A Case of Sesame Seeds:
Growing and Nurturing Credentials in the Face of Mimicry

Name:	Arne Padmos
Student NR:	100693458
Supervisor:	Dr. Lizzie Coles-Kemp

Submitted as part of the requirements for the award of the MSc in Information Security at Royal Holloway, University of London.

I declare that this thesis is my own work and that I have acknowledged all quotations from the published or unpublished works of other people. Also, I declare that I have read the RHUL Regulations on Assessment Offences 2010/11, and I submit this project report as my own work.

Date: 2nd September 2011

Place: Meppel, the Netherlands

Signature:

A Case of Sesame Seeds:
Growing and Nurturing Credentials in the Face of Mimicry

Arne Padmos

2nd September 2011

Oh, what a tangled web we weave
When first we practise to deceive!
—Walter Scott

The purpose of this paper is to put the study of mimicry on the information security research map. Mimicry in humans has received little scholarly attention. Sociologist Diego Gambetta has constructed a framework that enables reasoning about episodes of mimicry based on trust in signs. By looking at the problem of phishing the applicability of this framework to problems of mimicry in information security system was tested. It was found that while the framework offers valuable insights, it needs to be updated since the assumptions that it makes do not hold in practice. A new framework is proposed, built on the core ideas of Gambetta's framework, and extended with results from a literature study of phishing and other sources. This framework has been used for finding possible solutions to problems in web browser interface design. Because the nature of authentication was found to be the observation of discriminatory signals the paper also discusses the ethical issues surrounding the use of credentials.

We hope that this paper will help system designers in finding and choosing appropriate credentials for authentication. By using the proposed framework a system can be analysed for the presence of credentials that enable the discrimination between genuine users and impostors. The framework can also serve as a method for identifying the dynamics behind user verification of credentials. The two problems that the framework can help address are the impersonation of providers and the impersonation of users. Like much other security research the results of this paper can be misused by attackers. It is expected that the framework will be more useful for defenders than attackers, as it is of an analytical nature, and cannot be used directly in any attacks.

Since this study is of an exploratory nature the findings of the study need to be verified through research with greater validity. The paper contains directions for further research.

Contents

1 Introduction

The story *Ali Baba and the Forty Thieves* clearly illustrates the threat of mimicry and the dangers of easy to fake credentials. Were it not for this seed of trouble the thieves might still be alive. Tales of mimicry are plenty in art and in real life. Since the 19th century a lot of work has been done to understand mimicry in animals. This is in contrast to mimicry in humans which has received scant scholarly attention [42]. Work on mimicry in the information security field appears similarly thin. There has been some mathematical work on subverting intrusion detection systems [116] and detecting stolen user credentials [110], as well as investigations into the ease of spoofing biometrics [31, 39, 123, 100]. We are not aware of an in-depth look at mimicry in information security and seek to provide it here.

1.1 Research Focus

In addition to dealing with insecurity from a risk management perspective, humans also approach it as a trust problem [44]. Besides implementing measures to decrease their risk, they engage in trust acts, seeking others that they think are unlikely to take advantage of their vulnerability. At the moment the information security field is mainly concerned with risk management and assurance, even though placing trust correctly can lead to a better and more efficient culture [36]. This is evident in the use of the word 'trust', which more often than not merely means 'assurance'. Take cryptography: in contrast to the prevalent discourse it cannot support trust directly [118]. However, even in situations where assurance is claimed to obviate the need for trust (e.g. through contracts), trust must still be placed in those who operate the control mechanisms [16]. In information security, trust plays an important role since interactions between entities often involves a degree of reliance and vulnerability [5].

Because of the importance of trust in information security systems we thought it important to develop a better understanding of trust dynamics. To do so several works [9, 42, 44, 43] by sociologist Diego Gambetta were consulted. These works find that people do not trust in others directly, but instead 'trust in signs' [9] that mediate the knowledge of the other's trustworthiness, which is supported by many anecdotes [42] and descriptive studies of taxi drivers [44] and criminals [43]. The main trust problem is finding reliable signs that evidence someone's trust warranting characteristics, not determining what these characteristics are. Often the characteristics of a trustworthy person are known, but the signs that mediate them are difficult to rely on because mimics corrupt them. Mimics pass themselves off as trustworthy individuals by displaying signs that get them mistaken for trustworthy individuals. Because distinguishing real from fake is the biggest obstacle to placing trust correctly, the threat of mimicry is the foremost trust problem.

Phishing is a model problem for researching usability concerns in information security, with insights applicable to a wide range of other usability problems [23]. It also seems a model problem when looking at mimicry, with similar generalizability. However, researching phishing can be problematic: observational studies take a lot of time, laboratory studies suffer from unmotivated participants, interviews do not give an accurate picture of the actual behaviour of users, and live experiments bring ethical and legal problems [30]. Deceiving participants without informed consent is deemed unethical [7], although this issue has been brushed aside by some researchers [32] which is irresponsible and a very slippery slope. This does not mean that empirical research is impossible, but current approaches do have their issues and more work appears necessary in determining appropriate guidelines and viable research methodologies. For this research a theoretical approach was found to be appropriate since it is an exploratory study. The project is a descriptive study of mimicry during authentication. Although this does not allow a proper scientific test of hypotheses, it does provide potentially fruitful material for further research [44]. The results of the study can be tested in empirical experiments by those who want to sail these waters, but they should know where the rocks are to prevent running aground.

By looking at mimicry in the information security field this project responds to calls for an interdisciplinary approach to the study of mimicry [42]. The information security discipline can contribute insights that will enrich the framework by providing input from another perspective, so that assumptions such as rationality can be revisited. It has been conceded that the framework needs to be updated to take into account irrational behaviour in various contexts, such as the effectiveness of the tactics of con artist [44]. As has been said previously, the information security fields also has a lot to gain by looking at the problem of mimicry. Even if the assumptions of the framework need some polishing [70], its perspective is thought to offer valuable insights. These can provide ideas that may lead to new solution or the improvement of old ones.

