) = 96. All registered Internet bankers.
Honorarium	Personal cheque for £30.
Session time	30 minutes.

Table 5.1: Experiment Design Details

Unlike in the other experiments discussed earlier, the proportions of women to men was designed to be 60:40, Table 5.1, to reflect the fact that it has been reported that women are more likely to be duped by phishing email attacks than are men (Jagatic et al., 2007; Sheng et al., 2010). The measurements collected in the experiment included demographic and technographic characteristics of participants (e.g. Internet banking usage); click data; gaze tracking data; debriefing interview responses. The null hypothesis was tested by means of analysis of participants' button click data as gathered during their email sessions.

5.1.3. Results

5.1.3.1. Demographic and Technographic Results

A total of 96 participants were recruited. However, the data for 3 participants were found to be incomplete, resulting in a final total of 93, 35 males and 58 females, with 45 participants being aged 18 to 44; and 48 being aged 45 and over.

All of the participants were Internet banking users, with 77% having used Internet banking within the week prior to the experiment and 69% reporting that they used

Internet banking at least once per week. Overall, the cohort was considered valid as part of the target market sector to take part in the experiment and elicit views on Internet banking security issues.

Participants were also asked where they most often logged in to Internet banking and a large majority (80%) reported that they used Internet banking primarily from home. Asked how many emails they received per week that appeared to be from their bank, some 63% said that they never received any emails from their bank, while 21% said that they seemed to receive an email from their bank once a week.

5.1.3.2. Email Task Results

The (phishing response) button click activity of participants was recorded during their email reading session. Participants' propensity to be duped by, and respond to, the phishing emails is detailed in Table 5.2. Note that responding in this case relates to the participant clicking on the phishing link in the email but excludes any instances of participants sending an email in reply.

	Account Changes	Bill Payment	Pending Overdraft	Overall
Clickable Hyperlink Style	42.4%	67.7%	72.4%	60.2%
Clickable Button Style	53.3%	64.5%	53.1%	57.0%
Embedded HTML Style	33.3%	51.6%	65.6%	50.5%
Overall	43.0%	61.3%	63.4%	55.9%

Table 5.2: Propensity to Respond to Phishing Emails, by Style and Urgency

Younger participants showed a slightly higher propensity (57.6%) to respond to phishing emails than did older participants (53.2%). Women showed a slightly higher propensity (57.7%) to respond to phishing emails than did men (53.1%).

The lower urgency of the 'confirm account changes' email is reflected in the lower overall number of participants (43.0%) who responded to it compared to those who responded to the 'bill payment' email (61.3%) and 'pending overdraft charge' email (63.4%). These results suggest that message urgency has an effect on (duping) response rates to phishing email attacks.

The embedded HTML form style of response also resulted in lower uptake (50.5%) overall compared to the clickable hyperlink (60.2%) and clickable button (57.0%)

styles. This is possibly due to more people being aware of the security risks in completing Web forms in such emails.

Statistical analysis (Cochran's Q) of the propensity to respond results confirmed that the effect of content urgency in the phishing emails was statistically significant, $\chi^2(2)=13.625$, $p=0.001$. The result of Cochran's Q test on the link style of phishing emails shows that there was no statistically significant effect of link style in the email message on the uptake of phishing emails.

Pairwise comparisons using McNemar's test showed that the differences between the uptake of the 'confirm account changes' email and each of the other two content urgencies were highly significant ($p=0.005$) where the less urgent 'confirm account changes' email had significantly less uptake than the other two phishing emails, but the difference between the 'bill payment' email and the 'pending overdraft' email was not statistically significant.

Without exception, all of the participants who were duped by a phishing email and clicked on the hyperlink or completed the HTML form then proceeded to log on to Internet banking. No one attempted any alternative approaches such as typing the bank's URL into the address bar of the Web browser and no one backed out during the log on sequence. Only 21.5% of participants did not respond to any of the phishing emails.

Some 37% of participants clicked on the link in the genuine email from the bank that offered details on a new Internet Saver account. Of the 5 participants who sent a reply to one of the emails, 1 replied to the (benign) store email "*As per your message*"; 2 replied to the investment opportunity spam email with "*Thanks for this information*" and "*Thanks for info*"; 1 participant sent an email reply to the 'bill payment reminder' phishing email "*Thank you for your email with reminder*"; and 1 participant sent an email reply to the 'pending overdraft charge' phishing email.

There were 12 participants who attempted to delete emails without first reading them. Here 6 participants indicated that they would always try to delete emails according to sender and subject rather than looking at them first; 2 replied that they would normally move such emails to the deleted items folder after a quick scan and could always retrieve them if they were important; 2 replied that they "*wouldn't read* *adverts*" or would "*delete when I see people offering loans and things*"; 1 replied that they would rely on the spam filter and so wouldn't normally click on emails that they didn't recognise; and 1 replied that "*in my experience I don't get emails from the bank.*"

After working through their list of 6 emails, when participants were told "*I'm not sure if you were aware in what you've just done, but not all of those emails were genuine*", 29.0% of participants showed no obvious reaction, 40.9% said they had worked that out, 16.1% said they had had not noticed but were concerned by the revelation and 14.0% said they had not noticed and were not concerned.

Only 20 participants (7 males and 13 females) did not respond to any of the phishing emails in the experiment and only 7 (2 males and 5 females) of those participants who said they had worked out that some of the emails were not genuine did not respond to any of the phishing emails in the experiment.

When asked, 49 of the 93 participants (52.7%) (22 were older males) claimed they already knew what phishing emails were before they arrived for the experiment, "*Emails that look like they're from a bank – looking for your details.*" However, only 12 of these did not respond to any of the phishing emails in the experiment. Interestingly, 11 of the 49 (9 were women) wrongly described spam emails as being typical phishing attacks.

These data clearly indicate that participants' self-reporting on their knowledge of phishing is at odds with their exhibited resistance to being duped by the types of phishing emails encountered in the experiment.

5.1.3.3. Gaze Results

During the experiment gaze data were logged, and the proportion of participants who gazed at particular regions of the emails is shown in Table 5.3, Table 5.4 and Table 5.5 for the three email types.

Gaze Region	Confirm Account Changes	Bill Payment Due	Potential Overdraft Charge
Sender’s email address	75.8%	78.6%	64.5%
Bank’s logo	39.4%	35.7%	48.4%
Message title	57.6%	71.4%	81.0%
Message text	84.9%	78.6%	90.3%
Sender’s signature	63.6%	50.0%	64.5%

Table 5.3: Gaze Records (% of Participants) – Text Hyperlink

Gaze Region	Confirm Account Changes	Bill Payment Due	Potential Overdraft Charge
Sender’s email address	93.3%	84.4%	86.7%
Bank’s logo	46.7%	34.4%	20.0%
Message title	86.7%	75.0%	87.0%
Message text	96.7%	87.5%	93.3%
Sender’s signature	93.3%	84.4%	83.3%

Table 5.4: Gaze Records (% of Participants) – Button Hyperlink

Gaze Region	Confirm Account Changes	Bill Payment Due	Potential Overdraft Charge
Sender’s email address	86.2%	93.8%	90.3%
Bank’s logo	41.4%	28.1%	25.8%
Message title	93.1%	87.5%	87.0%
Message text	72.4%	75.0%	71.0%
Sender’s signature	58.6%	31.3%	29.0%

Table 5.5: Gaze Records (% of Participants) – HTML Form

These gaze data actually reveal no new insights into customers’ behaviour with the phishing emails, apart from confirming that the key regions of the emails were generally adequately studied in the experiment.

5.1.3.4. Interview Comments

When asked what practical steps the 49 ‘phishing aware’ participants had been taking to avoid problems from phishing emails, 25.8% reported that they used anti-virus or Internet security software; others that they simply employ good practices such as filtering messages based on subject (25.8%) or sender (10.8%). Only 9.7% of participants had not taken any practical steps to avoid threats from phishing emails.

Only 1 participant had been the victim of a phishing email but 10 participants knew of someone who had been a victim.

5.1.4. Discussion on Experiment Results

The null hypothesis (H_0) for this experiment was:

> *"The content urgency and the response style of phishing emails have no effect on the participants' propensity to click on phishing links within the email."*

The evidence presented here is sufficient to refute this null hypothesis since higher levels of urgency in the phishing emails was shown to have a significant effect on the propensity of participants to click on clickable hyperlinks within phishing emails. However, the style of the clickable hyperlink (text hyperlink or button hyperlink or embedded HTML form) in phishing emails had no significant effect on response rates.

Although 57% of participants were 'phishing aware', claiming that they knew what phishing emails were before the experiment, only 21.5% of participants did not respond to any of the phishing emails in their experiment session.

The perceived urgency in the message of the phishing emails resulted in the expected relative increase in duping / uptake of participants: 43.0% responded to the 'confirm account changes' email, 61.3% responded to the 'bill payment reminder' email and 63.4% responded to the 'pending overdraft charge' email.

In the experiment, participants exhibited low resilience to email phishing attacks and these data clearly indicate that participants' self-reporting on their knowledge of phishing is at odds with their exhibited resistance to being duped by the types of phishing emails encountered in the experiment.

Users' behaviour with phishing emails in the experiment confirms that although phishing is now an established form of eCrime, a large proportion of Internet users still know little or nothing about it and a better general awareness of online security, possibly developed by customer education about the need for Internet security would be of benefit. This tutorial approach is considered in the next experiment.

5.2. The Internet Security Tutorial Experiment

To further the observations on the absence of user skills in combating phishing attacks catalogued above, this experiment was designed to explore the effectiveness of online tutorials covering Internet banking security in driving changes in customer behaviour.

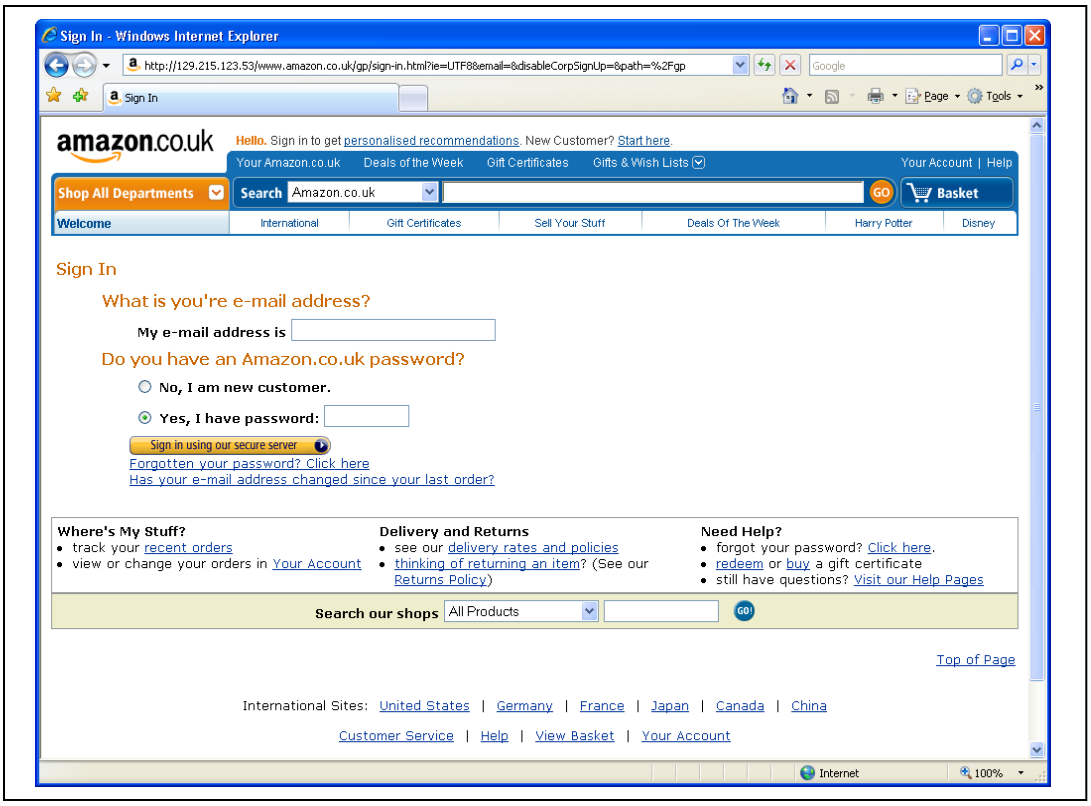
5.2.1. Experiment Design

There were three parts to the experiment – test to establish a baseline measure, give an on-line tutorial, and then re-test to assess learning impact. The test was based on a popular online phishing IQ test[1], and used security guidelines available on the bank's Internet banking site. In all parts of the experiment, all participants were monitored by Tobii 1750 gaze-tracking cameras (with participants' prior consent) to deliver data on the security features they actually checked for in the web pages they encountered in the experiment.

In the first part of the experiment the participant undertook the test viewing a set of 10 different web sites being asked to identify which ones they thought were fraudulent and which were genuine.

In addition to the bank's genuine Internet Banking web site, 9 other example web sites were used, taken from popular eCommerce sites. For each of the 10 web sites, 3 versions, exhibiting a range of typical fraud features, were created for use as the fraudulent sites in the experiment. All three of each of the fraudulent versions included a bad URL in the address bar of the web browser; in addition 2 failed to show an '**`https`**' secure connection with a corresponding absence of the padlock icon in the browser toolbar; and 1 also contained bad grammar in the content of the web page, Figure 5.4, ("*you're*" instead of "*your*", "*I am new customer*" with missing article.) For each participant, 4 of the 10 sites they saw were genuine and 6 were fraudulent (2 with bad URL only; 2 with bad URL but no '**`https`**' or padlock; and 2 with bad URL but with no '**`https`**' or padlock and bad grammar).

[1] The MailFrontier Phishing IQ Test at http://www.sonicwall.com/furl/phishing.



http://129.215.123.53/www.amazon.co.uk/gp/sign-in.html?ie=UTF8&email=&disableCorpSignUp=&path=%2Fgp

Hello. Sign in to get personalised recommendations. New Customer? Start here.

Sign In

What is you're e-mail address?

My e-mail address is

Do you have an Amazon.co.uk password?

○ No, I am new customer.

● Yes, I have password:

Forgotten your password? Click here
Has your e-mail address changed since your last order?

Where's My Stuff?
- track your recent orders
- view or change your orders in Your Account

Delivery and Returns
- see our delivery rates and policies
- thinking of returning an item? (See our Returns Policy)

Need Help?
- forgot your password? Click here.
- redeem or buy a gift certificate
- still have questions? Visit our Help Pages

Search our shops

International Sites: United States | Germany | France | Japan | Canada | China

Customer Service | Help | View Basket | Your Account

Figure 5.4: Web Site used in Test – with Bad URL, no 'https', Bad Grammar

In the second part of the experiment, participants studied a tutorial based on 7 'top tips' on Internet security, Figure 5.5, which focused on identifying fraudulent emails and web sites. Each of the tips was presented on a separate web page with either an example illustrating the concept or a further explanation. Participants were asked to choose whether they thought each tip was helpful or not by clicking on the appropriate button at the bottom of each page, which would then advance to the next page. However, this was done simply in order to encourage participants to read the tips, rather than for later analysis.

In the third part of the experiment participants were shown a second set of different examples based on the same 10 web sites (again, 4 genuine and 6 fraudulent) and again asked to identify the fraudulent ones. The overall aim was to assess how effective their learning had been in the tutorial. In both the 'before' and 'after' quizzes, one of the Web sites was their bank's Internet banking site.

1. **Don’t click on links in emails that are unsolicited or seem suspicious in any way.**
2. **Check that the website is using a secure connection.**
3. **Check that the website has a genuine site certificate.**
4. **Check the address of any website that asks for important personal details.**
5. **When in doubt, type in the address of a website directly.**
6. **Beware of pop-ups.**
7. **Beware of bad English.**

Figure 5.5: Seven ‘Top Tips’ Used in Internet Security Tutorial

5.2.2. Summary of Experiment Metrics

The same participants as the previous (phishing) experiment were involved with this experiment, which was run immediately after their session in the previous phishing emails experiment. The details of the experiment are summarised in Table 5.6.

The Internet Security Tutorial Experiment	
Experiment purpose	Exploration of customer awareness of Internet security.
Null hypothesis (H_0)	The Internet Security tutorial based on 7 ‘top tips’ will have no effect on detection of fraudulent web site test scores.
Experiment design	Participants are asked to identify the fraudulent web sites from a set of 10 before being shown a tutorial on Internet security. They are then shown another set of 10 web sites and asked to identify the fraudulent sites.
Dependent variables	Web site quiz results. Gaze-tracking data.
Other data	Demographic and technographic data.
Independent variables	Experiment: web site order (randomised). Participant: gender (2 genders, in proportions women:men 60:40), age group (2 groups, balanced, split aged 18 to 44 / 45 and over).
Cohort	96 (2 age groups × 2 genders × over-sampling ratio 24:1) = 96. All registered Internet bankers involved in the previous experiment.
Honorarium	Included in previous experiment.
Session time	30 minutes.

Table 5.6: Experiment Design Details

The measurements collected included demographic and technographic characteristics of participants (e.g. Internet banking usage); quiz data; gaze tracking data.

5.2.3. Results

5.2.3.1. Demographic and Technographic Results

All of the 93 participants involved in the previous (phishing emails) experiment took part in this experiment (see Section 5.1.3.1).

5.2.3.2. Fraudulent Web Site Test Results

Every participant was given a score out of 10 after being tested on the first set of web sites, and then also given a score out of 10 after being tested again on the second set. The means of these first test and second test scores were calculated for each of the four age and gender groups and also across all participants, Table 5.7.

	First test score	Second test score	Improvement
Males, 18 to 44	7.26	8.47	+1.21
Males, 45 and over	5.87	7.06	+1.19
Females, 18 to 44	5.19	8.12	+2.93
Females, 45 and over	5.16	7.55	+2.39
All participants	5.73	7.81	+2.08

Table 5.7: Mean Test Scores

Each of the four age and gender groups improved their test scores after being given the tutorial on Internet security. Younger males tended to have a higher initial score than older males and females: and females, notably younger females, showed the largest improvement in their second score.

A repeated-measures ANOVA was carried out on these test scores with the first and second test scores as the within-subjects variable, and taking age and gender as between-subjects factors. The resulting ANOVA table is shown in Table 5.8.

Source	Sum of Squares	*df*	Mean Square	*F*	Sig.
Scores	159.849	1	159.849	72.735	0.000
Scores × Age	0.841	1	0.841	0.383	0.538
Scores × Gender	22.816	1	22.816	10.382	0.002
Scores × Age × Gender	0.708	1	0.708	0.322	0.572
Error	193.398	88	2.198		

Table 5.8: ANOVA Test Results for Mean Test Scores

Results of the ANOVA tests confirm that the difference in scores between first and second tests was highly significant, $F(1,88)=72.735$, $p<0.001$, Table 5.8. There was also a significant interaction with gender, $F(1,88)=10.382$, $p=0.002$, Table 5.8. Whilst males scored significantly higher than females in the first test, this difference was not significant in the second test. This suggests that the tutorial had a more beneficial impact on women than on men in the experiment.

The percentages of genuine web sites that were correctly classified in the quiz are shown in Table 5.9, separated by age and gender. Only 58.9% of older female participants were able to correctly identify the genuine web sites before being given the tutorial, yet 81.5% were able to correctly identify the genuine web sites after the tutorial.

Group	First test	Second test
Males, 18 to 44	82.9%	86.8%
Males, 45 and over	76.6%	82.8%
Females, 18 to 44	78.8%	81.7%
Females, 45 and over	58.9%	81.5%
All participants	72.6%	82.9%

Table 5.9: Percentage of Genuine Web Sites Correctly Classified

Similarly, the percentages of fraudulent web sites that were correctly classified in the quiz are shown in Table 5.10, separated by age and gender. Only 47.8% of participants managed to correctly identify the fraudulent web sites before the tutorial, increasing to 73.7% after the tutorial.

Group	First test	Second test
Males, 18 to 44	65.8%	83.3%
Males, 45 and over	46.9%	62.5%
Females, 18 to 44	37.8%	71.2%
Females, 45 and over	45.7%	75.8%
All participants	47.8%	73.7%

Table 5.10: Percentage of Fraudulent Web Sites Correctly Classified

There was a much larger improvement across all participants in identifying the fraudulent web sites (+25.9%) after the tutorial when compared to identifying the genuine web sites (+10.3%).

Before being given the tutorial, only 4% of participants were able to correctly classify all 10 web sites. After the tutorial, 12% of participants correctly classified all 10 web sites, an 8% improvement.

These results are calculated for only 92 of the participants since 1 of the participants failed to experience the second test. There were 7 instances of a test terminating prematurely, mainly due to the gaze tracker software failing, but the participant was able to re-start the test from the same point.

5.2.3.3. Gaze Results

Behaviour of participants was assessed based on gazes to the important security features of the browser address bar (‘`https`’ presence) and padlock icon which were recorded as participants viewed the 10 web sites in both the first and second tests, as reported in Table 5.11 (for the 85 participants with complete gaze data). Although 2 of the 4 versions of each web site did not have a padlock icon, gazes to where it would have been were recorded in an attempt to gauge if participants still checked for its presence.

Group	Address bar		Padlock icon	
	First test	Second test	First test	Second test
Males, 18 to 44	74.2%	80.5%	11.6%	14.2%
Males, 45 and over	56.2%	77.7%	8.5%	30.0%
Females, 18 to 44	46.8%	71.6%	6.0%	22.0%
Females, 45 and over	48.6%	85.4%	7.5%	26.8%
Overall	54.1%	79.0%	7.9%	23.1%

Table 5.11: Percentage of Participants Who Checked Web Site Security

The results in Table 5.11 indicate the effectiveness of the Internet security tutorial in raising user awareness of the need to examine the details of the target URL and presence of the padlock icon in providing reassurance of the authenticity of a web site.

Repeated-measures ANOVAs were carried out using the gaze data for address bar as the within-subjects variable in one case; and the gaze data for padlock icon in the other. Age group and gender were taken as between-subjects variables. The within-subjects variable (test) was measured on an integer scale of 0 to 10 representing the number of web sites in which the participant gazed at least once at the relevant screen region (address bar or padlock).

The number of participants who gazed at the address bar rose from 54.1% on the first test to 79.0% on the second test and the results of the ANOVA test, Table 5.12, confirm that the difference was statistically significant, $F(1,81)=26.696$, $p<0.001$. There was no effect due to age group but the results indicate that the effect for gender was near significance with women showing a greater increase between tests than did men.

Source	Sum of Squares	*df*	Mean Square	*F*	Sig.
Within-Subjects Effects					
Test	1.949	1	1.949	26.696	0.000
Test × Gender	0.277	1	0.277	3.797	0.055
Test × Age	0.180	1	0.180	2.470	0.120
Error (Test)	5.912	81	0.073		
Between-Subjects Effects					
Gender	0.320	1	0.320	3.246	0.075
Age	0.007	1	0.007	0.071	0.791
Gender × Age	0.323	1	0.323	3.276	0.074
Error	7.989	81	0.099		

Table 5.12: ANOVA Test Results for Gaze Data (Address Bar)

The number of participants who gazed at the padlock icon rose from 7.9% on the first test to 23.1% on the second test and the results of the ANOVA test, Table 5.13, confirm that the difference was statistically significant, $F(1,81)=22.239$, $p<0.001$. In this case there were no effects due to age group or gender.

Source	Sum of Squares	*df*	Mean Square	*F*	Sig.
Within-Subjects Effects					
Test	0.861	1	0.861	22.239	0.000
Test × Gender	0.030	1	0.030	0.777	0.381
Test × Age	0.120	1	0.120	3.098	0.082
Error (Test)	3.136	81	0.039		
Between-Subjects Effects					
Gender	0.001	1	0.001	0.026	0.871
Age	0.088	1	0.088	2.457	0.121
Gender × Age	0.010	1	0.010	0.279	0.599
Error	2.886	81	0.036		

Table 5.13: ANOVA Test Results for Gaze Data (Padlock Icon)

5.2.4. Discussion on Experiment Results

The null hypothesis (H_0) for this experiment was:

> *"The Internet Security tutorial based on 7 'top tips' will have no effect on detection of fraudulent web site test scores."*

The evidence presented here is sufficient to refute this null hypothesis since there was an overall improvement in participant scores in identifying fraudulent web sites, rising from 47.8% in the first test to 73.7% in the second test (after the tutorial). Younger males tended to show a higher initial degree of knowledge about Internet security than the other gender and age groups. Females showed the largest improvement after being shown the top tips tutorial. The tutorial was clearly effective in educating users about web site security issues.

The gaze data also backs up this finding, where the number of participants who gazed at the browser address bar rose from 54.1% on the first test to 79.0% on the second test. Similarly, the number of participants who gazed at the screen area where the padlock icon is located in the web browser rose from 7.9% on the first test to 23.1% on the second test. In both cases the difference between first and second test results were statistically significant.

This experiment has demonstrated that online tutorials can be highly effective in educating users about the threat of phishing attacks and the nature of fraudulent web sites.

5.3. The Malware Detection Software Usability Experiment

Malware is the name given to malicious software that fraudsters embed in a target customer's computer by duping the customer into opening an embedded link in an email, or by other means, which then covertly downloads a small program which is injected into the customer's computer, usually in their **system directory**. The role of the malware program is to lurk in the system directory and only when a specific (banking) web site is being accessed, to 'awake' and monitor the security details that the customer is using to access their account; later sending these security details to the fraudster.

The approach to dealing with such malware threats follows the same principles, only overtly. The bank advises the customer to download the bank's approved malware detection software. This program sits in the system directory of the customer's computer and when the customer starts to log in to the bank's Internet banking site to access their accounts, the malware detection program awakes and checks the customer's computer for the presence of malware, sending a report to the bank to take remedial action for that session. The challenge facing the banks is that customers are hesitant in allowing the bank to do exactly the thing they are trying to prevent the fraudster from doing.

Loading malware detection software on the customer's computer allows the bank to assess the risk from malware on a multi-level basis where benign infections such as worms might register as low risk; malware that does not impact the Internet banking session would be medium risk; and malware such as **rootkit** infections that would contaminate the session registered as high risk. Targeted solutions like this allow the bank to have greater control over the security of their Internet banking site and to better profile the customer and their computer when used to access accounts on the bank's Internet banking site.

The key issues being addressed by this experiment were therefore concerned with the levels of customers' understanding about malware and its potential consequences for them as Internet banking users; and customers' propensity to accept the download and installation of software[2] to test for the presence of malware on their computer on an on-going basis each time they access the bank's Internet banking site. Another issue was how customers react to the branding / source of the software they were being asked to download – either coming from the bank or from a software vendor company.

In addition, this experiment was designed to determine the extent to which customers want to be made aware of the malware scan in operation or the results of the malware detection software as it scans during log on to their Internet banking session: and to assess customers' attitudes to the bank's intervention with transactions when a malware threat is detected.

5.3.1. Experiment Design

A cohort of 192 Internet banking customers was recruited to take part in the experiment. Participants were selected as having used their bank's Internet banking site at some point during the 3 months prior to the experiment.

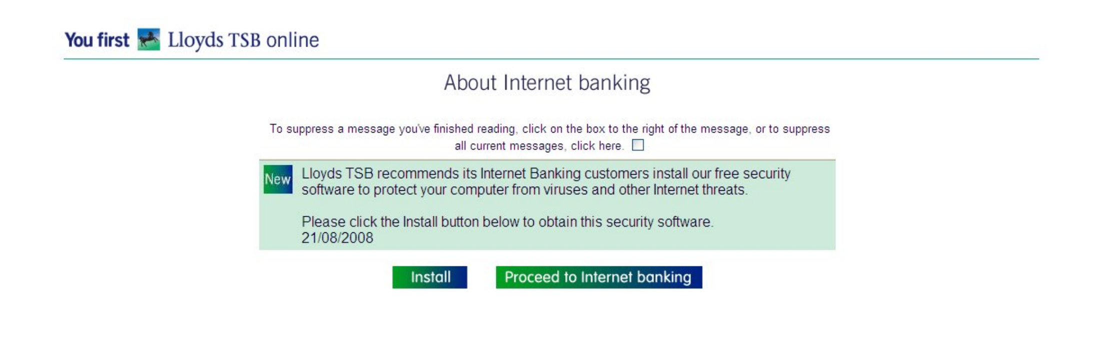
The participant sample was recruited with women and men in proportions 60:40, and in two age groups (ages 18 to 44 and ages 45 and over). Also, to model increasing customer awareness of phishing emails and malware infections by Internet users, half of the cohort were 'phishing aware', being invited back for this experiment having previously experienced the 'top tips' tutorial of the previous experiment dealing with the threat of phishing attacks.

In their experiment session, participants were asked to adopt a fictitious persona with appropriate bank account and Internet banking log on details.

In the first part of the experiment, participants were asked to log on to Internet banking and check 'their' account balance. After log on they encountered a web

[2] The product is only installed once; it is the scan that occurs every time the customer logs in. After installation, the scan can be covert or overt.

page, which recommended they download the bank's approved security software to check the security of their computer, Figure 5.6. For realism in the experiment, the security software appeared to take 2 seconds to download and 15 seconds to complete the initial scan after the log on process. If the participant chose not to download the software, the researcher made a note of this, asked them why they chose not to install the software, and then asked them "*in the interests of the experiment*" to accept the installation of the security software, explaining that it was actually necessary in order for them to continue with the remainder of the experiment. Any participant who continued to decline the offer would be dismissed from the experiment – and paid their honorarium (none declined).



About Internet banking

To suppress a message you've finished reading, click on the box to the right of the message, or to suppress all current messages, click here. □

Lloyds TSB recommends its Internet Banking customers install our free security software to protect your computer from viruses and other Internet threats.

Please click the Install button below to obtain this security software.
21/08/2008

Figure 5.6: Log on Screen Recommending Installation of Security Software

Participants were then asked to again log on to Internet banking and make a payment of a large amount of money (e.g. £500) to a third party account which they have previously set up for their account. After log on, the malware scan process took 5 seconds. By design, the third party funds transfer was not completed by the bank and the participant encountered an Internet banking page asking them to call the bank to authenticate the transaction.

Participants repeated this funds transfer task a total of three times, on each occasion experiencing different levels of overtness of the malware scanning procedure.

The branding of the security software varied between participants. Half experienced download and installation from the bank: the other half experienced download and installation from a leading security software vendor (Symantec).

5.3.1.1. Implementation Details

The web browser used in the experiment was IE6, with the security software download and installation being simulated to avoid issues in the experiment with unsigned software and/or invalid security certificates in Windows. The **hosts** file was modified on each experiment machine so as to redirect the bank URL to the local web server. Data from click logs was used to report the proportions of participants who avoided the installation of the software, and also those that clicked on the security certificate information during the installation of the software.

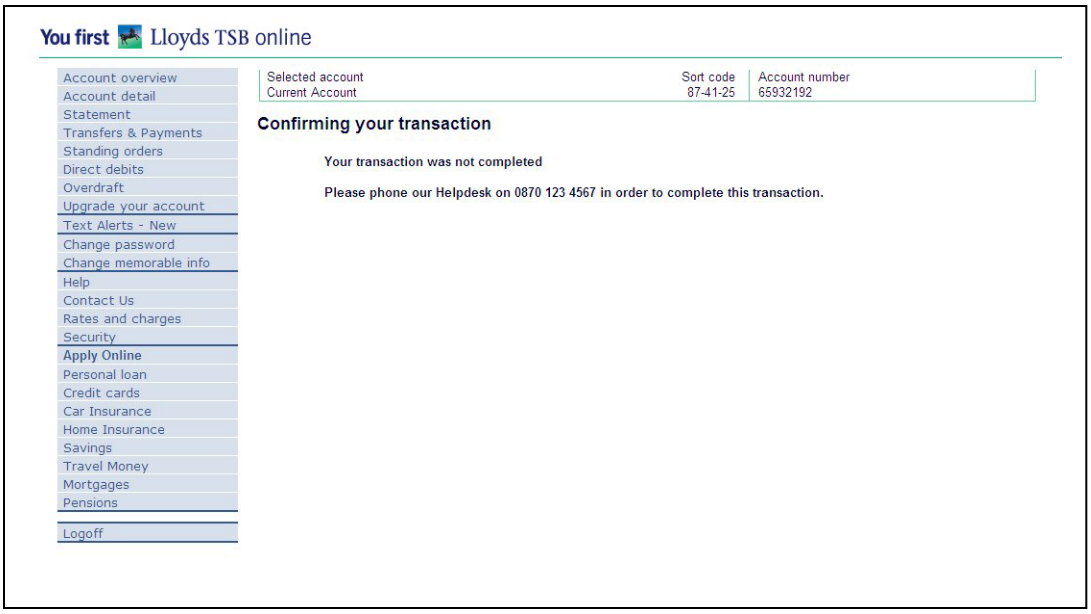
During the sessions, all participants were monitored by gaze tracking cameras (with their consent) to deliver data on what content in the Internet banking pages they actually concentrated on; and on what security features they actually checked during the installation and operation of the security software and reports.

5.3.2. User Interface Designs

Three user interface designs were assessed in the experiment, exhibiting different levels of overtness and user involvement in the malware scanning procedure.

The **Covert design** was characterised by the absence of on-screen information on malware scanning or detections and no specific information about security threats was presented to the user during their Internet banking session. The recommendation to call the bank was presented as matter-of-fact "*Your transaction was not completed. Please phone our Helpdesk on 0870 123 4567 in order to complete this transaction*", Figure 5.7.

The **Overt design** was characterised by malware detection and threat information being presented as a persistent security status bar graphic with warning text in the left-hand menu panel of the Internet banking page. The recommendation to call the bank drew the customer's attention to the risk that had been identified, Figure 5.8, "*Your transaction was not completed. <The bank> have detected a security threat on this computer. Please refer to the panel on the left to see the extent of the threat. Please phone our Helpdesk on 0870 123 4567 in order to complete this transaction*".



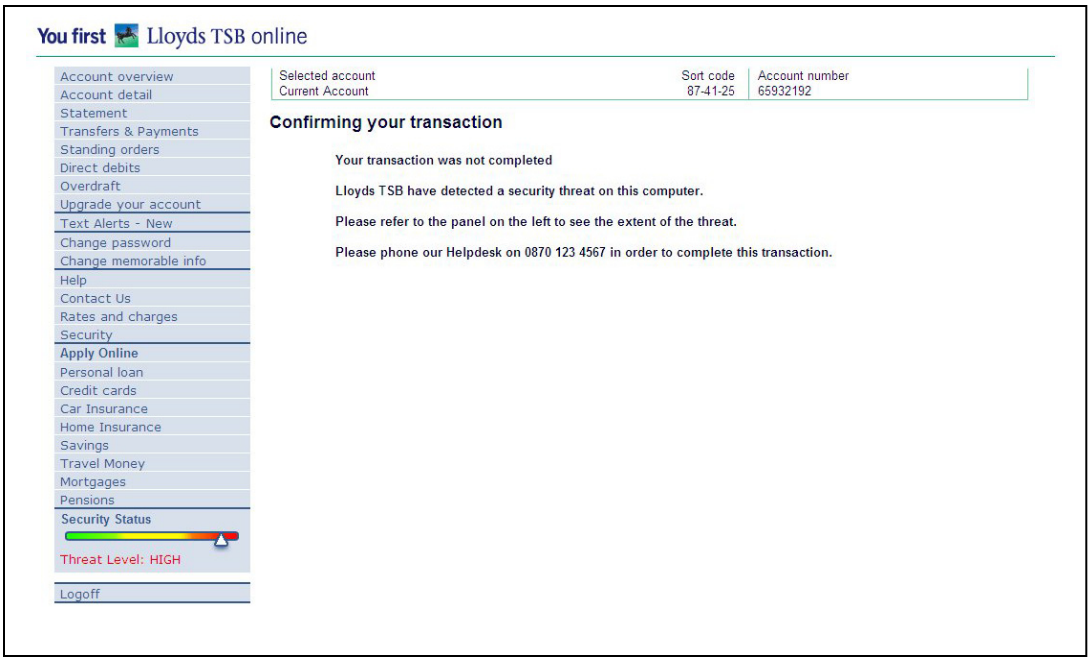
Selected account	Sort code	Account number
Current Account	87-41-25	65932192

Confirming your transaction

Your transaction was not completed

Please phone our Helpdesk on 0870 123 4567 in order to complete this transaction.

Figure 5.7: Covert User Interface Design: 'Transaction was not completed'



Selected account	Sort code	Account number
Current Account	87-41-25	65932192

Confirming your transaction

Your transaction was not completed

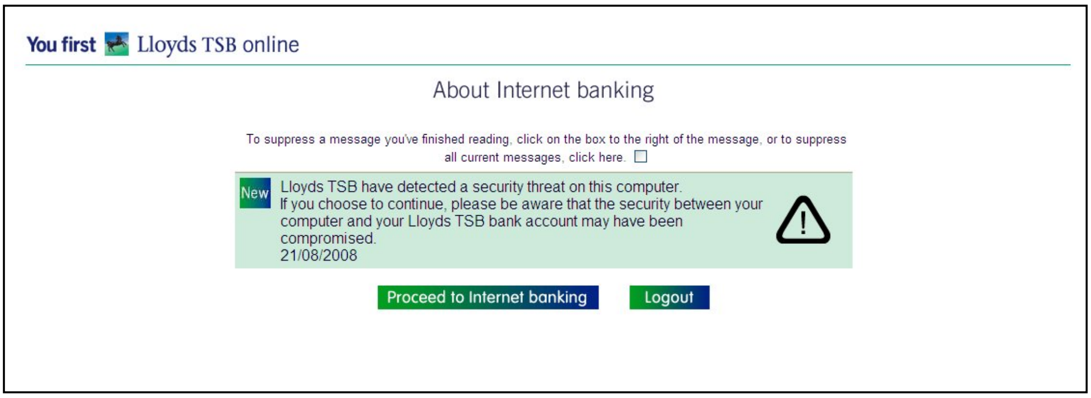
Lloyds TSB have detected a security threat on this computer.

Please refer to the panel on the left to see the extent of the threat.

Please phone our Helpdesk on 0870 123 4567 in order to complete this transaction.

Figure 5.8: Overt User Interface Design: 'Transaction was not completed'

The **Intrusive design** was characterised by malware detection and threat information which was being presented at log on as a dialogue box, Figure 5.9, that interrupted the flow of the user interaction, and needed to be read and closed by the user before proceeding to the Internet banking site. In addition, as with the Overt design, the threat information was presented as a persistent security status bar graphic with warning text in the left-hand menu panel of the Internet banking page. The recommendation to call the bank reminded the customer of the existing risk they had seen when they logged on, Figure 5.10, "*Your transaction was not completed. When you logged on to Internet Banking we advised you of a security threat to this computer. Please refer to the panel on the left to see the extent of the threat. Please phone our Helpdesk on 0870 123 4567 in order to complete this transaction*".

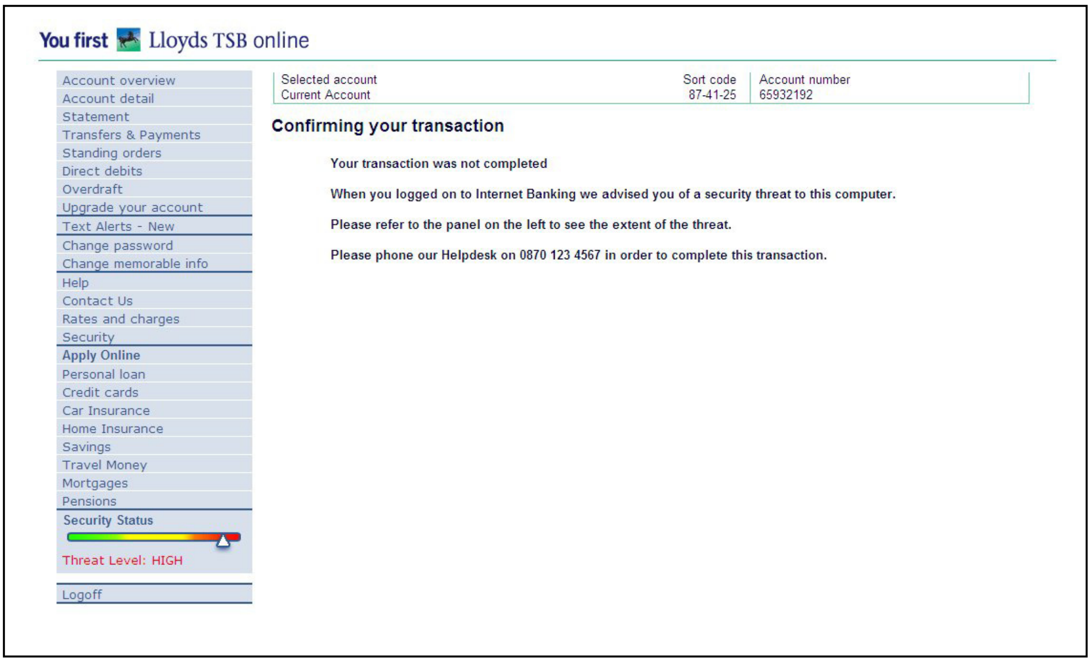


About Internet banking

To suppress a message you've finished reading, click on the box to the right of the message, or to suppress all current messages, click here. □

New Lloyds TSB have detected a security threat on this computer.
If you choose to continue, please be aware that the security between your computer and your Lloyds TSB bank account may have been compromised.
21/08/2008

Figure 5.9: Intrusive User Interface Design: 'Detected a security threat'



Selected account	Sort code	Account number
Current Account	87-41-25	65932192

Confirming your transaction

Your transaction was not completed

When you logged on to Internet Banking we advised you of a security threat to this computer.