1.2 Terminology

Trust is the action or willingness of the truster to make themselves vulnerable in the face of uncertain payoffs of the trustee: if the trustee is honest then trust pays, but if their trust is betrayed they would have done better by not trusting [9]. *Mimicry* is the deliberate display of a perceivable property (manifestum) by an entity that wants to be taken as the possessor of a non-visible property (krypton) of another entity while not actually possessing this property [9, 42, 44]. A *credential* is evidence of one's characteristics, such as authority, background, identity, or predispositions [35, 89]. *Authentication* is the checking of credentials before deciding on some action [58, 34, 33]. This often takes the form of verifying claims about identity, origin, or correctness. *Access control* consists of the authentication of entities and determining whether they are allowed access to assets based on some policy [58, 35].

The topic of this project is *mimicry during authentication*, which consists of two problems: the possibility of the authenticater or the authenticatee not being who they claim to be. Although all authentication is taken to consist of observing credentials, not all observation of credentials consists of authentication, since mimics may be at the receiving end of a communication of credentials. Similarly, when a credential is communicated this may not be rock solid evidence of being an authentic entity, since the entity may be vulnerable to mimicry. Identifying how to deal with these two problems is the goal of this project, with phishing as focal point of the investigation.

The definition of phishing is contested, with some sources defining it as any form of pilfering of user information [72, 62, 65], while other employ a more narrow definition limited to the vectors of email and websites [47]. There is also confusion about the relationship between pharming and phishing [111, 62]. Another issue is the use of cheesy terminology such as 'phish' [53]: continuing this trend would mean victims of pharming are 'phegetables'. Even though there is all this confusion we will continue to use the term phishing for the sake of backward compatibility. *Phishing* is taken to be a double dose of mimicry during authentication: both the sender of the email and the publisher of the website are not who they claim to be.

1.3 Examples

In order to provide a more down to earth explanation and to contextualise the definitions in the previous subsection it seems helpful to provide a few scenarios to better explain the terms.

When users read their email they cannot be expected to have a thorough understanding of the risks they face. They engage in *trust* behaviour, by looking for signs that show that the email comes from a valid receiver. One *credential* is the presence of an email address that users recognise. This credential can be *mimicked* by an imposted by spoofing the email address. If a user received an email asking for some information the user would *authenticate* the sender of the email based on various credentials of the email, and engage in *access control* to determine whether the entity has access to the information. On top of trust strategies, whereby users decide whether to make themselves vulnerable or not in the face of certain benefits, they engage in risk management, namely the selective sharing of their email address and the use of spam filters.

People's password policies show that they employ both risk management and trust strategies. Examples of risk management are evaluating how strong a password needs to be and deciding whether or not it is safe enough to write a password down. When sharing passwords users seem to engage in *trust* behaviour for various reasons, ranging from efficiency to sociability [36]. Research has shown that users tend to share their passwords with coworkers as well as people outside the company [107]. When choosing to share a password the *credentials* take the form of signals mediating the characteristics of years at the company, helpfulness, administrative role, previous interaction, and others. *Mimics* can display these credentials as well, and social engineering attacks leading to password compromise have been widely reported. After *authenticating* the credentials the users will engage in *access control* of their passwords, and will decide what passwords to share with the other entity.

1.4 Objectives

The objectives of this project are to

- put the study of mimicry on the information security research map;
- compile the theories of Gambetta about mimicry;
- evaluate the current state of phishing research;
- describe existing models of phishing and their relevance;
- show the relevance of Gambetta's theories to the phishing problem;
- test their usefulness by applying them to information security issues;
- discuss the ethical dilemmas surrounding credentials;
- make testable predictions based on the covered theory; and
- sketch a roadmap for further research into mimicry.

1.5 Overview

The layout of this paper follows the objectives that are defined in the previous subsection. The next section is a theoretical look at trust, mimicry, and credentials. A literature study of phishing research, including a description and critique of existing models is the focus of the second section. Another section is dedicated to the relevance of Gambetta's theories. Following this is a section that combines theory and practice by looking for solutions to current problems. A section discussing the ethical dimension of credentials is also provided. Guidance for practitioners and a roadmap for future research conclude the paper.

2 Trust, Mimicry, and Credentials

In this section the theories of Diego Gambetta pertaining to trust and mimicry are described. The material here is based on work describing [9] and extending [42] the theoretical framework, and two ethnographic studies of taxi drivers [44] and criminals [43] that test the framework. Because divergent terminology is used in these works for the same concepts, it was thought appropriate to discuss the framework using a coherent terminology. The goal of this section is to explain Gambetta's theories. A critique and possible improvements are saved for a later section, before which a thorough investigation of phishing is completed, so that the evaluation has something to build on. The main source of information is the original article [9] where the framework was first proposed, and this is the source of the material in this section unless noted otherwise.

2.1 Basic Trust Games

Gambetta and Bacharach built a framework for describing and evaluating trust decisions, based on certain assumptions. They take actors to be rational entities, and assume that all choices are decomposable into constituent decisions, whose outcome can be evaluated using game theoretic (win/lose) logic. These assumptions come from the origins of the theory, which are game theory and signalling theory. The theory portrays trust as a game of asymmetric information, with the pay-offs for breaking trust unknown to the truster, but known to the trustee. Raw and all-in payoffs are proposed, with the raw payoffs representing the rewards for the players not taking into account any contextual and historical factors such as their upbringing or future expectations. All-in payoffs are those that take into account any correcting factors. Raw payoffs are assumed to be known to all players, while not all all-in payoffs are. For an example trust game see figure 1.

Gambetta and Bacharach see trust as *trusting someone to do something* (and thus also the corollary, trusting someone not do so something). This perspective is argued to capture many trust scenarios, although it isn't designed for the analysis of every kind of trust. Trusting someone to do something is defined as acting under the expectation that the other will do what is asked of them, and both individuals are aware that when the trustee fails to do this something, the truster would have been better off not trusting the person, and both also know that the truster acting in the way they do will give the trustee a (selfish) reason to betray the truster (to not do that something). A more precise description of this perspective is provided by the basic trust game, which is a two-player, strategic-form, noncooperative game.