Please refer to the panel on the left to see the extent of the threat.

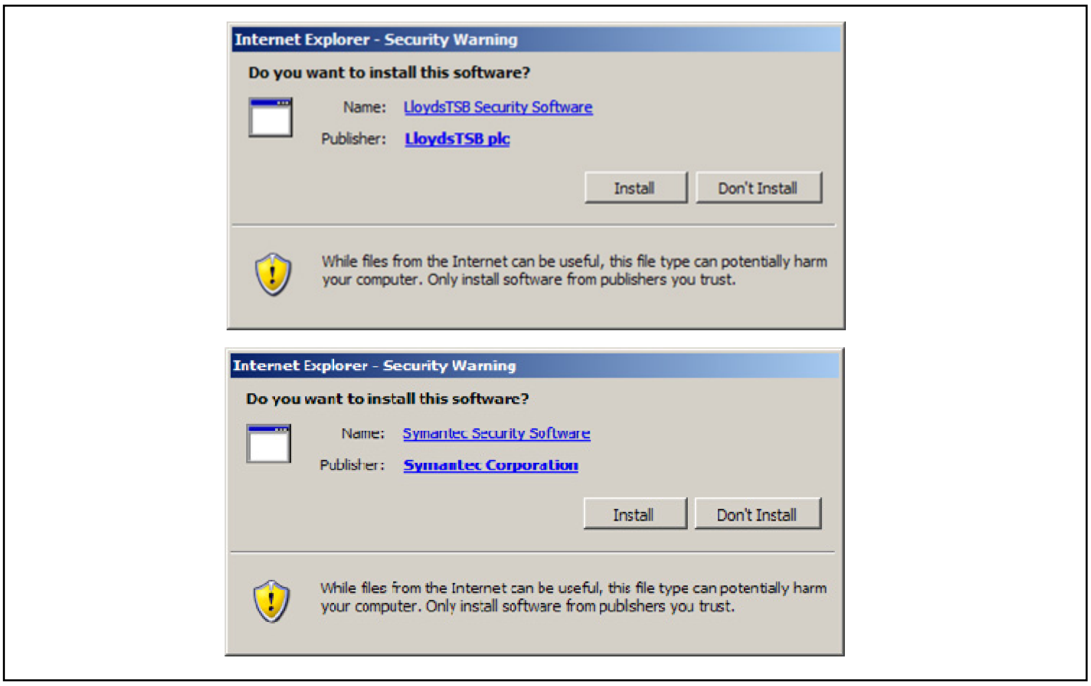
Please phone our Helpdesk on 0870 123 4567 in order to complete this transaction.

Security Status

Threat Level: HIGH

Figure 5.10: Intrusive User Interface Design: 'Transaction was not completed'

The 'name' and 'publisher' fields of the Internet Explorer download screen were set to be either the bank or a leading security software vendor, Figure 5.11, with 50% of participants experiencing each.



Internet Explorer - Security Warning

Do you want to install this software?

Name: LloydsTSB Security Software

Publisher: LloydsTSB plc

While files from the Internet can be useful, this file type can potentially harm your computer. Only install software from publishers you trust.

Internet Explorer - Security Warning

Do you want to install this software?

Name: Symantec Security Software

Publisher: Symantec Corporation

While files from the Internet can be useful, this file type can potentially harm your computer. Only install software from publshers you trust.

Figure 5.11: Malware Detection Software Branding During Installation

Both the fields in the installation window were clickable to pop up a security certificate window detailing the authenticity of the software to be downloaded from either source. Both the download prompt and security certificate window were written to bypass the browser security limitations and so that button click data could be logged in the experiment.

5.3.3. Summary of Experiment Metrics

The measurements collected included: demographic and technographic characteristics of participants (e.g. Internet banking usage); attitude toward usability; quality rating and ratings for the different user interface designs on a 30-point linear scale ('best' to 'worst'); preferences (rank order of the procedures) from the quality ratings; click data; gaze tracking data; interview responses with qualitative comments on specific issues.

After each of their hands-on experiences with each of the three user interface designs participants were asked to complete a short usability questionnaire. The usability questionnaire was based on a modified version of those used in the previous experiments. The subject in each sentence was changed to 'service' and statements on 'ease of understanding', 'helpfulness', 'convenience', 'user friendliness' and 'security information' were added, Table 5.14.

Question Statement (polarity)	Abbreviated Form
I had to concentrate hard to use this service (-)	Concentration
I got flustered when using this service (-)	Flustered
I felt under stress while using this service (-)	Stressed
I found this service frustrating to use (-)	Frustration
I thought this service was too complicated (-)	Complication
When using this service I always knew what to do next (+)	Knew what to do next
I felt in control while using this service (+)	In control
I thought this service was slow (-)	Quick
I would be happy to use this service again (+)	Use again
I felt this service was reliable (+)	Reliable
I felt that this service needed a lot of improvement (-)	Needs improvement
This service did not match my expectations (-)	Matched expectations
I found this service user-friendly (+)	User friendly
I liked using this service (+)	Liked
I did not enjoy using this service (-)	Enjoyment
I found this service trustworthy (+)	Trustworthy
I felt confident in the security of this service (+)	Security
I found this service confusing to use (-)	Confusion
The instructions for completing this service were clear (+)	Instructions
This service was easy to use (+)	Ease of use
This service was difficult to understand (-)	Easy to understand
I felt that this service was unhelpful (-)	Helpful
I found this service convenient to use (+)	Convenient
I felt informed about the security of this service (+)	Informed about security

Table 5.14: Usability Questionnaire Statements

Details of the experiment design are shown in Table 5.15.

The Malware Detection Software Usability Experiment	
Experiment purpose	Exploration of customer attitude to Internet security malware detection software.
Null hypothesis (H_0)	The attitudes of the participants towards usability will be the same for all three of the user interface designs with malware detection software.
Experiment design	Participants experience three different user interface designs using malware detection software. Repeated measures (balanced order).
Dependent variables	Usability attribute scores. Button click data. Gaze-tracking data.
Other data	Demographic and technographic data. Exit interview data.
Independent variables	Experiment: exposure order (3 user interface designs, 6 possible orders, balanced), software branding (2 brands – bank and software vendor). Participant: gender (2 genders, in proportions women:men 60:40), age group (2 groups; ages 18 to 44 / ages 45 and over, balanced), phishing awareness (2 groups; top tips tutorial or not).
Cohort	192 (2 age groups × 2 genders × 2 phishing awareness × 6 treatment orders × 2 software branding × over-sampling ratio 2:1) = 192. All experienced Internet bankers.
Honorarium	Personal cheque for £50.
Session time	90 minutes.

Table 5.15: Experiment Design Details

5.3.4. Results

5.3.4.1. Demographic and Technographic Results

A total of 168 participants were recruited, with 65 males and 103 females participating in the experiment: 86 were aged 18 to 44 and 82 were aged 45 and over.

All of the participants were Internet banking customers who had used Internet banking within the 3 months prior to the experiment, with 93% having used it in the last month and 75% within the last week. The frequency of their usage of Internet Banking was also noted, with 74% using Internet banking at least once a week and 20% using it every day. Overall, the cohort was considered valid as part of the target market sector to take part in the experiment and elicit views on Internet banking security issues.

Participants were also asked where they most often logged in to Internet banking. A large majority (77%) used Internet banking at home and 14% used it at work.

Of the 97 participants that used either glasses or contact lenses, 71 (73%) were wearing them during the experiment. There were 3 (2%) participants who stated that they had colour-blindness issues, although this did not prove to be an issue during the experiment.

By design, 77 of the participants (46%) had taken part in the previous (Internet security tutorial) experiment to highlight typical pitfalls in email and web security. This was treated as a between-subjects variable in the analysis to determine the effect, if any, of phishing awareness on customer attitude to malware detection software.

5.3.4.2. Software Download Results

Participants' button click data were recorded at the security software download prompt. Of the 168 participants, only 50 participants (30%) chose to install the software first time. Another 103 participants did not click on anything and were asked to click "OK" in order to continue with the experiment. Here comments included "*Didn't see the install button*", "*Normally I get stuff like this and just ignore it*", "*I already have security software*".

Of the remainder, 10 participants tried clicking on the "Save" button (presumably in an attempt to examine the file before installation) before being asked to click "OK" and complete the download and installation: and 2 participants tried clicking on the "Cancel" button at the download prompt, but were asked to click "OK" – they then proceeded as required. Of these, 1 participant stated that she would never normally perform such an installation, and the other said that she would probably install it the next time.

There were 3 participants who chose not to install the software, 1 of whom attempted to cancel the download as well and who said that they would never download such software. The other 2 said that they would normally proceed and install it anyway, but for some reason chose not to during the experiment. All 3 completed the installation for the purposes of the experiment.

The button click data reveal that for the 50 participants who completed the download without intervention, 95.6% of those who had not been given the tutorial on Internet security performed the expected installation compared to 85.7% of those who had been given the tutorial: evidence of only a slightly higher awareness of Internet security issues for those who had been given the phishing training.

None of the participants investigated the security certificate of the software download provider, either through clicking the appropriate fields in the download prompt or by clicking the padlock icon in the web browser.

Although each participant had gaze data for at least one of the four tasks in the experiment (download software, 3 login uses), only 91 participants had a full set of gaze data for each of the four tasks. This was due mainly to the participant looking away from the screen to refer to their login details and then not resuming the correct seating position afterwards, with the result that the gaze tracking camera lost registration. The percentages of all participants who looked at the relevant screen regions at least once, and the corresponding percentages of the 91 participants with full gaze data, are shown in Table 5.16.

Screen region	% of all participants (168)	% of those with full gaze data (91)
Browser toolbar	25.6%	33.0%
Browser address bar	26.2%	31.9%
Status bar	11.9%	15.4%
Padlock icon	0.6%	0.0%

Table 5.16: Gaze Reports during Security Software Installation

Here 25.6% of all participants (33.0% of those with full gaze data) looked at the browser toolbar and 26.2% of all participants (31.9% of those with full gaze data) looked at the URL of the web page during the software installation download, perhaps looking for reassurance that the download prompt was genuine: and 11.9% of all participants (15.4% of those with full gaze data) also gazed at the web browser status bar at this time, perhaps looking for similar reassurance. Only one participant looked at the padlock icon during this time, although that could be due to lack of familiarity with IE6 (in IE7 it is located at the address bar) or the fact that the icon was so small that it could not be accurately gaze tracked.

5.3.4.3. Usability Results

Table 5.17 shows the usability questionnaire means for each malware detection user interface design confirming that the usability scores were lowest for the Covert user interface design and highest for the Intrusive user interface design.

Design	Mean Score	Std. Dev.	N
Covert	5.0751	1.06257	168
Overt	5.2989	1.06079	168
Intrusive	5.3105	1.04791	168

Table 5.17: Usability Questionnaire Means by User Interface Design

A repeated-measures ANOVA was carried out on these usability data, with user interface design (Covert vs. Overt vs. Intrusive) as the within-subjects factor and gender (65 male; 103 female), age group (86 aged 18 to 44; 82 aged 45 and over), phishing knowledge (77 had experienced the Internet security tutorial and 91 had not) and software branding (87 bank software download and 81 software vendor) as the between subjects factors. Several combinations of variables including treatment order were explored beforehand. No significant effects for order were noted; therefore order was eliminated from further analysis.

Results of the ANOVA test, Table 5.18, confirm that there was a significant main effect for user interface design, $F(2,304)=10.697$, $p<0.001$. Pair-wise comparisons show that the difference in usability scores between the Covert user interface design (5.08) and the Overt user interface design (5.3) was statistically significant, $p<0.001$: and that the difference in usability scores between the Covert user interface design (5.08) and the Intrusive user interface design (5.31) was also statistically significant, $p<0.001$, with the Covert user interface design scoring significantly lower for usability in both cases. The difference between the Overt user interface design and the Intrusive user interface design was not statistically significant.

Results of the ANOVA test, Table 5.18, indicate a between-subjects effect of age group with the older age group (ages 45 and over) scoring usability higher for all three user interface designs than did the younger age group (ages 18 to 44), $F(1,152)=4.883$, $p=0.029$, Table 5.18, and the difference was statistically significant.

Source	Sum of Squares	*df*	Mean Square	*F*	Sig.
Within-Subjects Effects					
Design	6.391	2	3.195	10.697	0.000
Design × Gender	0.192	2	0.096	0.322	0.725
Design × Age	0.073	2	0.037	0.123	0.885
Design × Phished	1.086	2	0.543	1.819	0.164
Design × Branding	0.428	2	0.214	0.717	0.489
Error (Design)	90.811	304	0.299		
Between-Subjects Effects					
Gender	5.594	1	5.594	2.166	0.143
Age	12.612	1	12.612	4.883	0.029
Phished	0.087	1	0.087	0.034	0.854
Branding	1.643	1	1.643	0.636	0.426
Gender × Age	0.864	1	0.864	0.334	0.564
Gender × Phished	0.680	1	0.680	0.263	0.609
Gender × Branding	0.051	1	0.051	0.020	0.889
Age × Phished	1.227	1	1.227	0.475	0.492
Age × Branding	9.422	1	9.422	3.648	0.058
Phished × Branding	23.541	1	23.541	9.115	0.003
Error	392.554	152	2.583		

Table 5.18: ANOVA Test Results for Usability Attributes

The ANOVA test results, Table 5.18, also show an interaction between the effect of phishing knowledge and branding of the malware detection software, $F(1,152)=9.115$, $p=0.003$. The nature of this interaction is illustrated in Table 5.19. Those with phishing knowledge from the security tutorial gave higher scores to the download from the bank compared to the software vendor. Conversely, those without phishing knowledge from the tutorial gave higher scores to the download from the software vendor.

Phishing Knowledge	Branding	Mean
Phishing aware	Bank	5.500
	Vendor	4.920
Unaware	Bank	5.013
	Vendor	5.351

Table 5.19: Usability Scores, by Phishing Awareness and Download Branding

Pairwise analyses (with Bonferroni correction) for each of the 24 individual usability attributes in the usability questionnaire (Table 5.14) were carried out and reported in Table 5.20. Only main effects were considered in the analysis.

For each of the 11 usability attributes listed in Table 5.20 (as with the mean usability scores reported above) there were no significant pairwise differences between the Overt and the Intrusive user interface designs. There were 7 usability attributes where the Overt user interface design scored significantly higher than the Covert user interface design: and there were 9 usability attributes where the Intrusive user interface design scored significantly higher than the Covert user interface design.

Attribute	Significant Effects	Details
Mean	*df*=2.000; *F*=10.697; *p*<0.001 $p_{Covert\text{-}Overt}$<0.001: $p_{Covert\text{-}Intrusive}$=0.001	Covert < Overt Covert < Intrusive
Frustration	*df*=1.834; *F*=4.012; *p*=0.022 $p_{Covert\text{-}Intrusive}$=0.043	Covert < Intrusive
Use again	*df*=2.000; *F*=7.766; *p*=0.001 $p_{Covert\text{-}Overt}$=0.007: $p_{Covert\text{-}Intrusive}$=0.002	Covert < Overt Covert < Intrusive
Needs improvement	*df*=1.906; *F*=4.933; *p*=0.009 $p_{Covert\text{-}Intrusive}$=0.018	Covert < Intrusive
Matched expectations	*df*=2.000; *F*=10.322; *p*<0.001 $p_{Covert\text{-}Overt}$=0.001: $p_{Covert\text{-}Intrusive}$=0.001	Covert < Overt Covert < Intrusive
User friendly	*df*=1.884; *F*=4.106; *p*=0.019 $p_{Covert\text{-}Overt}$=0.039	Covert < Overt
Liked	*df*=2.000; *F*=7.006; *p*=0.001 $p_{Covert\text{-}Overt}$=0.003: $p_{Covert\text{-}Intrusive}$=0.010	Covert < Overt Covert < Intrusive
Enjoyment	*df*=2.000; *F*=4.623; *p*=0.011 $p_{Covert\text{-}Intrusive}$=0.024	Covert < Intrusive
Helpful	*df*=1.774; *F*=9.846; *p*<0.001 $p_{Covert\text{-}Overt}$<0.001: $p_{Covert\text{-}Intrusive}$=0.009	Covert < Overt Covert < Intrusive
Trustworthy	*df*=2.000; *F*=6.765; *p*=0.001 $p_{Covert\text{-}Intrusive}$=0.001	Covert < Intrusive
Informed about security	*df*=1.881; *F*=34.612; *p*<0.001 $p_{Covert\text{-}Overt}$<0.001: $p_{Covert\text{-}Intrusive}$<0.001	Covert < Overt Covert < Intrusive
Convenient	*df*=1.863; *F*=4.283; *p*=0.017 $p_{Covert\text{-}Overt}$=0.031	Covert < Overt

Table 5.20: Summary of Significant Differences in Usability Attributes

Taken together these data suggest that the Covert user interface design is not the correct approach.

For the three sets of usability questionnaire data, reliability analysis on the scale of each (based on Cronbach's α) reported a minimum value of α=0.956 (for the usability questionnaire for the Covert user interface design), indicating that the questionnaire reliability was high.

5.3.4.4. Quality and Preference Results

The results for the quality scores (30-point linear scale) are shown in Table 5.21, again confirming that the Covert user interface design scores lowest.

Design	Mean Score
Covert	14.94
Overt	18.13
Intrusive	19.81

Table 5.21: Mean Quality Scores by User Interface Design

Results of preliminary repeated-measures ANOVA tests with the quality scores for each of the three user interface designs as the within-subjects variable and with order of experience, gender and phishing awareness as between-subjects variables indicated that gender and phishing awareness had no significant effects or interactions in the quality rating scores so these were omitted from subsequent analysis.

The resulting ANOVA test used the quality scores for each of the three user interface designs as the within-subjects variable and with age group, order of experience and branding as the between-subjects variables. The results of the ANOVA test are shown in Table 5.22. Mauchly's test indicated that the assumption of sphericity had been violated; therefore the degrees of freedom were corrected using the Greenhouse-Geisser estimates of sphericity, as reflected in the non-integer values for degrees of freedom in Table 5.22.

Source	Sum of Squares	*df*	Mean Square	*F*	Sig.
Within-Subjects Effects					
Interface	1916.048	1.784	1074.265	21.864	0.000
Interface × Order	884.001	8.918	99.126	2.017	0.038
Interface × Age	33.509	1.784	18.787	0.382	0.659
Interface × Brand	54.730	1.784	30.685	0.625	0.519
Error (Interface)	12619.511	256.837	49.134		
Between-Subjects Effects					
Order	487.049	5	97.410	1.149	0.337
Age	1060.223	1	1060.223	12.510	0.001
Brand	224.040	1	224.040	2.643	0.106
Error	12204.228	144	84.752		

Table 5.22: ANOVA Test Results for Quality Scores

The results of the ANOVA test confirm that there was a main effect for quality scores of user interface design, *F*(1.784, 256.837)=21.864, *p*<0.001, Table 5.22. Pair-wise comparisons show that there was a significant difference, *p*<0.001, between the Covert user interface design with a quality score of 14.94 and the Overt user interface design with a quality score of 18.13: and also there was a significant difference, *p*<0.001, between the Covert user interface design and the Intrusive user interface design with a quality score of 19.81, with the Covert user interface design scoring significantly lower overall in terms of the quality scores in both cases. There was also a moderately significant difference between the Overt user interface design at 18.13 and the Intrusive user interface design with a quality score of 19.81, *p*=0.034, favouring the Intrusive user interface design in the quality scores.

The ANOVA results indicate a significant interaction with treatment order for the three user interface experiences, emphasising the importance of the care taken to carefully randomise and balance for treatment order across the cohort in execution of the experiment to minimise the impact of order effects on the data.

There was a significant between-subjects effect due to age group, *F*(1,144)=12.510, *p*=0.001, Table 5.22, where the older age group (ages 45 and over) scored all three user interfaces significantly higher than did the younger group (ages 18 to 44).

The quality scores were also used to generate a rank order of preference, Table 5.23. These data indicate a significant bias toward the Intrusive user interface design (Chi-square test for those participants with a preference: $\chi^2(2)=29.918$, $p<0.001$).

User Interface Design	Ranked Best	%
Covert	24	14.3%
Overt	30	17.9%
Intrusive	69	41.1%
No Preference	45	26.7%
Total	168	100.0%

Table 5.23: Preferences for the Alternative User Interface Designs

Participants explained their overall scores, comparing the three designs: "[Intrusive] *gave me a warning at the start. Others – can't remember a difference*", "*To not successfully pay the bill was very frustrating – none of them worked*", "[Intrusive] *felt flustered,* [Covert] *was more comfortable*", "*I was more aware of what was going on during the* [Covert] *session*."

5.3.4.5. Gaze Tracking Data from Different User Interface Designs

As with the security software download and installation task in this experiment, gazes at the browser toolbar, browser address bar, browser status bar and padlock icon were recorded during Internet banking login (bill payment) tasks. In addition, any gazes at the left hand navigation menu in the Internet banking web pages were logged and, in the Overt and Intrusive user interface designs, any gazes at the security status bar located within the navigation menu were also recorded. The gaze results are shown in Table 5.24.

Screen region	% of participants		
	Covert	Overt	Intrusive
Left hand menu	44.0%	50.0%	49.4%
Security status bar	n/a	13.1%	14.3%
Browser toolbar	5.4%	16.7%	10.7%
Browser address bar	8.3%	16.1%	10.7%
Status bar	2.4%	1.8%	1.2%
Padlock icon	0.0%	0.0%	0.0%

Table 5.24: Screen Region Gazes (All 168 Participants)

For the Overt design, participants looked at the browser toolbar and browser address bar more than with the other two designs, although only some 16% of participants gazed at these targets. There were surprisingly few participants who gazed at the security status bar in both Overt and Intrusive user interface designs (some 13%), typically gazing for just half a second, although some gazes at the security status bar lasted for 3 seconds.

The gaze results for the subset of 91 participants who had a full set of gaze data across all of the four tasks are similar to those for the full cohort of 168 participants, Table 5.25.

Screen region	% of participants		
	Covert	Overt	Intrusive
Left hand menu	62.6%	70.3%	63.7%
Security status bar	n/a	17.6%	15.4%
Browser toolbar	6.6%	25.3%	16.5%
Browser address bar	11.0%	25.3%	15.4%
Status bar	2.2%	1.1%	1.1%
Padlock icon	0.0%	0.0%	0.0%

Table 5.25: Screen Region Gazes (91 Participants with Full Set of Gaze Data)

It is a physiological fact that the pupils in the eye dilate, typically by up to 15%, during periods of increased cognitive load, such as heightened awareness, interest or anxiety (Klingner et al., 2008). To investigate this pupillary response in participants with the malware detection software in operation, the pupil size (mm) at each gaze event was recorded, averaged for each user interface design, Table 5.26. Each room was set to similar lighting levels to minimise the effects of differing light intensity on pupil size. Pupil dilation was calculated across the whole cohort in the one case, and also just for those 91 participants with a full set of gaze data. The mean pupil diameter should be interpreted relatively across designs and reflects trends.

Cohort	Pupil diameter (mm)			
	Installation	Covert	Overt	Intrusive
168 participants	3.20	3.11	3.15	3.14
91 participants	3.32	3.20	3.24	3.24

Table 5.26: Mean Pupil Diameter across Designs

The installation task appears to have resulted in larger average pupil dilation when compared to the bill payment tasks, possibly reflecting the additional anxiety caused by the software installation. The Covert design also resulted in smaller pupil dilation when compared to the other user interface designs.

5.3.4.6. Interview Comments

When asked about the differences between the three Internet banking sessions, a total of 39 (23%) participants did not notice any differences in their three bill payment tasks. The remaining 76% noticed some or all of the differences between the designs: 25% of participants remarked only on the security warnings at the end; 20% of participants remarked only on the security status bar; 11% of participants remarked only on the security warning at login; 10% of participants mentioned all of the differences between designs; 5% of participants mentioned all of the differences except they thought that the security status bar was only in one design; 4% of participants remarked only on the security warnings at login and at the end (and failed to mention the security status bar).

When asked what they liked in the Internet banking sessions, a total of 42 (25%) participants said that they specifically liked the security threat slider bar, "*Like the threat level indicator, very visible*", "*Being told about security threat. More security is always good*".

When asked about the factors they consider when deciding on a download, the main considerations for participants were "*Who is the download from?*", "*How big is the file?*", "*How will it interact with my existing software?*"

When asked about who they thought provided their download in the experiment, the vast majority (89%) had failed to notice. Of the 13 participants who thought that the bank had provided the download, 5 gave the wrong answer – their software had been branded as Symantec.

When asked how they felt about the downloaded software scanning their computer and sending the results to the bank, just over half of the participants (54%) either thought it was good thing, or had no problem with the scan; 32% of the participants were either not happy about it, or felt uncomfortable about it. The remainder either did not know, or wanted more information about the scan.

5.3.5. Discussion on Experiment Results

The null hypothesis (H_0) for this experiment was:

> *"The attitudes of the participants towards usability will be the same for all three of the user interface designs with malware detection software."*

The evidence presented here is sufficient to refute this null hypothesis since the differences in mean usability scores between the Covert user interface design and the Overt user interface design; and between the Covert user interface design and the Intrusive user interface design were statistically significant.

Only 30% of participants performed the security software download and installation as instructed by the login sequence without any prompting from the researcher: 9% of participants examined the details of the download or aborted the installation process, being generally very wary of what they would install on the computer. None of the participants examined the security certificate information that was available to them at the download prompt. Some 53% of participants would normally refuse to download or install such software.

The button click data reveal that for the 50 participants who completed the download without intervention, 95.6% of those who had not been given the tutorial on Internet security performed the expected installation compared to 85.7% of those who had been given the tutorial: evidence of a only a slightly higher awareness of Internet security issues for those who had been given the phishing training.

Usability attitudes were highest for the Intrusive user interface design (5.31) and the Overt user interface design at 5.30. These two designs, offering user information about the malware detection operation, scored significantly higher than the Covert design at 5.08. An age group effect revealed that older participants were much more positive in their attitudes than were younger participants towards all three designs, possibly suggesting stronger awareness of the impact of theft in the older age group.

There was also evidence to suggest that those with higher awareness of Internet security and phishing threats gave higher usability scores with bank-branded software than those who had not taken part in a previous tutorial training session on Internet security. Conversely, those who had not taken part in the previous tutorial training session gave higher usability scores with the download from the software vendor.

In terms of individual usability attributes, the Overt user interface design and the Intrusive user interface design for the malware detection software scored higher than did the Covert user interact design where they were considered to be more likely to be used again, more helpful and liked. They also better matched expectations and were considered to be more informative about security issues.

Additional benefits of the Intrusive user interface design when compared to the Covert user interface design included enjoyment, trustworthiness, less frustration and considering it to need less improvement. Similarly, for the Overt user interface design when compared to the Covert user interface design, additional benefits included convenience and user-friendliness.

Although the difference in usability scores between the Overt user interface design and the Intrusive user interface design was not statistically significant, the overall quality scores strongly favoured the Intrusive user interface design, scoring 19.8 on the 30-point scale, significantly higher than either of the alternative user interface designs.

The rank ordering of the three user interface designs, again clearly favoured the Intrusive user interface design, with 41% of participants ranking it best of the three.

Obtaining consistent eye tracking results from all participants was only partially successful (full data for 91 out of a cohort of 168), probably due to some participants looking away from the screen to refer to their login details and then not resuming the correct seating position afterwards. However, relative comparisons of gaze data between designs could still be performed.

The browser toolbar, address bar and status bar were all gazed at far more during the installation task than during the bill payment tasks, perhaps reflecting the increased interaction with the browser and heightened awareness of the surrounding operating system during the installation.

Interestingly, with the Overt user interface design, participants looked at the browser toolbar and address bar more than with the other two user interface designs, possibly due to the lack of information given about the security status bar after login and resulting in participants checking that the URL was still correct.

The pupil dilation data extracted from the gaze data suggest higher levels of cognitive load for participants during the software download and installation task, possibly reflecting higher levels of awareness or anxiety; and relatively lower levels of cognitive load in the bill payment task, notably with the Covert user interface design.

Some 23% of participants did not notice the difference between the three user interface designs (Covert, Overt and Intrusive). After being reminded of the three different user interface designs, the majority of participants were positive about their bank providing extra online security and 25% liked the security status bar, although a few suggested that it provided more detail. Some 20% of participants commented that they disliked the way the Covert design gave no reason for the bill payment failure.

Approximately half of the participants would have either ignored or purposely not installed the security software download. Their main worries appeared to be over who was providing the download, how big it was and how it would interact with their existing software. Only 25% would have installed it without asking any further questions.

5.4. The CodeSure Sender Authentication Experiment

Use by the banks of modern communication channels such as email, Facebook, text messaging, Twitter and automated voice messaging are currently limited by the prevalence of fraudulent messages from criminals which purport to come from the bank. Research shows that some 42% of UK Internet banking users have been the target of fraudulent emails (Ensor, 2005). In an attempt to reassure customers, many banks currently include part of the customer’s postcode in email message, but this has limited effect, and most banks are now telling customers that they will no longer receive any emails from their bank – and if they do then they are likely to be fraudulent.

Given results of the phishing experiment and the Internet security tutorial experiment considered earlier in this chapter it is clear that Internet banking users remain at risk from these types of security threats and that any usable security solution considered by the banks will need to address the issue of customers needing to validate communications from their bank. The usable security solution based around use of the CodeSure card offers a way forward since the CodeSure card supports sender authentication thereby allowing the bank to supply a unique identification code within an email, SMS message or outbound telephone call, which can be decoded by that customer's CodeSure card, and only that card. Each code can only be used once.

The CodeSure card's Verify mode of operation allows the customer to take an 8-digit security code in a message from their bank and key this into their CodeSure card. The 8-digit code will have been constructed by the bank with knowledge of the secret security key (although not the PIN) for that customer's CodeSure card such that the card will analyse the input code and display a VERIFIED or FAILED response to authenticate the sender.

5.4.1. Experiment Design

An experiment was undertaken to investigate usage of the CodeSure card in sender authentication tasks with emails, text messages and automated telephone calls being received by participants. Participants were presented with a bundle of emails, SMS messages and phone calls by an automated voice service and were then asked to evaluate the authenticity of each, initially without the CodeSure card and then with the CodeSure card.

A cohort of 96 Internet banking and telephone banking customers were recruited to take part in the experiment sessions. The cohort was balanced for age group and for gender. Participants were selected as having used Internet banking and telephone banking at some point during the 3 months prior to the experiment. In their experiment session, participants were asked to adopt a persona and told to deal with batches of incoming messages via email, SMS text messaging and via telephone.

To establish a baseline of behaviour for each participant, they first received a batch of 3 emails, 3 text messages and 3 phone calls from their bank without reference to any authentication considerations. The order of the 3 message channels was randomised across participants. Each participant received their 3 emails in sequence, one after the other: similarly for their text messages and phone calls from the bank. The messages included ‘their’ surname or last 3 characters of ‘their’ postal code for re-assurance. After each message the participant was asked to rate the message for authenticity.

Three different message content types were used in the experiment, exhibiting different message urgency (new product announcement, payment due alert and suspicious card use alert) and the order of message type was randomised across channels for each participant.

The participant was then given a letter “from their bank” detailing the use of the CodeSure card in sender authentication for each of the communications channels to help combat phishing and identity fraud. They were also given a tutorial sheet on how to use the Verify mode of the CodeSure card, Figure B.7 in Appendix B.

They were then asked to use their CodeSure card for sender authentication with a second batch of 6 emails, 6 text messages and 6 phone calls from their bank. Half of these messages included valid CodeSure sender codes; the other half included bogus sender codes as if from a fraudster. Again the order of the 3 message channels and the urgency of each message were randomised across the sample.

Participants gave a confidence score for each message: “*On a scale from 0 to 10 where 10 means absolutely confident, how confident are you that that message is genuine and comes from the bank?*”

5.4.2. Implementation Details

As with the phishing emails experiment presented in Section 5.1, the email client used was IE6 running a Webmail client, modified to prevent easy deletion of emails without first reading them. The Webmail client was connected to an hMailServer IMAP server running on a Windows 2003 server, and a dummy Sendmail agent was used to prevent any participants sending out real email replies (participants were not asked to respond to the emails so any emails they sent were ignored). Participants were forced to read the emails in the correct order.

The batch of emails that the participants received consisted of a low urgency message (discount on a loan), medium urgency message (card payment due) and high urgency message (unusual transactions on the account), Figure 5.12.

Greeting: "*Dear Mr/Mrs <**name**> [Postcode ending 5ND]*"

Low urgency: "*As a long-standing customer of <**bank**> we can offer you a significant discount on our range of personal loans. For information contact us online at http://www.<**bank**>.com or on the telephone on 08457 3000000.*"

Medium urgency: "*The minimum payment to your credit card is due. To avoid being charged interest, please make this payment as soon as possible. You can make your payment online at http://www.<**bank**>.com or on the telephone on 08457 3000000.*"

High urgency: "*We have noticed several unusual foreign transactions using your debit card recently. If this might be a concern for you, please check online at http://www.<**bank**>.com or on the telephone on 08457 3000000.*"

[With sender code]: "*You can confirm that this message is genuine by entering <**sender code**> using the keypad on the reverse of your bank card [Verify mode]. You should then see VERIFIED displayed on the card.*"

Figure 5.12: Emails used in Sender Authentication Experiment

The CodeSure sender authentication message was appended to the end of each of the three emails when the CodeSure card was being used. The (valid) 8 digit sender authentication codes were generated by the CodeSure card server.

An IVR telephone banking service similar to the one used in the CodeSure multi-channel experiment considered in Chapter 4 was used, written in **C** using Nuance speech recognition technology and the Nuance Dialog Builder API. When the participant was ready to accept a telephone call, the ‘bank’ dialled the customer and took the participant through the steps required for the task – depending on whether or not a CodeSure card was being used, Figure 5.13.

Greeting: “*Hello, this is an automated call from <**bank**> for Mr/Mrs <**name**> with account number ending <**account**>. If you are Mr/Mrs <**name**> please press ‘1’ to continue…*”

Low urgency: “*As a long-standing customer of <**bank**> we can offer you a significant discount on our range of personal loans. For information contact us online at www.<**bank**>.com or through telephone banking now…*”

Medium urgency: “*The minimum payment to your credit card is due. To avoid being charged interest, please make this payment as soon as possible. You can make your payment online at www.<**bank**>.com or through telephone banking now…*”

High urgency: “*We have noticed several unusual foreign transactions using your debit card recently. If this might be a concern for you, please check online at www.<**bank**>.com or through telephone banking now…*”

[With sender code]: “*You can confirm that this message is genuine by using the Verify mode on the reverse of your bank card. Using the keypad, enter the code <**sender code**>. Here’s that code again <**sender code**>. You should then see VERIFIED displayed on the card. To hear this message again press ‘1’; or to transfer to our telephone banking service, press ‘2’…*”

Figure 5.13: Telephone Messages used in Sender Authentication Experiment

The CodeSure sender authentication message was played after each of the three telephone messages when the CodeSure card was being used. As before, the (valid) 8 digit sender authentication codes were generated by the CodeSure card server.

The SMS service was written in Visual Basic and was run on a computer running Windows Server 2003 connected to a Sony Ericsson K700i mobile phone. The SMS service treated the mobile phone as a modem over which it could communicate and automate the sending of text messages to the three experiment mobile phones. The three types of SMS messages sent were as shown in Figure 5.14. As with the other 2 channels, the SMS service contacted the CodeSure card server to generate (valid) 8 digit sender authentication codes when the CodeSure card was being used.

Low urgency: "*As a customer we can offer you discounts on our personal loans. More info at www.<**bank**>.com or call us on 084573000000.*"

Medium urgency: "*The minimum payment for your account ending <**account**> is due. To avoid interest charges, make a payment online at www.<**bank**>.com or call us on 084573000000.*"

High urgency: "*We have noticed some unusual foreign transactions for your account ending <**account**>. Please check online at www.<**bank**>.com or call us on 084573000000.*"

[With sender code]: "*Confirm this message is genuine by entering <**sender code**> using the keypad on the reverse of your bank card (Verify Mode). You should then see VERIFIED displayed.*"

Figure 5.14: Text Messages used in Sender Authentication Experiment

The details of the experiment are summarised in Table 5.27.

The CodeSure Sender Authentication Experiment	
Experiment purpose	Exploration of using sender authentication with the CodeSure card to verify emails, SMS messages and phone calls from the bank.
Null hypothesis (H_0)	The use of sender authentication to verify the sender of a message has no effect on participants' trust in the message.
Experiment design	Participants were asked to assess the authenticity of 3 emails, 3 SMS messages and 3 telephone calls without sender authentication: and 6 other emails, 6 SMS messages and 6 telephone calls with the CodeSure card used for sender authentication. Repeated measures (randomised order of messages), within-subjects.
Dependent variables	Button click data. Perceived authenticity scores for each message.
Other data	Exit interview data.
Independent variables	Experiment: task order (randomised, balanced). Participant: gender (2 genders, balanced), age group (2 groups, ages 18 to 44 / ages 45 and over).
Cohort	96 (2 age groups × 2 genders × over-sampling ratio 24:1) = 96. Selected from banking customers who have used Internet banking and telephone banking at least once in the preceding 3 months.
Honorarium	Personal cheque for £30.
Session time	60 minutes.

Table 5.27: Experiment Design Details

5.4.3. Results

5.4.3.1. Demographic and Technographic Results

The customer sample consisted of 109 participants, balanced for gender (50% male and 50% female) and for age groups (51% aged 18 to 44 and 49% aged 45 and over). All had used their bank's Internet banking and telephone banking services in the 3 months prior to the experiment.

5.4.3.2. Message Confidence Results

Participants gave a confidence score for each message: "*On a scale from 0 to 10 where 10 means absolutely confident, how confident are you that that message is genuine and comes from the bank?*"

The confidence scores for each type of message (no sender code, valid sender code, bogus sender code) were averaged for each participant and rounded to the closest integer score, Table 5.28.

Confidence Score	No Code	Valid Code	Bogus Code
0	0.0%	0.0%	38.5%
1	0.0%	0.0%	17.4%
2	0.9%	0.9%	10.1%
3	0.9%	0.0%	9.2%
4	2.8%	0.0%	3.7%
5	9.2%	0.9%	7.3%
6	12.8%	1.8%	2.8%
7	26.6%	12.8%	4.6%
8	21.1%	22.0%	4.6%
9	12.8%	28.4%	1.8%
10	12.8%	33.0%	0.0%

Table 5.28: Message Confidence Results

The most common message confidence score without sender code was 7 (26.6% of participants). With a valid CodeSure sender code the most common message confidence score was 10 (33% of participants). With a bogus code the most common confidence score was 0 (38.5% of participants).

The mean and modal message confidence scores by channel are given in Table 5.29 (mean) and Table 5.30 (mode). Table 5.31 shows the frequencies of participants scoring messages with a confidence of 9 or 10.

	SMS	Email	Phone	Overall
No sender code	7.291	7.221	7.547	7.353
Valid CodeSure sender code	8.661	8.698	8.723	8.694
Bogus CodeSure sender code	2.242	2.228	2.248	2.239

Table 5.29: Mean Message Confidence Scores

	SMS	Email	Phone	Overall
No sender code	8	8	10	8
Valid CodeSure sender code	10	10	10	10
Bogus CodeSure sender code	0	0	0	0

Table 5.30: Most Frequent (Modal) Message Confidence Scores

	SMS	Email	Phone	Overall
No sender code	29.1%	34.3%	41.0%	34.8%
Valid CodeSure sender code	67.3%	69.4%	68.8%	68.5%
Bogus CodeSure sender code	3.1%	2.8%	3.4%	3.1%

Table 5.31: Frequency of Message Confidence Scores of 9 or 10

To investigate the statistical significance of the mean message confidence scores of Table 5.29 by channel, a repeated measures ANOVA was run with message type (no sender code, valid CodeSure sender code or bogus CodeSure sender code) and channel (SMS, email or phone) as within-subjects factors and age group and gender as between-subjects independent variables, Table 5.32. Mauchly's test for sphericity was significant for message type (p<0.001), channel (p<0.001) and for the interaction between message type and channel (p<0.001), therefore because the assumption of sphericity had been violated the degrees of freedom were corrected using the Greenhouse-Geisser estimates of sphericity.

Results of the ANOVA test reveal, Table 5.32, that there was a highly significant overall effect, F(1.421,149.230)=326.011, p<0.001, for message type with all pair-wise differences between the messages types being statistically significant, p<0.001. There were no significant interactions between message type and channel and there were no significant between-subjects interactions or effects.