Trustee, E

		β	$\bar{\beta}$
Truster, R	α	3, 1	−3, 4
	$\bar{\alpha}$	0, 0	0, 0

Raw payoffs

Trustee, E

		β	$\bar{\beta}$
Truster, R	α	3, 2	−3, −4
	$\bar{\alpha}$	0, 0	0, 0

All-in payoffs

Figure 1: Example of raw and all-in payoffs, adapted from [9, 44]

$$\langle \alpha, \beta \rangle_R > \langle \bar{\alpha}, \beta \rangle_R \tag{1}$$

$$\langle \alpha, \bar{\beta} \rangle_R < \langle \bar{\alpha}, \bar{\beta} \rangle_R \tag{2}$$

$$\alpha\colon \begin{cases} \beta < \bar{\beta} & \text{for selfish payoffs} \\ \beta \gtrless \bar{\beta} & \text{for all-in payoffs} \end{cases} \tag{3}$$

$$\bar{\alpha}\colon \ \ldots \tag{4}$$

$$\beta\colon \ \alpha > \bar{\alpha} \tag{5}$$

$$\bar{\beta}\colon \ \alpha < \bar{\alpha} \tag{6}$$

The basic trust game consists of two players, the truster R and the trustee E, both with two strategies α and $\bar{\alpha}$, and β and $\bar{\beta}$ respectively. The payoffs for the players are represented by two payoff matrices, one for the 'raw' payoffs that are motivated by self-interest, and another for the 'all-in' payoffs that include things like ethics, upbringing, morals. Many other payoff matrices are possible, but Gambetta and Bacharach have chosen to simplify it to just two for simplicity's sake. The assumptions about the knowledge of the players are that both R and E know about the possible strategies (α, $\bar{\alpha}$, β, $\bar{\beta}$), that both players know R's raw and all-in payoffs, and that normally only E would know E's all-in payoffs, although both players know E's raw payoffs. The game is thus one of asymmetric information about E's payoffs.

Gambetta and Bacharach identify two problems with the above formalisation of trust. Firstly, selfish payoffs may not be selfish at all, for example, when someone donates to charity instead of paying off a loan. The argument for calling one of the payoff matrices selfish is for simplicity's sake. Secondly, the basic trust game only includes one peril ($\bar{\beta}$). However, at the cost of additional complexity it can be extended to include multiple perils (β_1, β_2, ...). It may be noted that one could move even further to also include multiple profit and loss scenarios for R, by enabling multiple trusting choices of R, as well as increasing the number of positive actions of E.

Figure 1 is an example of possible payoff matrices. The exact values of the numbers are not important, since they are purely illustrative. It only matters that equations 1 and 2 for the payoffs of R under the given strategies are correct for both matrices, and that equations 3 to 6 are are satisfied. This ensures that R will prefer strategy β (trusting act) when E is expected to choose β (trusted-in act), but would have done better not to trust ($\bar{\alpha}$) when their trust is betrayed ($\bar{\beta}$). E is trustworthy if E chooses β, believing R trusts them. An assumption is that all-in payoffs govern choice, thus for E to be trustworthy their all-in payoffs must be different from their raw payoffs. If R thinks E is trustworthy, E thinks R trusts them, and R has all-in payoffs favouring strategy β, situation will appear where R trusts E and does α, and E thinks they are trusted and does β.

Trust-warranting properties seem to be present in most situations. There are many possible properties that can make a trustee trustworthy. These can be E's reasons to choose β or their preference for the $\langle \alpha, \beta \rangle$ outcome. The choice for β over $\bar{\beta}$ can be a conscious or subconscious mechanisms, and the value attached to $\langle \alpha, \beta \rangle$ can result from dispositions, norms, or morals. The trust-warranting properties of an individual transform their raw payoffs, replacing them by all-in payoffs. Raw payoffs can be seen as first-order payoffs, with all-in payoffs being of a higher order.

2.2 Mimic-Beset Trust Games

Gambetta and Bacharach identify two trust problems, the first-order trust problem and the second-order trust problem. The first-order trust problem is the uncertainty of the truster (R) over the payoffs of the trustee (E). Trust would not be necessary if all payoffs were known. The uncertainty over the payoffs leads R to look for signals that mediate trust-warranting properties. While all raw payoffs are assumed to be known to both parties, R does not know the all-in payoffs of E. The first-order trust problem is R's uncertainty over E's payoffs. All-in payoffs are contextual, and if R can discover how the properties of E relate to this context they can determine the all-in payoffs of E, and can solve the first-order trust problem by observing E's properties. While the discovery of these factors may be difficult, the rest of the framework assumes they are known to the players.

Once the relevant trust-warranting properties are known to both players the second-order trust problem appears. In this case E has the incentive to fake the properties so that they are more likely to be trusted. This is possible because in virtually all scenarios trust-warranting properties are unobservable, and 'signs mediate the knowledge of them'. Several concepts are added to the basic trust game for discussing this finding. *Krypta* are unobservable properties of people, and *t-krypta* are those that are trust-warranting. *Manifesta* are observable features of people that may be associated with krypta and t-krypta. Thus, manifesta may be signs of krypta, although their trustworthiness depends on the reliability of the manifesta.

Mimic-beset trust games are an upgraded version of the basic trust game, in which the knowledge of trust-warranting properties is mediated by signs and where an opportunist is present. The opportunist is a type of interactant that is not merely guided by raw payoffs, but actively seeks to deceive. A mimic of krypton k, through manifestum m, 'is a person who does not have k and deliberately displays m in order to be taken by another to have k'. When such a display is negative it is called camouflage. The second-order trust problem is finding manifesta that reliably signal t-krypta, k_t. When mimics are on the loose this is an important question. The problem is relevant t-krypta and krypta. When the game involves a temporal factor the reidentification of players is important. A player's identity is a krypton that is mediated through the display of a manifestum, which serves as the beacon for reidentification. In order to profit from another person's reputation for trust-warranting properties a mimic may actively try to impersonate them.