Source	Sum of Squares	*df*	Mean Square	*F*	Sig.
Within-Subjects Effects					
Message	7580.817	1.421	5333.969	326.011	0.000
Message × Gender	0.850	1.421	0.598	0.037	0.919
Message × Age	73.521	1.421	51.730	3.162	0.062
Error (Message)	2441.595	149.230	16.361		
Channel	2.963	1.697	1.746	0.923	0.386
Channel × Gender	8.508	1.697	5.013	2.649	0.082
Channel × Age	0.389	1.697	0.229	0.121	0.854
Error (Channel)	337.201	178.210	1.892		
Between-Subjects Effects					
Gender	11.936	1	11.936	1.124	0.292
Age	3.666	1	3.666	0.345	0.558
Gender × Age	2.214	1	2.214	0.208	0.649
Error	1115.171	105	10.621		

Table 5.32: ANOVA Results for Mean Message Confidence Scores by Channel

These results confirm that the use of the CodeSure card with sender codes is highly effective. Message confidence scores rose from 7.35 without sender codes to 8.69 with valid sender codes checked by means of the CodeSure card and also fell to 2.24 with bogus codes identified by use of the CodeSure card.

5.4.3.3. Interview Comments

When asked to identify the benefits to them of using the CodeSure card to verify messages, participants responded: "*It made me feel more confident in the messages I received.*", "*It could stop fraud. I would know if messages were fraudulent or genuine.*"

Some 75% of participants felt that banks should include a sender code in all messages: exclusions suggested were messages about sales and those that were not related to a bank account: 40% of participants expressed a preference for a 4-digit sender code, 43% for a 6-digit code and 16% for an 8-digit code.

5.4.4. Discussion on Experiment Results

The null hypothesis (H_0) for this experiment was:

> *"The use of sender authentication to verify the sender of a message has no effect on participants' trust in the message."*

The evidence presented here is sufficient to refute this null hypothesis: a valid CodeSure sender code was given a mean message confidence score of 8.7; a bogus CodeSure sender code was given a mean message confidence score of 2.2; and a message with no sender code was given a mean message confidence score of 7.4. The (pair-wise differences) were each statistically significant. Sender authentication using embedded CodeSure card codes in emails, SMS and phone calls from the Bank therefore boosts customer confidence in the validity of messages: the modal score without sender codes of 7 (26.6% of participants) rose to a modal score of 10 (33.0% of participants) with a valid CodeSure sender code and fell to 0 (38.5% of participants) with a bogus sender code.

5.5. Summary

These results of these experiments reveal several important findings in the field of online security, showing that whilst users claim to be aware of the general threats of phishing and malware, their claims are possibly misplaced since their behaviour in the experiments failed to indicate any resilience in avoiding phishing.

For the phishing threat, changes to the style of link employed in the 'lure' between text hyperlinks, button hyperlinks and embedded HTML forms had no effect on users' propensity to click on the link. However, heightening the apparent urgency in the message content of the email did have an effect. Educating users in phishing and online security significantly reduces their likelihood of being duped by a phishing email, and such education offers possibly the sole route available to banks to reduce identity theft and its associated costs, although care must also be taken in selecting the correct approach to educating customers (Desman, 2003).

The malware detection software experiment highlighted that users dislike having security information hidden from them, especially when it prevents them completing the task they are attempting to perform. If their bank were to offer them security software to detect malware on their machine, users would therefore prefer that it kept them informed about threats rather than hiding it from them and secretly communicating the scan results back to the bank. Although users would be more likely to be anxious during the installation of such software, that is an indicator of increased awareness, and a little education on the use of security certificates used in the installation of downloaded software could help to alleviate concerns.

For the (50) participants who completed the download without intervention, 95.6% of those who had not been given the tutorial on Internet security performed the expected installation compared to 85.7% of those who had been given the tutorial: evidence of a only a slightly higher awareness of Internet security issues for those who had been given the phishing training.

The CodeSure card is again shown to be a potential solution for banks to reduce the threat of phishing (and vishing) through the use of sender authentication procedures. Users have been shown here to be much more likely to be confident in the validity of communications from their bank, resulting in a lower success rate for phishing and could help bring back customer confidence in the bank using such channels to communicate with their customers.

Chapter 6. Discussion and Conclusions

The research reported here has centred on a series of 8 large-scale usability experiments investigating the role of two-factor authentication with the CodeSure card as the basis for a usable security solution for banking and eCommerce. The premise behind the concept of usable security is that by furnishing customers with a familiar device, a familiar form factor, to be used in a familiar, common modality for all purchases, financial transactions and account access in the physical world as well as in the digital world, this familiarity will demolish the existing barriers where channel security procedures are seen by customers as being an imposition, thereby achieving improved levels of actual security because security procedures with usable security will become a natural part of everyday financial activity. The results presented here have shown that the CodeSure card and procedures based on the card offer the basis of such a usable security option.

6.1. Discussion

Usability scores for the CodeSure card at 4.49 (on a 7-point scale) for first use were shown to rise to 4.66 on tenth use, confirming the device usability as suited for use in a usable security procedure. An enhanced card design, featuring improved mode selection and navigation cues and importantly, a keypad buffer to allow key-ahead, served to further boost usability scores for the CodeSure card to 5.01 with improved usage times and improved task times and success rates.

Participants recognised that the CodeSure card was convenient to use as reflected in the quality ratings and comments; and appreciated the role of the card as a multi-channel, usable security solution where 82% expected to be able to use it to access their bank's Internet banking service, 73% to use it for telephone banking and 88% to use it for online shopping. Some participants would even expect to use direct channels more often when using the CodeSure card.

The experiment results suggest that users are aware of a trade-off between the security of a procedure and its convenience. Some 66% of participants would prefer a common security procedure across all banking and eCommerce channels, which is encouraging for the future of the CodeSure card. Only 28% would prefer the security to remain the same in the banking channels they use, but that is perhaps due to being familiar with existing procedures rather than thinking about the benefits a common security procedure would bring. Although the CodeSure card scored lower than today's disparate security processes in terms of usability, the CodeSure card was still usable with mean usability scores above 5.0, and the main challenge for the bank would be to break users out of their comfort zone in using security processes that they are familiar with and move them over to the common security processes of the CodeSure card.

Experiment results also reveal several important findings in the field of online security, showing that whilst users claim to be aware of the general threats of phishing and malware, their claims are possibly misplaced since their behaviour in the experiments failed to indicate any resilience in avoiding phishing.

In terms of the phishing threat, changes to the style of link employed in the 'lure' between text hyperlinks, button hyperlinks and embedded HTML forms had no effect on users' propensity to click on the link. However, heightening the apparent urgency in the message content of the email did have an effect. Educating users in phishing and online security can reduce their likelihood of being duped by a phishing email, and such education offers possibly the sole route available to banks to reduce identity theft and its associated costs, although care must also be taken in selecting the correct approach to educating customers. However, experiment data actually confirm the limitations of reliance on user education since only 12% of participants who studied an on-line security tutorial on email and web site security were actually able to detect all of the fraudulent web sites presented to them in the experiment.

The experiment results also highlighted that users dislike having security information hidden from them, especially when it prevents them completing the task they are attempting to perform. If their bank were to offer them security software to detect malware on their machine, users would therefore prefer that it kept them informed about threats.

The CodeSure card is also shown to be a potential solution for banks to reduce the threat of phishing (and vishing) through the use of sender authentication procedures and users are much more likely to be confident in the validity of communications from their bank, resulting in a lower success rate for phishing and could help bring back customer confidence in the bank using such channels to communicate with their customers. The sender authentication mode of the CodeSure card proved to be a successful device in boosting customer trust in messages sent by the bank. This offers hope to banks in being able to re-establish trust in such channels and effectively use them as a means of communicating to their customers again. It also reduces the threat posed by phishing (and vishing) to the bank and offers customers a straightforward way to check the authenticity of messages without necessarily having been educated in Internet security.

6.2. Limitations and Practical Issues

Creating true-to-life prototypes of banking systems for these experiments was key to making the experience seem as real as possible for participants, thus making it possible to extrapolate experimental results into real world usage. Although still located in a usability laboratory setting, participants were immersed in the tasks and even when gaze tracking was being used it was unobtrusive and did not impede participants in their assigned tasks.

Several limitations on the use of gaze tracking cameras in such large-scale usability work were encountered leaving their use as such in question. For example, in the CodeSure Keypad and Display Synchrony experiment (Section 3.3) the gaze tracking cameras may sometimes have been blocked from tracking a participant's gaze if they inadvertently moved their arm to obscure the line of sight between the camera and their eyes. The experiment design aimed to minimise this by suitably adjusting the position of the CodeSure card on the screen depending on if it was being used by a left or right-handed person, but the limitation was still an issue. Also, calibration of the gaze tracker camera was only performed for a participant at the start of each batch of tasks and participants may have been seated differently between tasks, resulting in the gaze tracker camera finding it difficult to locate their eyes and therefore generate accurate gaze data, mandating the need for periodic re-calibration of the camera, which is at odds with attempts to portray the experiment scenarios as being 'normal'.

The CodeSure multi-channel usability experiment (Chapter 4) attempted to compare the usage of the CodeSure card with today's disparate authentication methods, resulting in a bias towards today's methods most likely due to familiarity. To counter this, the experiment could perhaps have been combined with a longitudinal study that measured attitudes towards usability over a prolonged period of time, perhaps asking participants to repeat the experiment a week later in a similar way to the CodeSure card baseline usability experiment (Section 3.1).

A limitation of the Phishing Emails experiment (Section 5.1) was that participants were constrained to using the supplied Webmail client to read and process the emails they were given. However, although participants were primed on the use of the Webmail client before the experiment, many would have been more familiar using other email clients or devices. As part of experiment procedure, participants were also restricted to reading the emails in the order they were given and had to open them before being allowed to delete them, a constraint that many participants would not have followed in real life. This highlights a general point to be made about all of the experiments presented in this research: whilst much effort was made to make experiment scenarios as realistic as possible, the experiments were still run within a laboratory setting with personas assigned to participants.

6.3. Ideas for Further Work

Whilst the experiments presented here investigated the use of the CodeSure card in a wide selection of banking channels, use of the card in any of the physical channels such as ATM or counter, where the CodeSure card can be used as a normal debit or credit card, was not examined. Although there is no difference in usage between using a CodeSure card and a normal bank card at an ATM, it was never entirely clear that participants in this research were mindful, or perhaps even aware of, the fact that the CodeSure card could still be used as a normal bank card at an ATM or at a POS terminal in a shop. This would serve as an interesting new thread of research investigation.

The aspect of the CodeSure card which drew the most negative criticism was the sensitivity of the keypad buttons. It would be interesting to investigate in a further usability experiment participants' attitude to the usability of the CodeSure card with different button technologies, such as capacitive touch sensitivity as used in most touch screen phones. Whilst the CodeSure card might have a better response to button presses with capacitive technology, there could also be a negative reaction to over-sensitivity of the buttons. An experiment such as this would provide insight into the extent to which negative reaction to the sensitivity of the buttons dominated the overall usability ratings for the CodeSure card.

With the growing use of mobile phones to access Internet banking, many banks offer an alternative mobile phone banking service which provides a more limited feature set compared to Internet banking yet is more suited for use with the smaller form factor of a mobile phone. Whilst not investigated in the experiments presented here, the reduced authentication that is typical for this banking channel would be interesting to investigate with the CodeSure card as an alternative to a predefined passcode.

Finally, further work to investigate how best to generate and analyse pupil dilation data (as derived in the Malware Detection Software usability experiment reported here) as measured whilst participants were busy completing practical tasks would be of value. Whilst the results shown for that experiment showed slight differences, any future research might benefit from using timestamps for key events during the tasks so that a more detailed representation of pupil dilation could be obtained for such key points. This might lead to a better understanding of where users become more anxious and might help to influence the design of a better security procedure.

6.4. Conclusions

In all of the experiments presented here the CodeSure card was rated acceptably high in terms of mean usability. The usability and convenience of the CodeSure card were rated at odds with security, reflecting the belief that more complicated procedures are more secure, yet also not fully appreciating the need for more secure procedures to protect themselves from eCrime. Perhaps this need will become more apparent in years to come as more users become aware of security issues and usable security solutions are deployed.

Overall, the research reported here is offered in support of the thesis that a usable security solution predicated on code-based multi-factor authentication will result in tangible improvements to actual security levels in banking and eCommerce services, and that the CodeSure card as described here can form the basis of such a usable security solution.

References

Ackerman, M. S. and Mainwaring, S. D. (2005), ‘Privacy issues and human-computer interaction’, in L. F. Cranor and S. Garfinkel (Editors), *Security and Usability*, O’Reilly, Sebastopol, CA, USA, Chapter 19, pp. 381-399.

Adams, A. and Sasse, M. A. (2005), ‘Users are not the enemy: Why users compromise security mechanisms and how to take remedial measures’, in L. F. Cranor and S. Garfinkel (Editors), *Security and Usability*, O’Reilly, Sebastopol, CA, USA, Chapter 32, pp. 639-649.

Augier, M. (2007), ‘Passwords’, in M. Jakobsson and S. Myers (Editors), *Phishing and Countermeasures*, Wiley, Hoboken, NJ, USA, Chapter 8, pp. 277-308.

Bardzell, J., Blevis, E. and Lim, Y.-K. (2007), ‘Human-centered design considerations’, in M. Jakobsson and S. Myers (Editors), *Phishing and Countermeasures*, Wiley, Hoboken, NJ, USA, Chapter 7, pp. 241-275.

Barton, B. F. and Barton, M. S. (1984), ‘User-friendly password methods for computer-mediated information systems’, *Computers and Security*, Volume 3(3), pp.186-195.

Beaumier, C. M. (2006), ‘Multifactor authentication: A blow to identity theft?’, *Bank Accounting and Finance*, February/March 2006, pp. 33-37.

Bennett, M. (2004), *Online banking customer authentication systems: PIN/TAN is not immune from phishing*, Technical Report, Forrester Research, Cambridge, MA.

Besnard, D. and Arief, B. (2004), ‘Computer security impaired by legitimate users’, *Computers and Security*, Volume 23(3), pp. 253-264.

Bicakci, K. and van Oorschot, P. C. (2011), ‘A multi-word password proposal (Gridword) and exploring questions about science in security research and usable security evaluation’, in *Proceedings of the New Security Paradigms Workshop (NSPW’11)*, ACM, New York, NY, USA, pp. 25-36.

Bishop, M. (2005), ‘Psychological acceptability revisited’, in L. F. Cranor and S. Garfinkel (Editors), *Security and Usability*, O’Reilly, Sebastopol, CA, USA, Chapter 1, pp. 1-11.

Bonneau, J., Preibusch, S. and Anderson, R. (2012), ‘A birthday present every eleven wallets? The security of customer-chosen banking PINs’, in *Proceedings of the 16th International Conference on Financial Cryptography*, pp. 25-40.

Braz, C. and Robert, J. (2006), ‘Security and usability: the case of the user authentication methods’, in *Proceedings of the 18th International Conference of* *the Association Francophone D'Interaction Homme-Machine (IHM'06)*, Volume 133, ACM Press, New York, pp. 199-203.

Brostoff, S. and Sasse, M. A. (2000), 'Are passfaces more usable than passwords? A field trial investigation', in *Proceedings of HCI 2000*, Sunderland, UK, pp. 405-424.

Brunk, B. (2005), 'A user-centric privacy space framework', in L. F. Cranor and S. Garfinkel (Editors), *Security and Usability*, O'Reilly, Sebastopol, CA, USA, Chapter 20, pp. 401-420.

Centeno, C. (2004), 'Adoption of Internet services in the acceding and candidate countries, lessons from the Internet banking case', *Telematics and Informatics*, Volume 21, pp. 293-315.

Christin, N., Egelman, S., Vidas, T. and Grossklags, J. (2011), 'It's all about the Benjamins: An empirical study on incentivizing users to ignore security advice', in *Proceedings of the 15th International Conference on Financial Cryptography*, pp. 16-30.

Claessens, J., Dem, V., Cock, D. D., Preneel, B. and Vandewalle, J. (2002), 'On the security of today's online electronic banking systems', *Computers and Security*, Volume 21(3), pp. 253-265.

Coolican, H. (1990), *Research Methods and Statistics in Psychology*, Hodder & Stoughton, London.

Coventry, L., Angeli, A. D. and Johnson, G. (2003), 'Usability and biometric verification at the ATM interface', in *Proceedings of SIGCHI on Human Factors in Computer Systems*, ACM Press, pp. 153-160.

Coventry, L. (2005), 'Usable biometrics', in L. F. Cranor and S. Garfinkel (Editors), *Security and Usability*, O'Reilly, Sebastopol, CA, USA, Chapter 10, pp. 175-197.

Davidson, N., Mclnnes, F. R. and Jack, M. A. (2004), 'Usability of dialogue design strategies for automated surname capture', *Speech Communication*, Volume 43(2), pp. 55-70.

Desman, M. (2003), 'The ten commandments of information security awareness training', *Security Management Practices*, Volume 11(6), pp. 39-44.

Dhamija, R., Tygar, J. D. and Hearst, M. (2006), 'Why phishing works', in *Proceedings of SIGCHI on Human Factors in Computing Systems*, ACM, Montreal, Canada. pp. 581-590.

Dourish, P. and Redmiles, D. (2002), 'An approach to usable security based on event monitoring and visualization', in *Proceedings of the New Security Paradigms Workshop (NSPW'02)*, ACM Press, Virginia Beach, Virginia, pp. 75-81.

Du, W., Jayaraman, K., Tan, X., Luo, T. and Chapin, S. (2011), 'Why are there so many vulnerabilities in web applications?', in *Proceedings of the New Security Paradigms Workshop (NSPW'11)*, Marin County, CA, USA, pp. 83-93.

Duchowski, A. T. (2007), *Eye Tracking Methodology*, 2nd Edition, Springer-Verlag, London.

Dutton, R. T., Foster, J. C., Jack, M. A. and Stentiford, F. W. M. (1993), 'Identifying usability attributes of automated telephone services', in *Proceedings of EUROSPEECH'93*, pp. 1335-1338.

Emigh, A. (2007), 'Phishing attacks: Information flow and chokepoints', in M. Jakobsson and S. Myers (Editors), *Phishing and Countermeasures*, Wiley, Hoboken, NJ, USA, Chapter 2, pp. 31-63.

Ensor, B. (2005), *What UK net users think about phishing (and what banks should do about it)*, Technical Report, Forrester Research, Cambridge, MA.

Falk, L., Prakash, A., Borders, K. (2008), 'Analyzing websites for user-visible security design flaws', in *Proceedings of the Fourth Symposium on Usable Privacy and Security (SOUPS'08)*, ACM, New York, pp. 117-126.

Faulkner, X. (2000), *Usability Engineering*, Macmillan Press, London, UK.

FFIEC (2005), *Authentication in an Internet Banking Environment*. Retrieved from http://www.ffiec.gov/pdf/authentication_guidance.pdf

Field, A. (2009), *Discovering Statistics Using SPSS*, 2nd Edition, SAGE Publications, London.

Florêncio, D. and Herley, C. (2010), 'Where do security policies come from?', in *Proceedings of the Sixth Symposium on Usable Privacy and Security (SOUPS'10)*, ACM, New York, pp. 10:1-10:14.

Forget, A., Chiasson, S., van Oorschot, P. C. and Biddle, R. (2008), 'Improving text passwords through persuasion', in *Proceedings of the 4th Symposium on Usable Privacy and Security (SOUPS'08)*, ACM, New York, pp. 1-12.

Foster, J. C., Mclnnes, F. R., Jack, M. A., Love, S., Dutton, R. T., Nairn, I. A. and White, L. S. (1998), 'An experimental evaluation of preferences for data entry method in automated telephone services', *Behaviour and Information Technology*, Volume 17(2), pp. 82-92.

Fraser, N. (2007), *Passwords 2.0 (white paper)*, Technical Report, Tricerion Group, London, UK, pp. 1-9.

Friedman, B., Hurley, D., Howe, D. C., Felten, E. and Nissenbaum, H. (2002), 'Users' conceptions of web security: A comparative study', *Proceedings of CHI 2002*, pp. 747-747.

Fu, A. Y., Deng, X. and Wenyin, L. (2007), 'Homograph attacks using Unicode', in M. Jakobsson and S. Myers (Editors), *Phishing and Countermeasures*, Wiley, Hoboken, NJ, USA, Chapter 3, pp. 81-89.

Furnell, S. and Clarke, N. (2005), 'Biometrics: No silver bullets', *Computer Fraud and Security*, August 2005. pp. 9-14.

Furnell, S. (2007), ‘A comparison of website user authentication mechanisms’, *Computer Fraud and Security*, Volume 2007(9), pp. 5-9.

Garfinkel, S. and Spafford, G. (2002), *Web Security, Privacy and Commerce*, 2nd Edition, O’Reilly, Sebastopol, CA, USA.

Garfinkel, S. (2005), ‘Sanitization and usability’, in L. F. Cranor and S. Garfinkel (Editors), *Security and Usability*, O’Reilly, Sebastopol, CA, USA, Chapter 15, pp. 293-317.

Gaw, S. and Felten, E. W. (2006), ‘Password management strategies for online accounts’, in *Proceedings of the Second Symposium on Usable Privacy and Security (SOUPS’06)*, ACM, New York, pp. 44-55.

Graeber, C. (2004), *Phishing concerns impact consumer online financial behaviour*, Technical Report, Forrester Research, Cambridge, MA.

Griffith, V. (2007), ‘Case study: Automated trawling for public private data’, in M. Jakobsson and S. Myers (Editors), *Phishing and Countermeasures*, Wiley, Hoboken, NJ, USA, Chapter 6, pp. 191-202.

Gross, R. (2001), *Psychology: The Science of Mind and Behaviour*, 4th Edition, Hodder & Stoughton, London.

Gunson, N., Marshall, D., Morton, H. and Jack, M. (2011), ‘User perceptions of security and usability of single-factor and two-factor authentication in automated telephone banking’, *Computers and Security*, Volume 30(4), pp. 208-220.

Gupta, M. (2007a), ‘Pharming and client side attacks’, in M. Jakobsson and S. Myers (Editors), *Phishing and Countermeasures*, Wiley, Hoboken, NJ, USA, Chapter 4, pp. 105-126.

Gupta, M. (2007b), ‘Spoofing and countermeasures’, in M. Jakobsson and S. Myers (Editors), *Phishing and Countermeasures*, Wiley, Hoboken, NJ, USA, Chapter 3, pp. 65-81.

Halderman, J. A., Waters, B. and Felten, E. W. (2005), ‘A convenient method for securely managing passwords’, in *Proceedings of the 14th International Conference on the World Wide Web (WWW’05)*, ACM Press, New York, pp. 471-479.

Hartson, R. (1998), ‘Human-computer interaction: Interdisciplinary roots and trends’, *Systems and Software*, Volume. 43, pp. 103-118.

Hayashi, E., Dhamija, R., Christin, N. and Perrig, A. (2008), ‘Use your illusion: Secure authentication usable anywhere’, in *Proceedings of the 4th Symposium on Usable Privacy and Security (SOUPS’08)*, ACM, New York, pp. 35-45.

Henry, P. A. (2006), ‘Two-factor authentication: A look behind the headlines’, *Network Security*, Volume 2006(4), pp. 18-19.

Hertzum, M., Jørgensen, N. and Nørgaard, M. (2004), ‘Usable security and e-banking: Ease of use vis-a-vis security’, *Australasian Journal of Information Systems*, Volume. 11, pp. 52-65.

Hiltgen, A., Kramp, T. and Weigold, T. (2006), 'Secure Internet banking authentication', *IEEE Security and Privacy*, Volume 4(2). pp. 21-29.

Hornbaek, K. (2006), 'Current practice in measuring usability: Challenges to usability studies and research', *International Journal of Human-Computer Studies*, Volume 64(2), pp. 79-102.

Howell, W. (1985), 'Engineering psychology', in E. M. Altmaier and M. E. Meyer (Editors), *Applied Specialties in Psychology*, Random House, Lawrence Erlbaum Associates, New Jersey, Chapter 10.

International Organisation for Standardization (1998), *Ergonomics of Human System Interaction Part 11: Guidance on Usability*. ISO 9241-11.

Ives, B., Walsh, K. R. and Schneider, H. (2004), 'The domino effect of password reuse', *Communications of the ACM*, New York, USA, pp. 75-78.

Jack, M. A., Foster, J. C. and Stentiford, F. W. M. (1993), 'Usability analysis of intelligent dialogues for automated telephone services', in *Proceedings of the Joint ESCA/NATO Workshop on Applications of Speech Technology*, pp. 149-152.

Jagatic, T. and Johnson, N. A. (2007), 'Case study: Using your social network against you', in M. Jakobsson and S. Myers (Editors), *Phishing and Countermeasures*, Wiley, Hoboken, NJ, USA, Chapter 6, pp. 202-210.

Jagatic, T., Johnson, N. A., Jakobsson, M. and Menczer, F. (2007), 'Social phishing', *Communications of the ACM*, ACM, New York, USA, Volume 50(10), pp. 94-100.

Jahankhani, H., Ali, L. and Jahankhani, H. (2010), 'Security and e-accessibility of online banking', in H. Jahankhani, D. Lilburn Watson, G. Me and F. Leonhardt (Editors), *Handbook of Electronic Security and Digital Forensics*, World Scientific Publishing, Singapore, Chapter 9, pp. 155-167.

Jakobsson, M. (2007), 'Adding context to phishing attacks: Spear phishing', in M. Jakobsson and S. Myers (Editors), *Phishing and Countermeasures*, Wiley, Hoboken, NJ, USA, Chapter 6, pp. 175-191.

Jakobsson, M. and Juels, A. (2009), 'Server-side detection of malware infection', in *Proceedings of the New Security Paradigms Workshop (NSPW'09)*, ACM, New York, NY, USA. pp. 11-22.

James, L. (2005), *Phishing Exposed*, Syngress Publishing, MA.

Johnson, M. E. and Willey, N. D. (2011), 'Usability failures and healthcare data hemorrhages', *IEEE Security and Privacy*, Volume 9, pp. 35-42.

Johnston, J., Eloff, J. H. P. and Labuschagne, L. (2003), 'Security and human computer interfaces', *Computers and Security*, Volume 22(8), pp. 675-684.

Jowitt, T. (2011), 'Lessons in security', *Chartered Banker*, February/March 2011, pp. 18-19.

Just, M. (2005), ‘Designing authentication systems with challenge questions’, in L. F. Cranor and S. Garfinkel (Editors), *Security and Usability*, O’Reilly, Sebastopol, CA, USA, Chapter 8, pp. 143-155.

Just, M. and Aspinall, D. (2009), ‘Personal choice and challenge questions: A security and usability assessment’, in *Proceedings of the 5th Symposium on Usable Privacy and Security (SOUPS’09)*, ACM, New York, pp. 8:1-8:11.

Kadirire, J. (2010), ‘Online transactions’ security’, in H. Jahankhani, D. Lilburn Watson, G. Me and F. Leonhardt (Editors), *Handbook of Electronic Security and Digital Forensics*, World Scientific Publishing, Singapore, Chapter 8, pp. 133-153.

Kainda, R., Flechais, I. and Roscoe, A. W. (2010), ‘Security and usability: Analysis and evaluation’, in *Proceedings of the 5th International Conference on Availability, Reliability and Security*, Krakow, Poland, pp. 275-282.

Karat, C.-M., Brodie, C. and Karat, J. (2005), ‘Usability design and evaluation for privacy and security solutions’, in L. F. Cranor and S. Garfinkel (Editors), *Security and Usability*, O’Reilly, Sebastopol, CA, USA, Chapter 4, pp. 47-74.

Karat, J. (1998), ‘Software evaluation methodologies’, in M. Helander (Editor), *Handbook of Human Computer Interaction*, Elsevier, Amsterdam, North-Holland, pp. 891-903.

Karlof, C. K. (2009), *Human Factors in Web Authentication (Doctoral dissertation)*, University of California, Berkeley. Retrieved from http://www.eecs.berkeley.edu/Pubs/TechRpts/2009/EECS-2009-26.pdf

Kim, J. O. (1975), ‘Multivariate analysis of ordinal variables’, *American Journal of Sociology*, Volume 81(2), pp. 261-298.

King, D. (2012), ‘Chip-and-PIN: Success and challenges in reducing fraud’, *The Innovator*, March 2012, pp. 31-54.

Klingner, J., Kumar, R. and Hanrahan, P. (2008), ‘Measuring the task-evoked pupillary response with a remote eye tracker’, in *Proceedings of Eye Tracking Research and Applications (ETRA’08)*, ACM, New York, pp. 69-72.

Kluever, K. A. and Zanibbi, R. (2009), ‘Balancing usability and security in a video CAPTCHA’, in *Proceedings of the 5th Symposium on Usable Privacy and Security (SOUPS’09),* ACM, New York, pp. 14:1-14:11.

Kumaraguru, P., Cranshaw, J., Acquisti, A., Cranor, L., Hong, J., Blair, M. A. and Pham, T. (2009), ‘School of phish: A real-world evaluation of anti-phishing training’, in *Proceedings of the 5th Symposium on Usable Privacy and Security (SOUPS’09)*, ACM, New York, pp. 3:1-3:12.

Kuo, C., Romanosky, S. and Cranor, L. F. (2006), ‘Human selection of mnemonic phrase-based passwords’, in *Proceedings of the Second Symposium on Usable Privacy and Security (SOUPS’06)*, ACM, New York, pp. 67-78.

Labovitz, S. (1970), ‘The assignment of numbers to rank order categories’, *American Sociological Review*, Volume 35, pp. 515-524.

LaLomia, M. J. and Sidowski, J. B. (1990), ‘Measurements of computer satisfaction, literacy, and aptitudes: A review’, *International Journal of Human-Computer Interaction*, Volume 2(3), pp. 231-253.

Landauer, T. K. (1998), ‘Research methods in human computer interaction’, in M. Helander, ed., *Handbook of Human Computer Interaction*, Elsevier, Amsterdam, North-Holland, pp. 905-928.

Larsen, L. B. (1999), ‘Combining objective and subjective data in evaluation of spoken dialogues’, in *Proceedings of the ESCA Workshop on Interactive Dialogue in Multi-Modal Systems*, Irsee, Germany, pp. 89-92.

Larsen, L. B. (2003), ‘Assessment of spoken dialogue system usability – what are we really measuring?’, in *Proceedings of EUROSPEECH’03*, pp. 1945-1948.

Lewis, J. R. (2001), ‘Current issues in usability evaluation’, *International Journal of Human-Computer Interaction*, Volume 13(4), pp. 343-349.

Lichtenstein, S. and Williamson, K. (2006), ‘Understanding consumer adoption of Internet banking: An interpretive study in the Australian banking context’, *Electronic Commerce Research*, Volume 7(2), pp. 50-66.

Likert, R. (1932), ‘A technique for the measurement of attitudes’, *Archives of Psychology*, Volume 140, pp. 1-55.

Lininger, R. and Vines, R. D. (2005), *Phishing: Cutting the Identity Theft Line*, Wiley, Indianapolis.

Liu, W., Deng, X., Huang, G. and Fu, A. Y. (2006), ‘An antiphishing strategy based on visual similarity assessment’, *Computer Security*, Volume 10(2), pp. 58-65.

Lopez, M. (2005), *Phishing spreads among consumers (many are oblivious, multiple lines of defense are likely the answer)*, Technical Report, September 2005, Forrester Research, Cambridge, MA, pp. 1-4.

Lopez, M. (2006), *Online teens are not immune to phishing (but many are willing to take precautions to protect themselves)*, Technical Report, April 2006, Forrester Research, Cambridge, MA, pp. 1-5.

Love, S., Dutton, R. T., Foster, J. C., Jack, M. A., Nairn, I. A., Vergeynst, N. A. and Stentiford, F. W. M. (1992), ‘Towards a usability measure for automated telephone services’, in *Proceedings of the Institute of Acoustics, Speech and Hearing Workshop*, Volume 14, pp. 553-559.

Manna, M. and van Oorschot, P. C. (2007), ‘Security and usability: The gap in real-world online banking’, in *Proceedings of the New Security Paradigms Workshop (NSPW’07)*, ACM, New York, NY, USA, pp. 1-14.

Martilla, J. A. and Garvey, D. (1975), ‘Four subtle sins of marketing research’, *Journal of Marketing*, Volume 39(1), pp. 8-15.

Menczer, F. (2007), ‘Case study: Using the autofill feature in phishing’, in M. Jakobsson and S. Myers (Editors), *Phishing and Countermeasures*, Wiley, Hoboken, NJ, USA, Chapter 6, pp. 219-221.

Miller, R. C. and Wu, M. (2005), ‘Fighting phishing at the user interface’, in L. F. Cranor and S. Garfinkel (Editors), *Security and Usability*, O’Reilly, Sebastopol, CA, USA, Chapter 14, pp. 275-292.

Monrose, F. and Reiter, M. K. (2005), ‘Graphical passwords’, in L. F. Cranor and S. Garfinkel (Editors), *Security and Usability*, O’Reilly, Sebastopol, CA, USA, Chapter 9, pp. 157-174.

Moody, J. (2004), ‘Public perceptions of biometric devices: The effects of misinformation on acceptance and use’, *Issues in Informing Science and Information Technology,* Volume 1, pp. 753-761.

Morton, H., McBreen, H. M. and Jack, M. A. (2004), ‘Experimental evaluation of the use of embodied conversational agents in ecommerce applications’, in Z. Ruttkay and C. Pelachaud (Editors), *From Brows to Trust: Evaluating Embodied Conversational Agents*, Kluwer, pp. 592-599.

Munshi, J. (1990), *A method for constructing Likert scales*, Technical Report, Sonoma State University. Retrieved from http://www.munshi.4t.com/papers/likert.html

Myers, S. (2007a), ‘Introduction to phishing’, in M. Jakobsson and S. Myers (Editors), *Phishing and Countermeasures*, Wiley, Hoboken, NJ, USA, Chapter 1, pp. 1-29.

Myers, S. (2007b), ‘Status quo security tools’, in M. Jakobsson and S. Myers (Editors), *Phishing and Countermeasures*, Wiley, Hoboken, NJ, USA, Chapter 5, pp. 139-159.

Nielsen, J. and Molich, R. (1990), ‘Heuristic evaluation of user interfaces’, in *Proceedings of Human Factors in Computer Systems CHI’90*, Seattle, WA, pp. 249-256.

Nielsen, J. (1993), *Usability Engineering*, Academic Press, USA.

Nilsson, M., Adams, A. and Herd, S. (2005), ‘Building security and trust in online banking’, in *Proceedings of CHI’05 Extended Abstracts on Human Factors in Computing Systems*, pp. 1701-1704.

Norman, D. (1988), *The Design of Everyday Things*, Basic Books.

O’Brien, R. M. (1979), ‘The use of Pearson’s r with ordinal data’, *American Sociological Review*, Volume 44, pp. 851-857.

O’Gorman, L. (2003), ‘Comparing passwords, tokens and biometrics for user authentication’, in *Proceedings of the IEEE*, Volume 91, pp. 2021-2040.

Ollmann, G. (2004), *The phishing guide (part 2): Understanding and preventing phishing attacks (white paper)*, Technical Report, Next Generation Security Software. Retrieved from http://www.technicalinfo.net/papers/Phishing2.html

Oppenheim, A. N. (1992), *Questionnaire Design, Interviewing and Attitude Measurement*, Pinter, London.

Patrick, A. S., Briggs, P. and Marsh, S. (2005), ‘Designing systems that people will trust’, in L. F. Cranor and S. Garfinkel (Editors), *Security and Usability*, O’Reilly, Sebastopol, CA, USA, Chapter 5, pp. 75-99.

Penn, J. (2004), *Defending against phishing attacks*, Technical Report, Forrester Research, Cambridge, MA.

Piazzalunga, U., Salvaneschi, P. and Coffetti, P. (2005), ‘The usability of security devices’, in L. F. Cranor and S. Garfinkel (Editors), *Security and Usability*, O’Reilly, Sebastopol, CA, USA, Chapter 12, pp. 221-242.

Piazzalunga, U. and Salvaneschi, P. (2006), ‘How to test the usability of security sensitive systems’, *Software Quality Professional*, Volume 8(3), pp. 34-42.

Polson, P. G., Rieman, J., Wharton, C. and Olson, J. (1992), ‘Usability inspection methods: Rationale and examples’, in *Proceedings of the 8th Symposium on Human Interface*, Kawasaki, Japan, pp. 377-384.

Preece, J., Rogers, Y. and Sharp, H. (2002), *Interaction Design: Beyond Human-Computer Interaction*, John Wiley & Sons, New York.

Quesenbery, W. (2003), ‘Dimensions of usability: Defining the conversation, driving the process’, in *Proceedings of the UPA 2003 Conference*, Lawrence Erlbaum Associates, Scottsdale, Arizona, USA.

Rabkin, A. (2008), ‘Personal knowledge questions for fallback authentication: security questions in the era of Facebook’, in *Proceedings of the 4th Symposium on Usable Privacy and Security (SOUPS’08)*, ACM, New York, pp. 13-23.

Ranger, S. (2005), *Banks must boost security to drive online banking*. Retrieved from http://www.zdnet.com/banks-must-boost-security-to-drive-online-banking-3040146904

Raskin, A. (2007), ‘Simulated browser attack’, in M. Jakobsson and S. Myers (Editors), *Phishing and Countermeasures*, Wiley, Hoboken, NJ, USA, Chapter 3, pp. 89-101.

Reavley, N. (2005), ‘Securing online banking’, *Card Technology Today*, Volume 17(10), pp. 12-13.

Renaud, K. (2005), ‘Evaluating authentication mechanisms’, in L. F. Cranor and S. Garfinkel (Editors), *Security and Usability*, O’Reilly, Sebastopol, CA, USA, Chapter 6, pp. 103-128.

Robila, S. A. and Ragucci, J. W. (2006), ‘Don’t be a phish: Steps in user education’, *Proceedings of the 11th SIGCSE Conference on Innovation and Technology in Computer Science Education*, ACM, New York, USA, pp. 237-241.

Robson, C. (1983), *Experiment, Design and Statistics in Psychology: An Introduction*, Pelican Books, GB.

Root, R. W. and Draper, S. (1983), ‘Questionnaires as a software evaluation tool’, in *Proceedings of CHI’83*, ACM Press, New York, pp. 83-87.

Rossi, P. H., Wright, J. D. and Anderson, A. B. (1983), *Handbook of Survey Research*, Academic Press, New York, USA.

Salvucci, D. D. and Goldberg, J. H. (2000), 'Identifying fixations and saccades in eye-tracking protocols', in *Proceedings of Eye Tracking Research and Applications Symposium*, ACM Press, pp. 71-78.

Sasse, M. A. and Flechais, I. (2005), 'Usable security: Why do we need it? How do we get it?', in L. F. Cranor and S. Garfinkel (Editors), *Security and Usability*, O'Reilly, Sebastopol, CA, USA, Chapter 2, pp. 13-30.

Schultz, E. E., Proctor, R. W., Lien, M.-C. and Savendy, G. (2001), 'Usability and security: An appraisal of usability issues in information security methods', *Computers and Security*, Volume 20(7), pp. 620-634.

Sheng, S., Magnien, B., Kumaraguru, P., Acquisti, A., Cranor, L. F., Hong, J. and Nunge, E. (2007), 'Anti-phishing Phil: The design and evaluation of a game that teaches people not to fall for phish', in *Proceedings of the 3rd Symposium on Usable Privacy and Security (SOUPS'07)*, ACM, New York, pp. 88-99.

Sheng, S., Holbrook, M., Kumaraguru, P., Cranor, L. and Downs, J. (2010), 'Who falls for phish? A demographic analysis of phishing susceptibility and effectiveness of interventions', *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM Press, New York, pp. 373-382.

Sinclair, S. and Smith, S. W. (2005), 'The TIPPI point: Toward trustworthy interfaces', *IEEE Security and Privacy*, Volume 3, IEEE Computer Society, Los Alamitos, CA, USA, pp. 68-71.

Smetters, D. K. and Grinter, R. E. (2002), 'Moving from the design of usable security technologies to the design of useful secure applications', in *Proceedings of the New Security Paradigms Workshop (NSPW'02)*, ACM Press, Virginia Beach, Virginia, pp. 82-89.

Smith, R. E. (2001), *Authentication: From Passwords to Public Keys*, Addison Wesley Longman, USA.

Smyth, B. (2010), *Privacy vs. usability: A failure of Barclays online banking?*, Retrieved from http://www.bensmyth.com/publications/10barc/BarclaysFailure.pdf

Stamm, S. and Jagatic, T. N. (2007), 'Case study: Browser recon attacks', in M. Jakobsson and S. Myers (Editors), *Phishing and Countermeasures*, Wiley, Hoboken, NJ, USA, Chapter 6, pp. 210-219.

Stanton, N. A. and Young, M. S. (1999), 'What price ergonomics?', *Nature*, Volume 399, pp. 197-198.

Sturm, J. and Boves, L. (2005), 'Effective error recovery strategies for multimodal form-filling applications', *Speech Communication*, Volume 45(3), pp. 289-303.

Tognazzini, B. (2005), 'Design for usability', in L. F. Cranor and S. Garfinkel (Editors), *Security and Usability*, O'Reilly, Sebastopol, CA, USA, Chapter 3, pp. 31-46.

Toledano, D. T., Fernandez Pozo, R., Hernandez Trapote, A. and Hernandez Gomez, L. (2006), 'Usability evaluation of multi-modal biometric verification systems', *Interacting With Computers*, Volume 18(5), pp. 1101-1122.

Tsow, A. (2007), 'Pharming with appliances', in M. Jakobsson and S. Myers (Editors), *Phishing and Countermeasures*, Wiley, Hoboken, NJ, USA, Chapter 4, pp. 126-133.