2.3 Signalling Games

Gambetta and Bacharach think that primary and secondary problems of trust often go hand-in-hand, and that the latter is often the one that needs to be solved to take care of the former. In order to delve into the secondary problem they recast the mimic-beset trust game into a signalling game. They also expand the simple signalling game to include second-order signals, namely identity.

The building blocks of signalling theory are signals and three types of players. Signals are any actions performed by the sender of the signal that are meant to change the receiver's perception of the probability of a certain event. The three player types are owners of a krypton (here K), non-owners of a krypton (here $\bar{K}$), and receivers seeking to determine which is which (here V, for verifier). It is assumed that V will try to separate K from $\bar{K}$ since responding solely to K is though best for V. The main result of signalling theory is that a signal s can convey a certain level of truth if is too expensive for some $\bar{K}$ to display and cheap enough for more K. A signal can be displayed when the expected benefit exceeds the cost. When all K's *can* emit a signal and all $\bar{K}$'s *cannot* emit a signal, a *sorting equilibrium* is present since V can correctly distinguish all players. Where V's predictions are correct only intermittently a *semi-sorting equilibrium* is present, as is the case in most animal and human system.

Gambetta and Bacharach identify the relationship between signalling theory and the trust problem as follows: 'mimic-beset trust games are instances of signalling games'. They take the display of m to be the signalling of s, which E (K or $\bar{K}$) may or may not display. By drawing these parallels they upgrade the mimic-beset trust game by identifying when mimics will display m. In the presence of a semi-sorting equilibrium some $\bar{K}$ display m. Signalling theory is portrayed as 'a powerful tool', but there is the concession that there 'are strong assumptions about the players' background knowledge'. Signalling theory can can give insight into the amount of false signalling, under the assumptions of players' knowledge about the cost of signals for K and $\bar{K}$ and the probabilities of meeting K and $\bar{K}$.

At this point the theory enables the analysis of *idealised cases of informative and misleading rational krypton signalling*. The partial and abstract nature of the model is conceded, noting that it focuses only on cost and benefits, neglecting any surrounding structure or explanation. The rest of Gambetta and Bacharach's paper [9] makes improvement to the framework to provide a more complete theory. These are the introduction of identity and sign management practices.

2.4 Identity Signalling

The structure of the model up to this point is $m \rightarrow k \rightarrow t$. Manifesta ($m$) are displayed, and are linked to krypta (k), which may or may not warrant trust (t). Identity can be introduced into the model by viewing it as another krypton (i), signalled via signatures (g), and related to other krypta (k). This four-layered inferential structure is takes the form $g \rightarrow i \rightarrow k \rightarrow t$. Introducing identity helps players in multi-round games with rematching to have some level of confidence in a player's history. Players can build up a reputation, and can call on it by signalling signatures linked to their identity.

Gambetta and Bacharach find that identity signalling might be more efficient than simple krypton signalling since it can build on previous interactions, potentially lessening the intensity of probing. These previous interactions create beliefs in the mind of the truster, which can take two forms: trait laws and reputations. Trait laws are descriptions of how various signals of group membership and personal habits generalise to further encounters, without specifying exact values for these mappings. Reputation is associated with an experience krypton: a krypton that becomes revealed after an interaction with the possessor of the krypton. Revealing a krypton the player lead to a reputation for either K or $\bar{K}$. Reaping the benefits of reputation requires attestation of identity, and for players that do not engage in such signalling 'reputation will be stillborn'. This signalling generally takes place through the display of what Gambetta and Bacharach call *signatures*: 'a manifestum that is drawn from some fixed family, called a stock'. Since signal corruption and the ease of mimicry increase as signatures increase in similarity, a trustee should pick signatures from the stock in a way that gives players unique manifesta.

Gambetta and Bacharach see identity signalling as one instantiation of a more general maxim: to increase efficiency signallers should build on preexisting beliefs linking one krypton to another, and finding which krypton's manifesta are the cheapest and most reliable signals, i.e. the more efficient ones. The $g \rightarrow i \rightarrow k \rightarrow t$ structure contains only four layers, with one manifestum-krypton link, and one krypton-krypton link. However, many more krypton-krypton layers are possible, providing even more indirect signalling mechanism.

A set of people can signal identity through a group signature. It should be displayed by all members of the group (x), and by none of the non-members ($\bar{x}$). The advantages of sets are that the perception of having k spreads much faster due to categorical trail laws, and groups may better at self-observance, enabling greater compliance with trait laws. However, signals need temporal invariance for them to be of use, which is easier for an individual to achieve than for a group. Such coordinations become simpler in the presence of social cohesion.

2.5 Sign Management

Gambetta and Bacharach note that although signalling theory provides a framework to analyse the problem of secondary trust it is 'abstract and incomplete'. They attribute this to the absence of a concrete semiotic structure in signalling theory. They sketch such a structure as 'a relentless semiotic warfare' that consists of the scope 'for creating new signs, for discovering latent one, and for protecting signs against mimicry', as well as stratagems to get around these protective measures. Their paper closes with a discussion of how manifesta obtain their significance, a taxonomy of manifesta, and techniques of sign management, namely protection and selection.

Signalling theory cannot explain the mechanics the contraction of m and its association with k, since it cannot support real-time processes. Gambetta and Bacharach state that, while signalling theory has nothing to say about it, in most cases 'model comes before mimic'. Note that there is a case of rip-off pottery design that was later adapted by the original artist [109], which makes the question of directionality a tough call, but it seems apparent that such cases are a tiny minority. The meaning of m is seen as a temporal associative belief, that is, $m \rightarrow k$ at time t is a function of experience krypta perceived before this time. Signalling theory is static because it is an equilibrium theory, thus failing to model causal relationships, and therefore cannot provide an explanation.