Viecco, C. (2007), 'Honeypots', in M. Jakobsson and S. Myers (Editors), *Phishing and Countermeasures*, Wiley, Hoboken, NJ, USA, cCapter 5, pp. 159-173.

Viega, J. (2005), 'Security – problem solved?', *Queue*, ACM, New York, NY, USA, Volume 3(5), pp. 40-50.

Weir, C. S., Douglas, G., Carruthers, M. and Jack, M. A. (2009), 'User perceptions of security, convenience and usability for eBanking authentication tokens', *Computers and Security*, Volume 28(1), pp. 47-62.

Weir, C. S., Douglas, G., Richardson, T. and Jack, M. A. (2010), 'Usable security: User preferences for authentication methods in eBanking and the effects of experience', *Interacting with Computers*, Volume 22(3), pp. 153-164.

Weirich, D. and Sasse, M. A. (2001), 'Pretty good persuasion: A first step towards effective password security in the real world', in *Proceedings of the New Security Paradigms Workshop (NSPW'01)*, ACM Press, Cloudcroft, New Mexico, pp. 137-143.

Whalen, T. and Inkpen, K. M. (2005), 'Gathering evidence: Use of visual security cues in web browsers', in *Proceedings of Graphics Interface*, Canadian Human-Computer Communications Society, Victoria, British Columbia, pp. 137-144.

Whiteside, J., Bennett, J. and Holtzblatt, K. (1998), 'Usability engineering: Our experience and evolution', in M. Helander, ed., *Handbook of Human Computer Interaction*, Elsevier, Amsterdam, North-Holland, pp. 790-817.

Wilkie, J., Jack, M. A. and Littlewood, P. (2005), 'System-initiated digressive proposals in automated human-computer telephone dialogues: The use of contrasting politeness strategies', *International Journal of Human Computer Studies*, Volume 62, Issue 1, pp.41-71.

Witt, A. J. D. and Kuljis, J. (2006), 'Is usable security an oxymoron?', *Interactions*, Volume 13(3), 41-44.

Wu, M., Miller, R. C. and Garfinkel, S. L. (2006), 'Do security toolbars actually prevent phishing attacks?', *Proceedings of CHI'06*, pp. 601-610.

Yan, J., Blackwell, A., Anderson, R. and Grant, A. (2005), 'The memorability and security of passwords', in L. F. Cranor and S. Garfinkel (Editors), *Security and Usability*, O'Reilly, Sebastopol, CA, USA, Chapter 7, pp. 129-142.

Yan, J. and El Ahmad, A. S. (2008), 'Usability of CAPTCHAS or usability issues in CAPTCHA design', in *Proceedings of the 4th Symposium on Usable Privacy and Security (SOUPS'08)*, ACM, New York, pp. 44-52.

Yee, K.-P. (2005), 'Guidelines and strategies for secure interaction design', in L. F. Cranor and S. Garfinkel (Editors), *Security and Usability*, O'Reilly, Sebastopol, CA, USA, Chapter 13, pp. 247-273.

Zhuang, L., Zhou, F. and Tygar, J. D. (2007), 'Case study: Acoustic keyboard emanations', in M. Jakobsson and S. Myers (Editors), *Phishing and Countermeasures*, Wiley, Hoboken, NJ, USA, Chapter 6, pp. 221-239.

Zurko, M. E. and Simon, R. T. (1996), 'User-centered security', in *Proceedings of the New Security Paradigms Workshop (NSPW'96)*, ACM, New York, NY, USA, pp. 27-33.

Zviran, M. and Haga, W. J. (1999), 'Password security: An empirical study', *Management of Information Systems*, Volume 15(4), pp. 161-185.

Appendix A. Description of Statistical Tests used in this Research

A.1. Statistical Significance

Probability, *p*, is defined as the likelihood (between 0 and 1) of a given event happening. The statistical significance of an experimental result is defined as the probability that the experiment would produce a result as strong as this by chance alone.

Results can be referred to as being 'significant' ($p<0.05$, or 1 in 20 chance of happening by chance alone), 'highly significant' ($p<0.01$, or 1 in 100 chance of happening by chance alone) or 'very highly significant' ($p<0.001$, or 1 in 1000 chance of happening by chance alone).

Test statistics each have an idealised distribution which allows the calculation of the probability of obtaining the resulting value in the general population. If a result is found to be statistically significant, there are two possibilities. The first is that it was caused by a Type I error (false positive) where the populations are actually identical and there really is no difference and, purely by chance, larger values were obtained in one group and smaller values in the other. If statistically significant is defined to mean $p<0.05$, then there will likely be a significant finding in 5% of analyses where there really is no difference. Alternatively, a statistically significant result could be because the populations really are different, so the conclusion is correct.

If a result is not found to be statistically significant, there is still a possibility that there is still an effect. This is termed a Type II error (false negative). There is a trade-off in that making the chance of getting a Type I error smaller, the chance of getting a Type II error increases.

A.2. Statistical Tests

The choice of test to use to analyse data depends on the type of data that was measured and the assumptions that each test makes about the data it is given. For example, parametric tests assume that the data is normally distributed and is interval data. After performing a particular test, the output can include the test statistic itself, a probability value *(p)* and the degrees of freedom used. The degrees of freedom *(df)* of a test statistic refers to the number of independent values that are free to vary in its calculation. Typically, the degrees of freedom will be the number of input parameters minus any that depend on other input parameters.

The *t*-test is a parametric test used to compare the means of a single dependent variable between two different sample groups to determine if the difference is not just due to chance alone. The groups are determined by the independent variable and if there is a significant difference between the means then there is a high probability that the difference was caused by the effect of the independent variable. This assumes that the sample groups came from the sample population, that the data is normally distributed and that it is measured at the interval level. If the sample groups came from a repeated measures experiment (i.e. participants experienced both designs) then the paired samples *t*-test is used, otherwise the independent measures *t*-test is used.

When comparing more than two sample groups or more than one independent variable, the analysis of variance (ANOVA) test is used to investigate how the variables interact and which effects are important. Like the *t*-test, it is a parametric test, but it is also able to compare the means of three or more sample groups of between-subject and within-subject factors. The *t*-test is unsuitable for use on more than two groups because the repeated tests would inflate the cumulative Type I error. The ANOVA test manages to minimise such errors but can only test for an overall effect and does not provide information about which sample groups were affected.

After performing an ANOVA test, post-hoc procedures using *t*-tests can be used to perform pairwise comparisons between selected combinations of the sample groups, thus determining where effects lie. The cumulative Type I error resulting from performing these tests is typically managed by applying Bonferroni adjustment

(Field, 2009) which adjusts the criterion for significance, and there are other adjustments that are more appropriate to use when very many tests are being performed. In repeated measures experiments with three or more sample groups, the ANOVA test has the additional assumption of sphericity, where variances across groups are assumed to be equal. Mauchly's test is used to determine if the assumption of sphericity has been violated, and if so Greenhouse-Geisser correction can be applied to the output of the ANOVA.

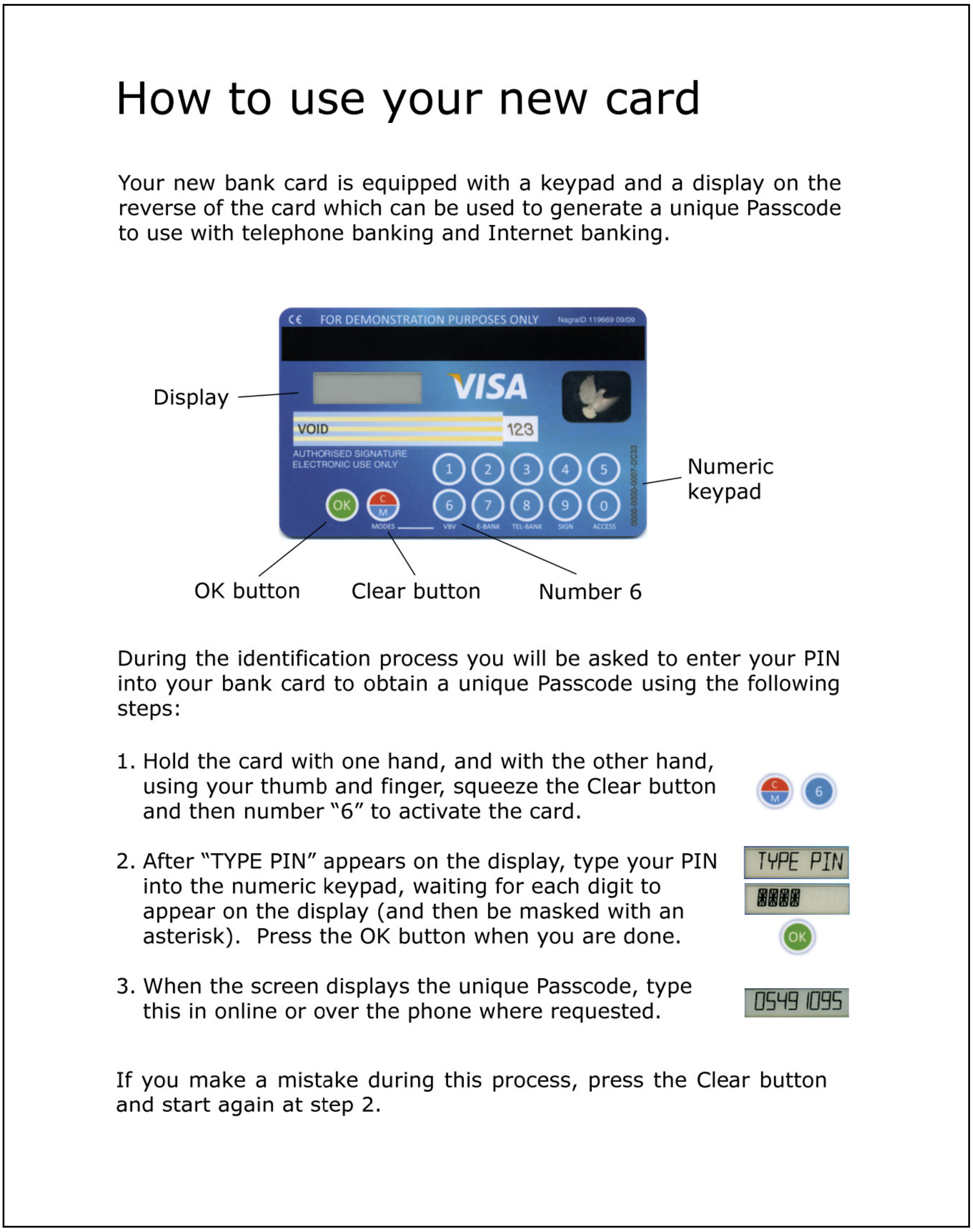
For two between-subjects variables measured on a category scale, a non-parametric test for associations between them can be performed with the chi-square test. The frequencies for each category are compared with those expected by chance and a probability score is calculated for the likelihood of an effect between the variables. The chi-square test requires that each of the frequencies in each category be greater than five and assumes that a participant will have contributed only once to just one variable. Analysis of success rates (pass or fail, encoded as 1 and 0) between two treatments in the same population can be performed with McNemar's test, or with Cochran's Q test for three or more treatments.

The measure of scale reliability of a questionnaire can be ascertained by calculating Cronbach's α. This compares all possible combinations of the split-half reliability (Field, 2009) of questionnaire responses for each participant to determine if all questions contribute equally well to the overall questionnaire. Values of α above 0.7 are typical acceptable values for questionnaire reliability.

All of the statistical analysis for the experiments presented in this work was carried out using SPSS[1].

[1] SPSS (Statistical Package for the Social Sciences).

Appendix B. User Guides for the CodeSure Card



How to use your new card

Your new bank card is equipped with a keypad and a display on the reverse of the card which can be used to generate a unique Passcode to use with telephone banking and Internet banking.

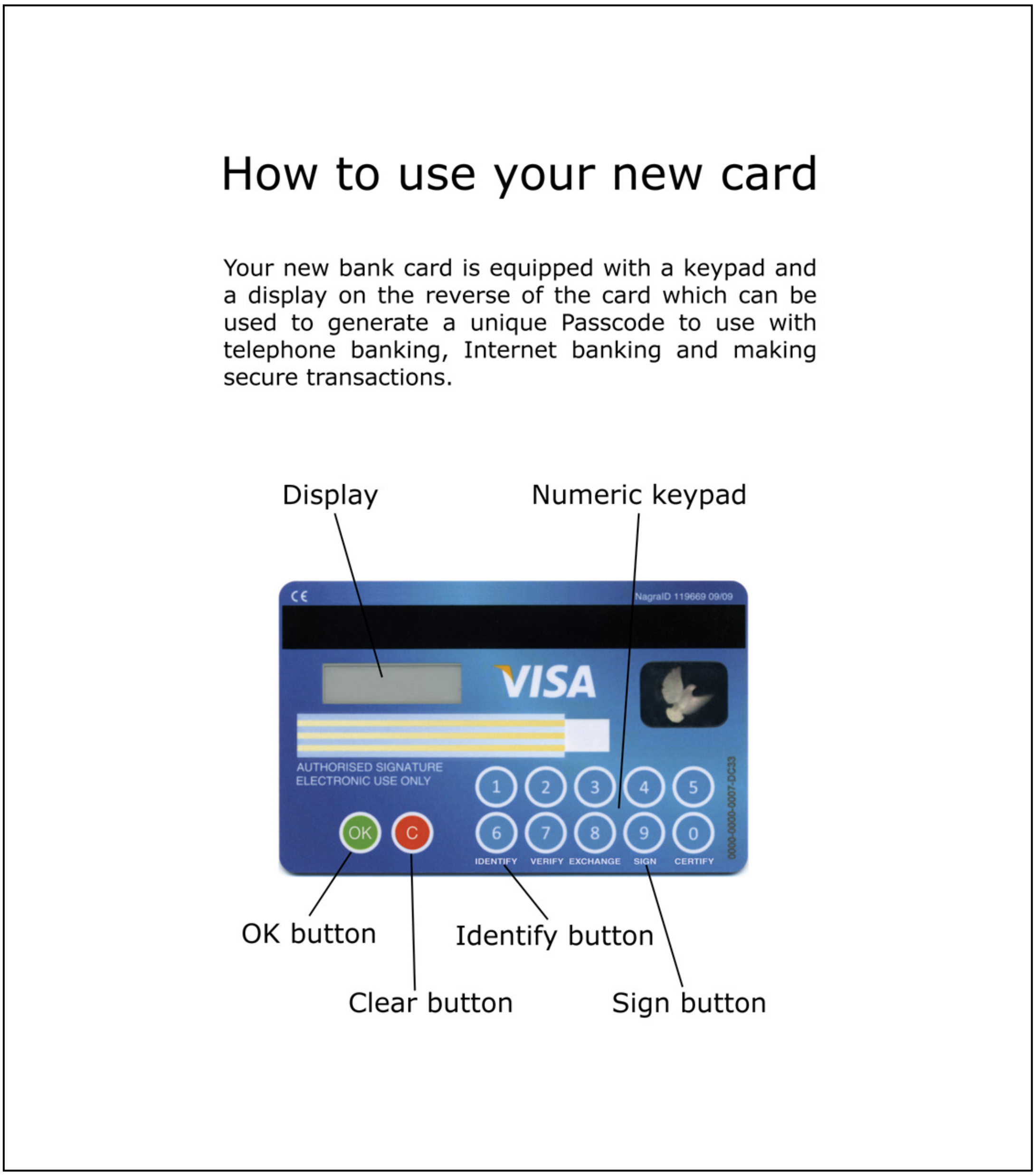
FOR DEMONSTRATION PURPOSES ONLY
NagraID 119669 09/09
VISA
VOID
123
AUTHORISED SIGNATURE ELECTRONIC USE ONLY
1 2 3 4 5
6 7 8 9 0
OK
C M
MODES
VBV
E-BANK
TEL-BANK
SIGN
ACCESS
Display
Numeric keypad
OK button
Clear button
Number 6

During the identification process you will be asked to enter your PIN into your bank card to obtain a unique Passcode using the following steps:

1. Hold the card with one hand, and with the other hand, using your thumb and finger, squeeze the Clear button and then number "6" to activate the card.
2. After "TYPE PIN" appears on the display, type your PIN into the numeric keypad, waiting for each digit to appear on the display (and then be masked with an asterisk). Press the OK button when you are done.
3. When the screen displays the unique Passcode, type this in online or over the phone where requested.

If you make a mistake during this process, press the Clear button and start again at step 2.

Figure B.1: The CodeSure Card Baseline Usability Experiment

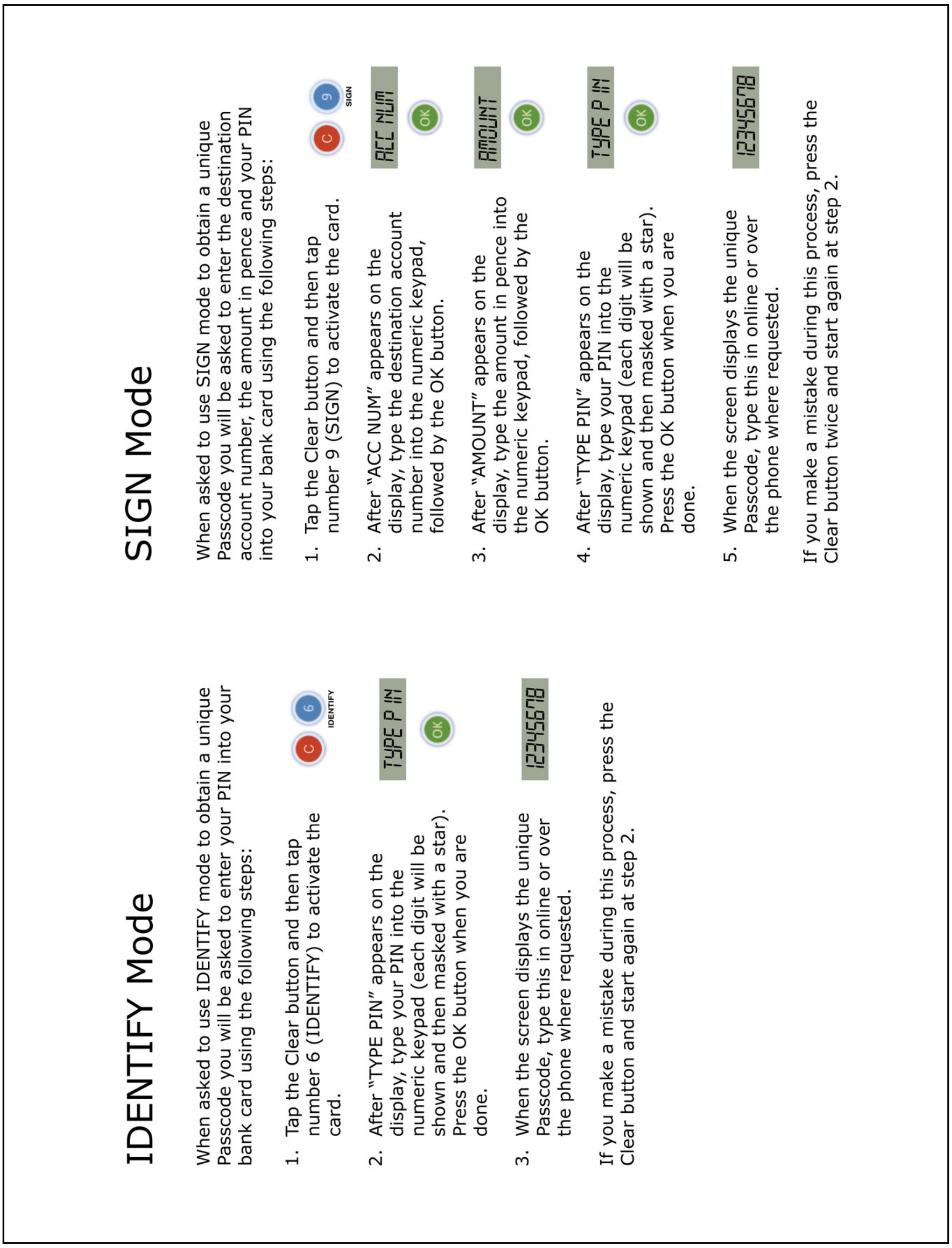


How to use your new card

Your new bank card is equipped with a keypad and a display on the reverse of the card which can be used to generate a unique Passcode to use with telephone banking, Internet banking and making secure transactions.