The taxonomy of manifesta is based the reliability of signs based on on the can and cannot conditions. Three types of manifesta are identified: cues, symbolic manifesta, and fakeable manifesta. Cues are signals that are costless to display for K, and have a non-zero cost for $\bar{K}$. Automatic cues, or marks [42], are displayed subconsciously by K and impose even less of a burden. The effectiveness of cues depends on the cost of s for $\bar{K}$, and the presence of cues in $\bar{K}$, and needs to be evaluated by R. The use of technology has made visible some cues, while others are the result of evolutionary pressure. Symbolic manifesta are those m that 'consist of a configuration of characters, however these may be physically realised', e.g. names and logos. A disadvantage is that they are very easy to copy, while an advantage is that they are cheap to display. Their evidential value is decreased because of the threat of copying, and Gambetta and Bacharach find that 'the expansion of the scope for ultra-cheap transmission of symbol-strings is indeed a major cause of the growth of mimicry in our time'. Still, symbols are often used in identity signalling, with various sign management strategies in place that increase the cost of mimicry. Fakeable manifesta are not true manifesta but 'quasi manifesta' whereby only of the part of the object is observable. Fakeable object can be replaced by another one with the same observable component. An example of a fakeable object is a college certificate.

The protection of signals takes place by changing the can and cannot conditions of the signs. Often the cannot condition fails, and protection need to be introduced that deters potential mimics from producing or displaying the signal. Such a strategy must be affordable to K, else the can condition is not satisfied. The total cost of producing, displaying, and protecting a sign should not be too high. It is possible to choose alternative manifesta that are difficult to mimic, but this may be too expensive. A common strategy is to combine cheap manifesta with strong protection. The protection strategy can build on catching mimics in the act or on subsequent prosecution and punishment, and it can be implemented by the truster, the model, or a coalition of the two. These protection strategies are illustrated in the empirical study of criminal communication [43].

Choosing the correct signs to use is as important as protecting them, and it often involves searching or designing new signals. The model, or trustworthy trustee, can often raise the cost of mimicry by making manifesta more expensive to display or by choosing employing new manifesta. When the model is able to switch to costlier signals at low enough cost the threat of mimicry can be dealt with. Often the adjustment of one manifestum is enough to increase the difficulty of mimicry. Another technique for defending against mimicry is creating a constellation of manifesta, since the coordination cost goes up as more manifesta need to be displayed.

2.6 Mimicry Systems

The paper [42] by Gambetta that updates the framework provides a more high-level overview of the framework than the 2001 paper. Besides a few changes in the taxonomy of signs the main addition is a description of various mimicry systems. This update and the applications [44, 43] of the framework show a move away from game theoretic formalisms towards a more liberal and higher-level conception. In a review [27] of the ethnographic study of criminal communication it was found that many of the findings remained to be integrated into fundamental game theoretic models. This explains the current high-level focus of the framework.

The mimicry systems that are identified are 'mimic versus dupe via model', 'mimic versus model via a dupe', and 'mimic versus dupe-model'. In the first the mimic has no conflict with the model but uses its appearance to harm the dupe. The second employs seeks to fool the dupe in an effort to harm the model. In the last the model and the dupe are of the same kind, for example passing oneself off as a police officer among police.

2.7 Applicability

In a simple analysis of several case studies the framework was found to be applicable. It pictures the problem of trust not as one where the truster needs to determine what properties make the trustee trustworthy, but instead involves the decoding of signs in a creaseless semiotic warfare with mimics.

An ethnographic study [44] of the trusting behaviours of taxi drivers in New York and Belfast found that while taxi drivers generally approach the problem through simple risk acceptance, they are always aware of the issues and they started to approach the problem as one of trust when they found that something did not feel right. The decisions made by the taxi drivers were evident of a high level of reasoning as predicted by the framework. However, Gambetta and Hamill did not investigate whether this was a result of self-selection of learning on the job.

The ethnographic study [43] of the communication of criminals found that criminals find it difficult to communicate their trustworthiness to other criminals due to the nature of their profession. In order to enable collaboration with other criminals they need to show in some way that they can be expected to keep their side of the bargain. Once tactic for this was found to be the exchange of hostages in the form of information about previous crimes. The way that the trust problem is dealt with is through making the trust-warranting properties close to irrelevant, by creating the proper contextual factors. This elucidates the differences between risk management and trust. Risk management, and in particular assurance, can be described as a means of dealing with uncertainty by implementing countermeasures. In the case of trust decisions this uncertainty is the unknown trust-warranting properties. However, even if countermeasures have been put in place an individual still has to trust in signals that show this. Measures such as auditing and information hostages create proxy signals, in which trust is needed. Thus, the level of trusting is shifted to another level.

The relevance of the framework to the ethnographic studies may be due to the high stakes that the players are in for. In both cases they put their life at risk. Whether the framework also applies in less life threatening situation is something that needs to be investigated. This is done by dipping into the phishing literature. However, first the relation of the framework to problems of mimicry during authentication is described.

2.8 Mimicry during Authentication

Going by the framework of Gambetta, authentication is seen as a trust decision. The authenticater is the truster and the authenticatee is the trustee, and the purpose of the authentication is to observe credentials of trustee that provide information about their characteristics. Note that credentials take the same role as signals, and characteristics are equal to krypta. If the problem of mimicry can be solved then the problem comes down to access control, whereby the characteristics of the user determine what assets are allowed to access.

It is interesting to look at cryptography from the perspective of the framework. Cryptography creates symbolic signals but still allows for strongly discriminatory signals. It does not make the copying or display of signs prohibitively expensive, but instead addresses the creation of signs, which is something not addressed by Gambetta. The keys can be said to be the krypta of the trustee. Since management assigned these krypta to users this allows them to link it to other krypta. Looking at a run of a cryptographic protocol it becomes clear that, provided a challenge-response protocol is used, the signal does not depend merely on the krypta of the user, but also on the environment. Protocols provide temporal links and boundary conditions, and create a context for sign creation.

To show the relevance of Gambetta's framework to information security, an example of how it fits into online services is helpful. We will consider phishing and the traditional MITM attack here. A more thorough analysis and application of the framework is provided after the section on phishing research.

In an MITM attack the attacker engages in two parallel mimicry runs: the attacker mimics the client to the server and the attacker mimics the server to the client. Credentials of the user are traditionally just their passwords, and those of the server the use of their secret key couple to a certificate. Alternative credentials that may offer some degree of discrimination between genuine entities and mimics are the location from which the user appears to connect, the apparent location of the server, and SSL protected locally stored cookie linked to the relevant webpage.

Phishing consists of the mimicry of the sender of the message and the mimicry of the publisher of the website, which happens serially. The credentials that users look for in an email range from the sender's email address to the accuracy of the spelling. These are not perfect in their ability to filter out phishing emails, but they do distinguish them to a certain degree. When authenticating the website many users focus on the appearance of the website, and check whether the URL is plausible. Again, these are not perfect but they do filter out some mimics.

While these examples show that some of Gambetta's assumptions may not hold, they also indicate the potential of a sociological view of authentication.

3 Phishing

The feasibility of the framework described in the previous section is tested by looking at the problem of phishing. This section will provide an overview of the current phishing literature, and the next section will see how well these findings match with the framework. The topics that are covered here are phishing statistics, or the lack thereof, various findings with respect to fighting phishing, and the theoretical frameworks that currently exist to explain phishing. The main ways in which phishing is being addressed are better user education, designing systems that are easier to use, constructing systems that are more secure by default, detecting phishing emails and websites, taking down of phishing websites, and discovering of new attack methods. The following subsections will provide an overview of these aspect, and the remaining problems.

3.1 Statistics

Phishing attacks can cause emotional and monetary harm. However, accurate quantitative information is hard to come by. The actors involved in the fight against phishing seek to portray themselves in a positive light, leading to biased results [80]. The number of URLs, domains, and hosts can be interpreted in multiple ways, monetary losses may be the result of activities other than phishing, gangs are not as clear-cur as they appear to be, monetary damage is limited through recovery efforts, and measurements of take-down time depends on who is asked. Although it can be said that attaching numbers to crimes is difficult, phishing statistics are easily manipulated by interested parties. Data from the Anti-Phishing Working Group (APWG) [8] and the UK Cards Association (UKCA) [113] illustrate the difficulty of evaluating the phishing problem. These are shown in table 1, and indicate the disparity between the results of the two organisations. The contradictory trends could be an effect of different focus, but as [80] notes, they may be the result of personal interests. Banks like portraying themselves as effective fighters of phishing, by pumping up the number of attacks and playing down any damages. APWG receives it's numbers from industry feeds, which may be influenced by paying customers. It seems that the phishing industry is in an ugly political tangle.

In order to better address the problem of phishing more accurate statistics are needed. Without such numbers it will be difficult to know whether countermeasures are working. Before pumping money into systems that may or may not work, companies, governments, and individuals need a way to evaluate the performance of such tools. Further research should focus on finding reliable ways of gathering broader and greater amounts of data.

Year	APWG	UKCA
2008	104,283	43,991
2009	182,395	51,161
2010	115,921	61,873

Table 1: Number of phishing attacks from 2008 to 2010

3.2 Training

The work on user education is not very rosy. One problem with user education is the large percentage (96%) of popular websites that use confusing indicators [108]. They have http pages with https forms, lock icons as favicons, hidden location bars, mismatched domain names, or complicated URLs. Through daily use of the Internet users are being educated to ignore security indicators. Research into new ways of displaying these indicators is needed. However, making users respond to indicators is one thing, they also need to known the meaning of such indicators. Research evaluating connection security found that people have wrong conceptions [38]. These conceptions were not always more accurate in participants with a technological background. Through interviews, experiments, and having users draw their perception of security it was found that users think that there is a 'place' that needs to be made secure, e.g. the lock area. Such research can help designers and educators find the misconception of users, although this is only the first step in finding a solution to the problem. Users need to be taught to look at security indicators and they need to understand what they mean.

Some effort has gone into the design of training systems that build on insights from learning theory [71]. Two such systems are PhishGuru and Anti-Phishing Phil. PhishGuru works by sending users periodic training emails that lead them to a comic tutorial if they fall for the email, and Anti-Phishing Phil is a game where people eat worms that represent URLs. The systems were found to be about as effective as traditional methods, provided that people actually read them. Whether the effectiveness of the two systems was due to the novelty factor or the design is not clear, and worthy of further research. The results do indicate that testing education systems in a natural setting is of utmost importance: when testing the effectiveness of anti-phishing education validity needs to ensured. Traditionally the focus of testing the effectiveness of anti-phishing education has been on student populations, which make generalisation to business users problematic [11]. Another issue is that role-playing users act less securely than when the use their own passwords [102]. Pushing out education packages that haven't been thoroughly tested is not constructive. More fruitful is testing the effectiveness of publicly available training material, and finding ways to use these most effectively [11]. One side of the coin is designing effective training methods, another is developing general guidelines for appropriate anti-phishing education.

The problems with user training are adapting to changes in technology, the constant need to keep educating users, expense issues, and questions of effectiveness. Most phishing education does not reach levels that can be called 'safe' [11]. As such, other avenues for dealing with the problem should be explored. This does not mean that education should be abandoned, but it does mean that it cannot be the sole solution to the problem.

3.3 Usability

It has been argued that 'users are not the enemy' but that designers which don't understand how their systems are used in practice are [4]. An example is the effectiveness of password policies, which fail when they do not take into account the needs and limitations of users. The needs of the user should be at the centre of security design [101]. However, the knowledge of users is often limited as is their motivation. Designers need to take this into account when designing systems. It is important to understand the psychology of the user, and to evaluate various possible solutions for the phishing problem [61]. Looking at the technical side of the problem is not enough, and more attention is needed for the human side of the problem. One experiment that measured users' abilities found that users are not suspicious of personalised messages, wellformed URLs with similar semantic meaning, and phone calls [63]. Another study found that even in conditions where users expect spoofed webpages a large percentage cannot distinguish phishing websites from legitimate websites [24]. Indicators were poorly understood, not well noticed, and easy to spoof. The problem of phishing should not be approached from a cryptographic angle but from a usability perspective. More work is necessary in studying users in a naturalistic settings, as well as investigations into the role of education versus usability.