Display
Numeric keypad
NagraID 119669 09/09
VISA
AUTHORISED SIGNATURE
ELECTRONIC USE ONLY
OK
C
1
2
3
4
5
6
7
8
9
0
IDENTIFY
VERIFY
EXCHANGE
SIGN
CERTIFY
0000-0000-0007-DC33
OK button
Identify button
Clear button
Sign button

Figure B.2: The Enhanced CodeSure Card Usability Experiment (Front Page)



IDENTIFY Mode

When asked to use IDENTIFY mode to obtain a unique Passcode you will be asked to enter your PIN into your bank card using the following steps:

1. Tap the Clear button and then tap number 6 (IDENTIFY) to activate the card.
2. After "TYPE PIN" appears on the display, type your PIN into the numeric keypad (each digit will be shown and then masked with a star). Press the OK button when you are done.
3. When the screen displays the unique Passcode, type this in online or over the phone where requested.

If you make a mistake during this process, press the Clear button and start again at step 2.

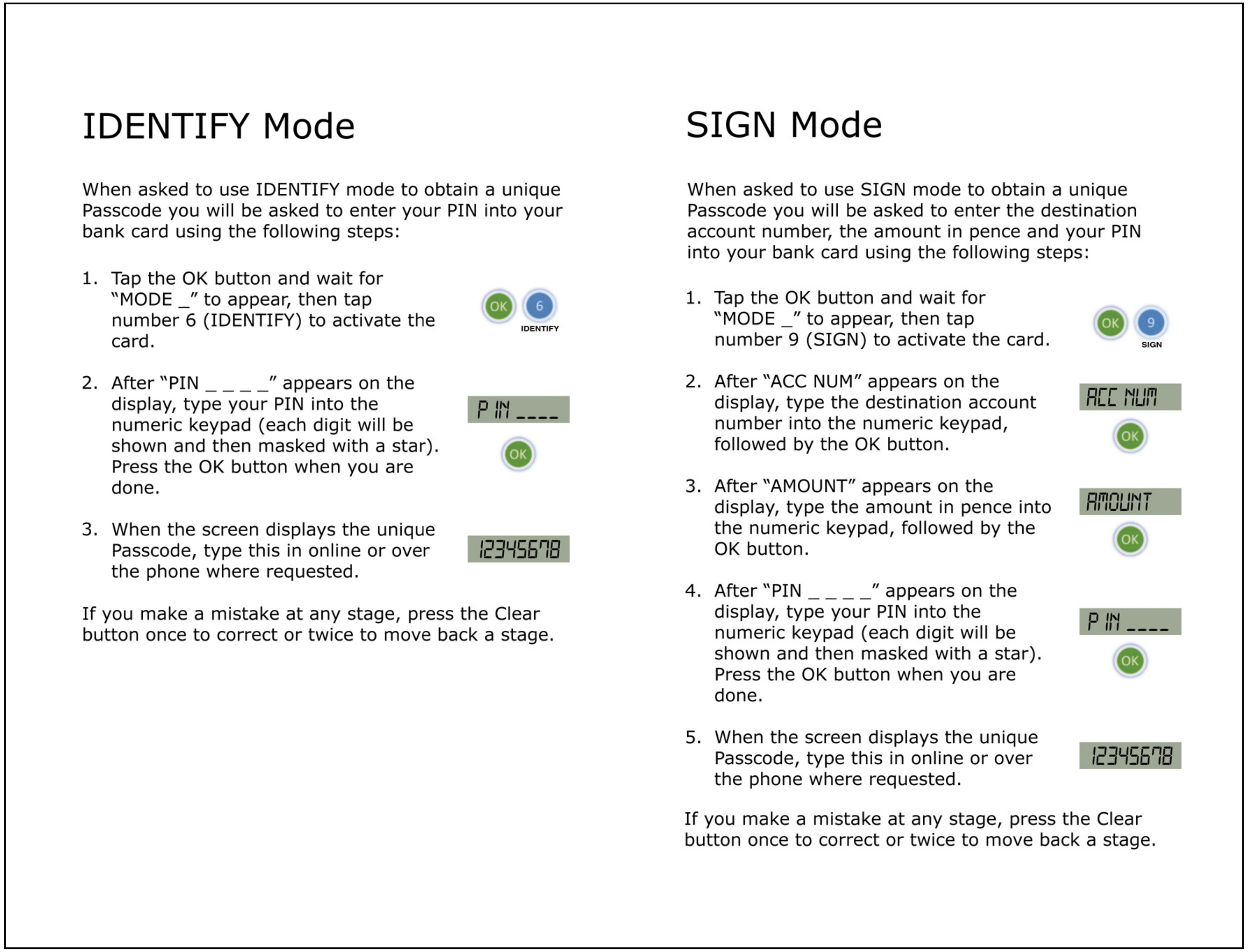
SIGN Mode

When asked to use SIGN mode to obtain a unique Passcode you will be asked to enter the destination account number, the amount in pence and your PIN into your bank card using the following steps:

1. Tap the Clear button and then tap number 9 (SIGN) to activate the card.
2. After "ACC NUM" appears on the display, type the destination account number into the numeric keypad, followed by the OK button.
3. After "AMOUNT" appears on the display, type the amount in pence into the numeric keypad, followed by the OK button.
4. After "TYPE PIN" appears on the display, type your PIN into the numeric keypad (each digit will be shown and then masked with a star). Press the OK button when you are done.
5. When the screen displays the unique Passcode, type this in online or over the phone where requested.

If you make a mistake during this process, press the Clear button twice and start again at step 2.

Figure B.3: The Enhanced CodeSure Card Usability Experiment (Baseline)



OK
6
IDENTIFY
P IN ____
OK
12345678
OK
9
SIGN
ACC NUM
OK
AMOUNT
OK
P IN ____
OK
12345678

IDENTIFY Mode

When asked to use IDENTIFY mode to obtain a unique Passcode you will be asked to enter your PIN into your bank card using the following steps:

1. Tap the OK button and wait for “MODE _” to appear, then tap number 6 (IDENTIFY) to activate the card.
2. After “PIN _ _ _ _” appears on the display, type your PIN into the numeric keypad (each digit will be shown and then masked with a star). Press the OK button when you are done.
3. When the screen displays the unique Passcode, type this in online or over the phone where requested.

If you make a mistake at any stage, press the Clear button once to correct or twice to move back a stage.

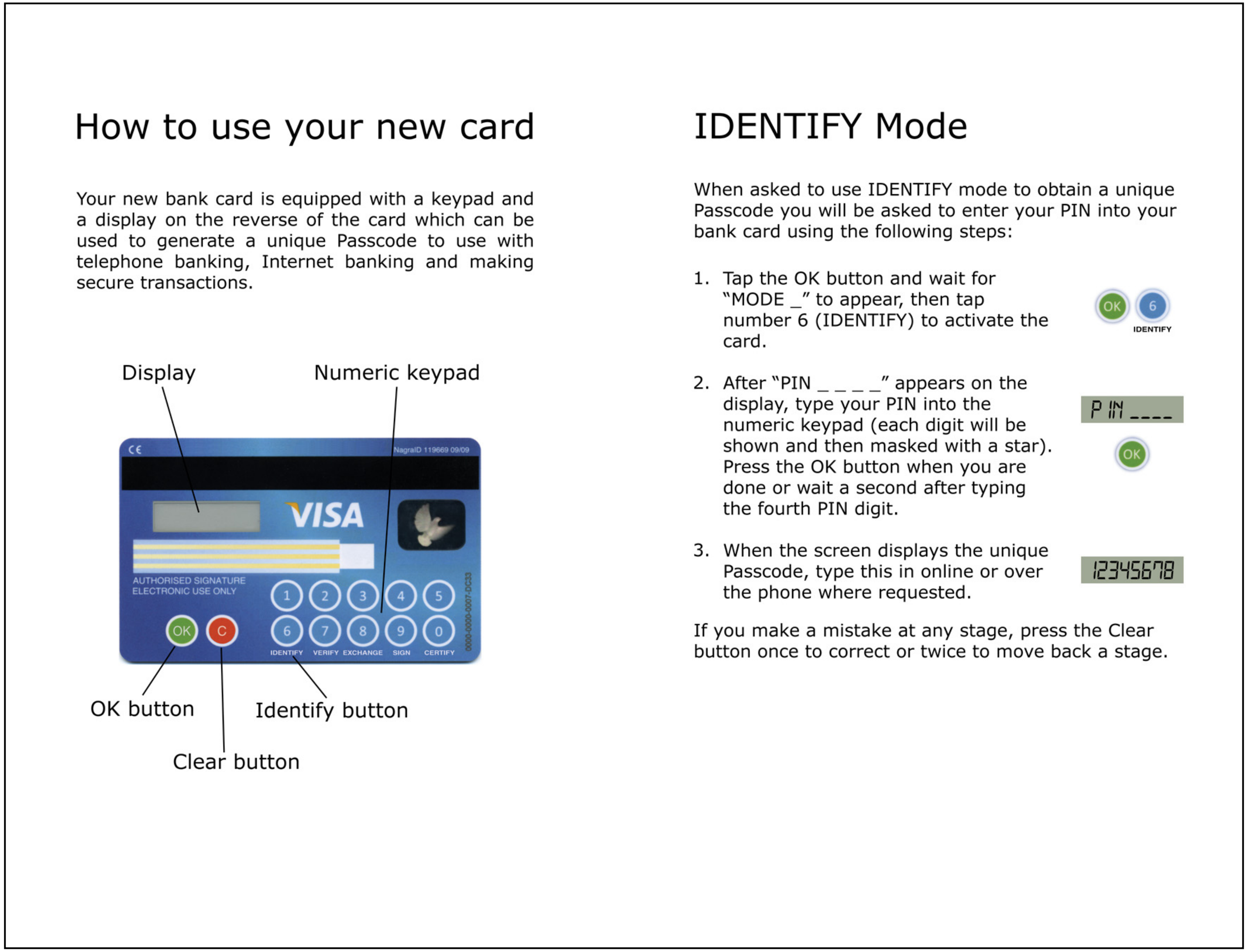
SIGN Mode

When asked to use SIGN mode to obtain a unique Passcode you will be asked to enter the destination account number, the amount in pence and your PIN into your bank card using the following steps:

1. Tap the OK button and wait for “MODE _” to appear, then tap number 9 (SIGN) to activate the card.
2. After “ACC NUM” appears on the display, type the destination account number into the numeric keypad, followed by the OK button.
3. After “AMOUNT” appears on the display, type the amount in pence into the numeric keypad, followed by the OK button.
4. After “PIN _ _ _ _” appears on the display, type your PIN into the numeric keypad (each digit will be shown and then masked with a star). Press the OK button when you are done.
5. When the screen displays the unique Passcode, type this in online or over the phone where requested.

If you make a mistake at any stage, press the Clear button once to correct or twice to move back a stage.

Figure B.4: The Enhanced CodeSure Card Usability Experiment (Enhanced)



How to use your new card

Your new bank card is equipped with a keypad and a display on the reverse of the card which can be used to generate a unique Passcode to use with telephone banking, Internet banking and making secure transactions.

Display
Numeric keypad
OK button
Identify button
Clear button
VISA
AUTHORISED SIGNATURE ELECTRONIC USE ONLY
IDENTIFY
VERIFY
EXCHANGE
SIGN
CERTIFY

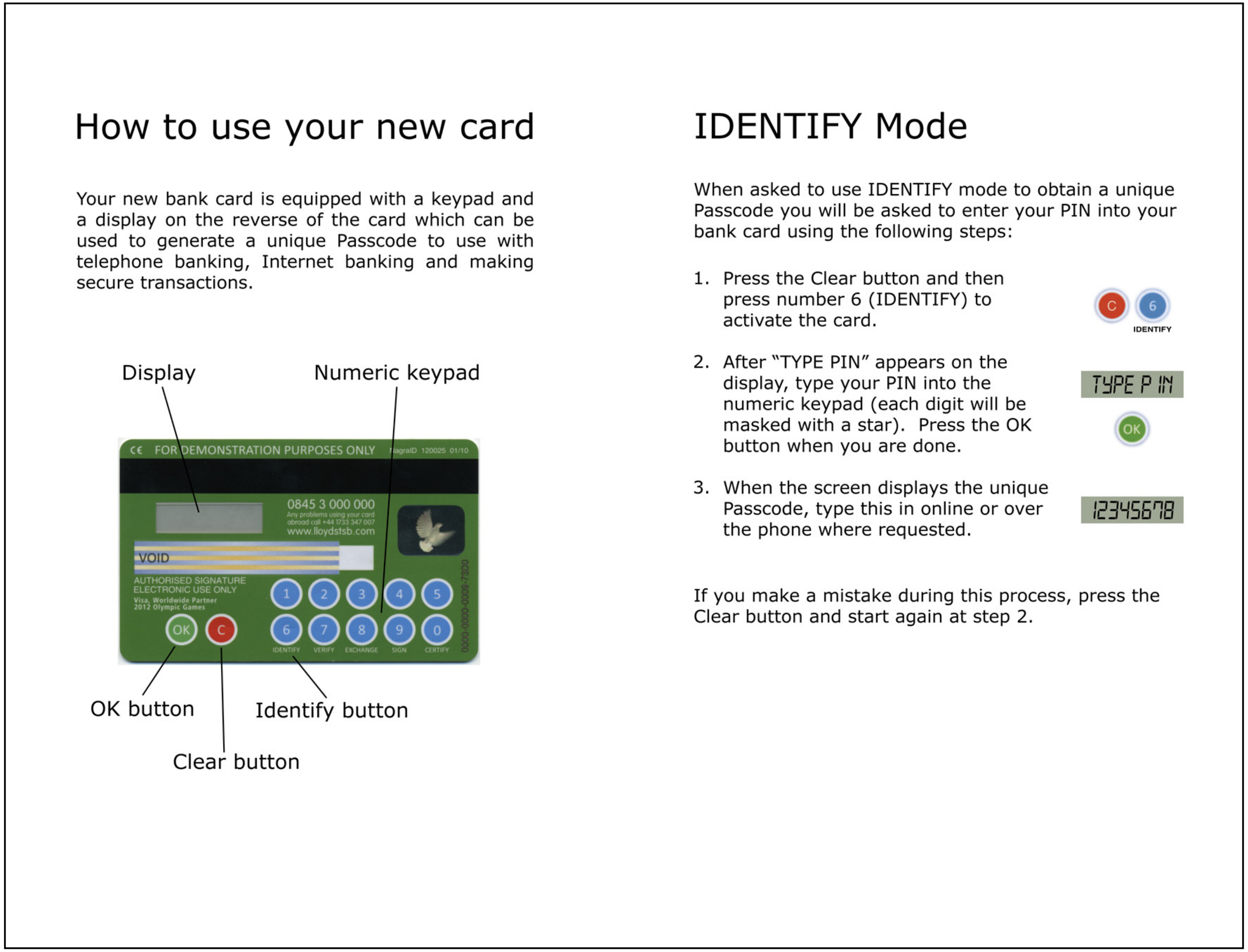
IDENTIFY Mode

When asked to use IDENTIFY mode to obtain a unique Passcode you will be asked to enter your PIN into your bank card using the following steps:

1. Tap the OK button and wait for "MODE _" to appear, then tap number 6 (IDENTIFY) to activate the card.
2. After "PIN _ _ _ _" appears on the display, type your PIN into the numeric keypad (each digit will be shown and then masked with a star). Press the OK button when you are done or wait a second after typing the fourth PIN digit.
3. When the screen displays the unique Passcode, type this in online or over the phone where requested.

If you make a mistake at any stage, press the Clear button once to correct or twice to move back a stage.

Figure B.5: The CodeSure Keypad and Display Synchrony Experiment



How to use your new card

Your new bank card is equipped with a keypad and a display on the reverse of the card which can be used to generate a unique Passcode to use with telephone banking, Internet banking and making secure transactions.

Display
Numeric keypad
FOR DEMONSTRATION PURPOSES ONLY
0845 3 000 000
www.lloydstsb.com
VOID
AUTHORISED SIGNATURE
ELECTRONIC USE ONLY
OK
C
IDENTIFY
VERIFY
EXCHANGE
SIGN
CERTIFY
OK button
Identify button
Clear button

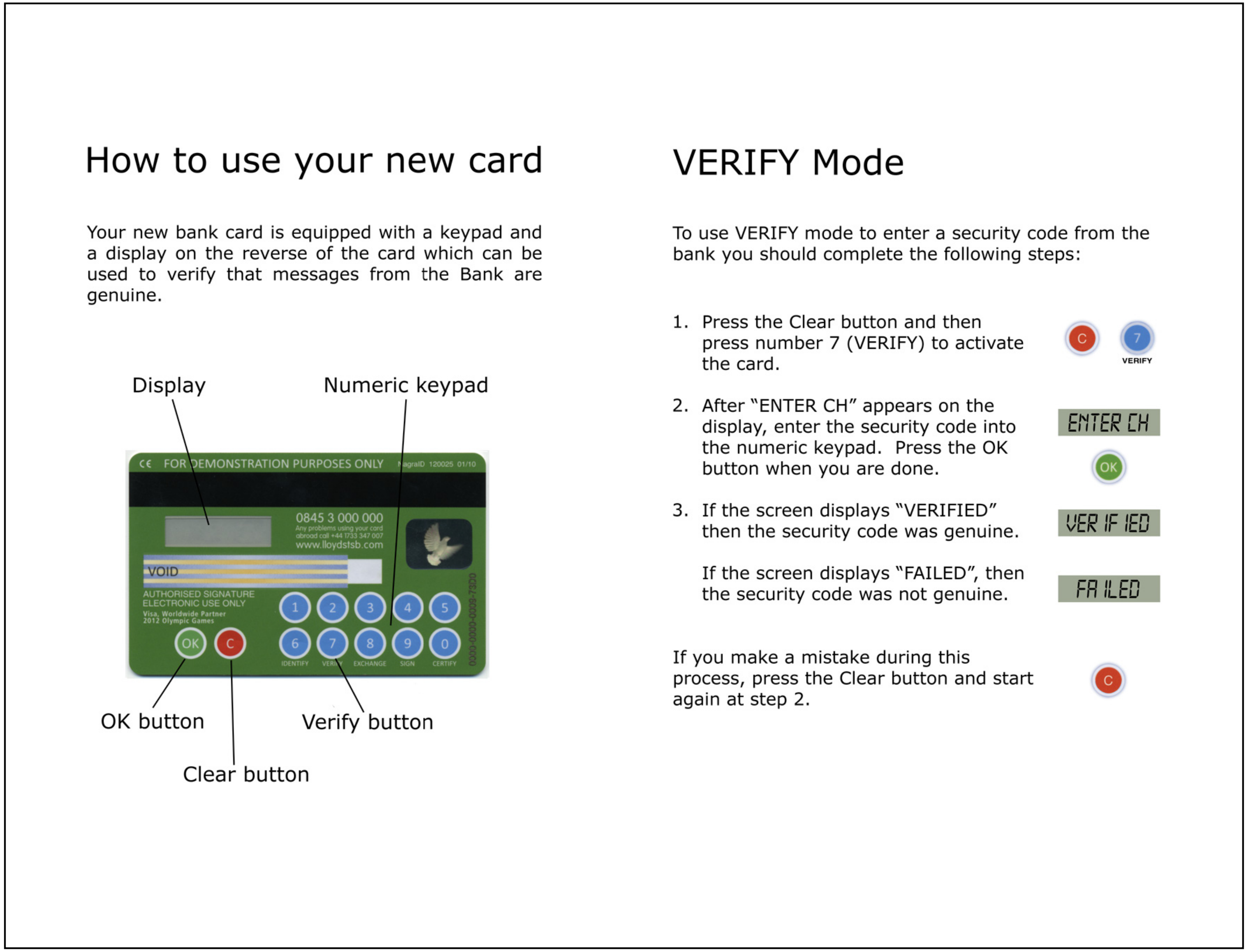
IDENTIFY Mode

When asked to use IDENTIFY mode to obtain a unique Passcode you will be asked to enter your PIN into your bank card using the following steps:

1. Press the Clear button and then press number 6 (IDENTIFY) to activate the card.
2. After "TYPE PIN" appears on the display, type your PIN into the numeric keypad (each digit will be masked with a star). Press the OK button when you are done.
3. When the screen displays the unique Passcode, type this in online or over the phone where requested.

If you make a mistake during this process, press the Clear button and start again at step 2.

Figure B.6: The CodeSure Multi-Channel Usability Experiment



How to use your new card

Your new bank card is equipped with a keypad and a display on the reverse of the card which can be used to verify that messages from the Bank are genuine.

VERIFY Mode

To use VERIFY mode to enter a security code from the bank you should complete the following steps:

1. Press the Clear button and then press number 7 (VERIFY) to activate the card.
2. After "ENTER CH" appears on the display, enter the security code into the numeric keypad. Press the OK button when you are done.
3. If the screen displays "VERIFIED" then the security code was genuine.

   If the screen displays "FAILED", then the security code was not genuine.

If you make a mistake during this process, press the Clear button and start again at step 2.

Figure B.7: The CodeSure Sender Authentication Experiment