Several approaches to make systems easier to use have been proposed in the literature. Visual hashes are described as the basis of a solution for certificate verification and authentication via image recognition, since humans are bad at checking random strings and cannot remember strong passwords [92]. To make online password entry more secure a verified password windows has been proposed that builds upon a personalised background image as a security skin, which is thought to make website authentication easier for the user [23].

Various passport management systems have been built, such as the browser extensions Passpet [126] and Web Wallet [121], which simplify proper password policies.

While efforts to make systems more usable are commendable, in general trying to increase the usability of a system can be problematic because it takes a lot of time and effort, requires training of designers and engineers, and may have problems with user take-up. Nonetheless, a helpful guiding principle is making an effort to reduce the schism between the designer's and user's interpretation of the workings of a system [125].

3.4 Security by Default

Systems can be designed to be secure by being usable, but they can also be designed to be secure by default. In online scenarios this means that a reliable path needs to be created between the server and the user. The path between the server and the browser is relatively secure, with the path between the browser and the user needed most work. To create a reliable path between the browser and the user the interface needs to be designed in such a way that information is correct and unspoofable [124]. Anti-phishing systems that are secure by default are not widely implemented, but various solutions have been designed. Some of these rely on significant changes to the operation of the browser, while others seek only minor adjustments.

One mechanism for creating a reliable path is using a second browser that employs protected links which are signed by providers that have to be whitelisted [105]. Interaction happens in a stripped down browser interface that is unconnected to the normal browser. While this solves the problem, when new applications can be installed the use of dedicated applications may be more advisable, as is currently happening in the tablet and smartphone ecosystem. Another proposal is the use of a secure device that communicates with the browser [91]. The device and the server share a secret that is used to authenticate the phone, and a secure bookmark on the phone serves to authentication the server. The issues here are the security of the phone, the likelihood of adoption, and the effectiveness with respect to fraud reduction across the board.

In order to defeat keyloggers an on-screen keyboard has been proposed [120], but it is still vulnerable to screen capture malware. Another anti-malware method based on tracking the intentions of users over time to build a user-intention based access control policies [103] seems relevant to fighting phishing: it could protect users based on automatically setting boundaries building on normal user behaviour. The effectiveness of such an approach depends the willingness of designers to implement it, as well as user acceptance of such a system. A less drastic change is the implementation of a strong locked same origin policy for browsers. Using such an approach it is possible to achieve provably secure cookie-based authentication protocols by employing a cookie that signals a site's identity to the user [40]. This assumes secure hardware and software, as well as the user's ability to distinguish the cookie signals. Work on these assumptions is needed.

The broad question with respect to designing security into the system is the relationship between innovation and control, and the limitations of security. Carefully designed and controlled systems tend to be easier to secure. The distributed nature of the Internet, and the different incentives of a broad range of actors, make a common solution difficult. Still, the call [91] to decrease the reliance on the user's operation of the system is a welcome one.

3.5 Detection

From the preceding subsections it becomes that many users cannot be expected to correctly authenticate websites. One way to deal with the problem is having the system detect suspicious emails and websites so that the user doesn't have to. Many different detection methods have been proposed, which differ with respect to their algorithms, heuristics, data sources, and location within the network. Common locations for filters are the organisation's proxy and the user's browser. Filters can be place on the user's computer or on a server. Filters can be placed on the user's computer or somewhere on the network, such as a company proxy system. Today all the recent browsers have some form of anti-phishing capability.

In a test of a large range of browser-based anti-phishing tools it was found that their effectiveness varies greatly, with many vulnerable to simple exploits [127]. A combination of heuristics and blacklists was recommended for future tools. Heuristics can be used to calculate risk scores, based on factors such as domain lifetime, domain age, location, and domain search ranking [69]. The issue with heuristics is that they may lead to too many false positives or negatives, while blacklists do not deal with zero-hour phishing websites. A system based on content analysis and simple heuristic, using the TF-IDF frequency algorithm to derive search terms that are used to check whether the website comes up sufficiently high in search engine ranking, was found to have a decent detection rate, although language issues and obfuscation need to be addressed [128]. Another system uses this same technique, coupled with whitelists, human-verified blacklists, and the shingling method allowing similarity detection, giving a system based on a potpourri of techniques that leads to reasonable true positive rates and very low false positive rates [122]. Research has shown that there is potential for adding a wide variety of machine learning methods to anti-phishing tools [10]. A notable unorthodox method is the offence-centric anti-phishing technique that builds on a list of fake usernames, and interrogates the website to check for accurate responses [18]. Although CAPTCHAs would prevent such a tactic, they are not currently deployed.

Network based anti-phishing filters are not as common as host-based filters. One such system is an institution-wide, two-stage email filter based on intrusion detection methods [90]. It filters emails and then crawls the phishing website to check whether the content matches that of sites that are the target of impersonation. By being server-based more uniform enforcement and updating are possible, although such approaches are more difficult in consumer systems. For heavy load systems more lightweight mechanisms may be needed. Efficient blocking of phishing emails can take place using simple heuristics such as the presence of host obfuscation [46], although phisher may be quick to adapt to such measures. An alternative to outright blocking is installing a sanitising proxy system, which works by checking for suspicious websites, and stripping all content from insecure webpages what may allow users to submit information [79]. This may be more effective in teaching users about the dangers of phishing websites, although this need further investigation.

While a wide variety detection methodologies have been proposed, and continue to be proposed, they also keep getting subverted. It is important for anti-phishing schemes to not just withstand current attacks, but to be designed with an eye for the future [37]. Research looking at web malware delivery found that attackers are quick to adapt [98]. As more and more countermeasures are adopted against phishing attacks the ingenuity of attacks can be expected to rise. In order to prevent an expensive game of incremental security improvements where designers are always on the defensive, designers need to sidestep the arms race by a sudden extensive increase in anti-phishing technology [91]. However, many users may not have access to accurate filters, or these may be subverted through spearphishing attacks, and as such alternative means to fight phishing should not be neglected.

3.6 Takedown

Although takedown is often perceived as as a game of whack-a-mole, research has shown that it does reduce damages, even if the removal happens slowly [82]. Because there is a window between phishing pages going live and being taking down, takedown is only part of the solution. Some service providers were found to have a faster response than others and some brands are more successful in fast takedown. It seems that for the larger part the effectiveness of takedown does not depend on the technologies used by phishers (such as fast-flux networks), but instead is determined by how the responsibility for issuing takedown request is distributed and what the incentives are for organisations to allocate appropriate resources for getting websites removed [84]. Legal frameworks, content types, and attack methods were found to have limited impact on takedown speed. For example, the takedown speed of child pornography is extremely slow because the responsibilities are divided over national boundaries, even though it has been universally criminalised. Fast-flux was found to play less of a role than incentives, looking at the long life of online pharmacies versus the short life of phishing websites.

When researching phishing various ethical dilemmas pop up [32]. In takedown research the questions are: do researchers have a reporting duty, should they notify victims, is it permissible to fabricate content to conduct 'pure' experiments (not merely observational experiments), should researchers collect world-readable data, how should criminals deal with the potential assisting of criminals with their analysis, should the investigatory technique be revealed (there is a scientific need for repetition but there may be a need to keep things secret which may hamper publication), when should data sets be made public (greater knowledge sharing, but also with phishers) or kept secret, is the fix realistic and does it consider the incentives of the participants (the market can be a beast), and what if the fix is worse than the problem (e.g. how does it impact free and open society)? Although more information is necessary about how victim selection takes place and how phishing websites could be disrupted [81], research needs to take place within an ethical framework. Unfortunately such a framework still needs to be agreed on. While the ethical issues have not been hammered out alternative methods may need to be investigated, such as poisoning the databases of phishers with fake fingerprinted credentials, which are used to observe their actions once they use these to login at which point they are directed to a phoneypot [41].

The effectiveness of takedown has been identified in fighting phishing. However, better incentive structures may be needed to decrease takedown time further. Effective research into takedown, and phishing in general, is hampered not just by ethical issues, but by the unclarity regarding these issues. More thorough online research into takedown will require both clear guidelines as well as innovative research methods.

3.7 Attacks

Much of the research into phishing has been directed at methods to fight phishing. A smaller amount of effort has gone into discovering new kinds of phishing attacks. While this can be argued to help with timely implementation of countermeasures is tantamount to opening Pandora's box. The debate over full-disclosure has been raging for a long time, and camps have formed on both sides of the rift. As with the issue of full-disclosure the question of morality probably needs to be analysed individually, since contextual factors play a big role. Another aspect that descriptions of phishing attacks need to deal with is the balance between malicious use of information and the legitimate academic inquiry.

Detailed coverage about various ways in which phishing attacks can be carried out are available [12]. Those looking for good examples of phishing websites can visit PhishTank (*www.phishtank.com*), a collaborative clearing house for phishing websites. Descriptions of the general tactics of phishers are widely available, as are more advanced attacks. Academic papers have been published that detail DNS cache poisoning attack on anti-phishing tools [1] and 'social phishing' attacks that makes use of social network information for personalising messages [59], among others. A distributed phishing attack has been described whereby the location of sites hosting the content is personalised per-victim [64]. This limits the effectiveness of takedown, which is argued to be the most effective defence today. The attack utilises cryptovirology, public key steganography, and covert broadcasts for later reconnaissance by the phishers, by having the cryptotrojan publish the loot on bulletin boards. A solution has been presented which requires client-side ISP to block fraudulent web hosts [112]. General tactics, such as the effectiveness of using search engines for compromising and recompromising host [83], have also been described in academic papers.

Generally academics publish attacks together with proposed countermeasures. However, such countermeasures may not be achievable by everyone affected by an attack. Also, attacks can be published with different levels of detail and ethical involvement. Efforts to improve online security are laudable, but there is a big difference between raising the potential of denial of service attacks of information visualisation tools [20] and describing and carrying out a phishing attack [59]. Analysing a system from the point of view of an attacker can provide valuable guidance in making systems more robust, but care needs to be taken that such forays do not result in excessive adverse impacts.

3.8 Analysis

The analysis of phishing phenomena takes three forms: real-world investigations, computer models, and theoretical frameworks.

There have been investigations into how phishers operate, finding that phishers may host websites on residential machines, use free services such as hosting URL shorteners, and that the structure of phishing URLs differs from the average legitimate URL [76]. The market for phishing has been analysed, and found to be have a low barrier of entry and low regeneration capacity, leading to the problem of the tragedy of the commons [55]. Payoffs for most phishers decrease until they payoff of each phisher is equal to their opportunity cost, giving low skill, low reward business, the impact of which has been overestimated because mythical numbers have a tendency to be repeated when there are forces at play that benefit from the exaggerated perception and no actors with an interest in accurate numbers. When analysing the problem of phishing it can be helpful to look at the problem in terms of protocols [19]. This can show simple changes that make the job of phishers much more difficult. Another important aspect is to know the purpose for which authentication takes place, otherwise reasoning about proper authentication mechanisms is impossible [75].

The distributed nature of the Internet makes microeconomic analysis and game-theoretic analysis as important as protocol analysis and cryptanalysis [85]. There is a need to understand the economic incentives of layers for a better understanding of cybercrime. Various computer models have been let loose on the problem of phishing. An analysis of the tactics of malware cleanup found that a coordinated response by top players is more effective than a larger random selection of agents [56]. A model of security decisions with respect to security investments showed that central planner led security investments were not always higher than those due to individual choice [50]. In an evaluation of the effect of information policies on users found that expert users never provide a positive improvement to system-wide security, and that benevolent agents are needed for strong security [51]. Observations of predicted and actual user behaviour found that traditional weakest-link games differ from weakest-link security games, which shows that care should be taken when applying models from another discipline [49].

There are only five theoretical models of phishing. These will be shortly described here. A more general model of trust in information security will also be covered, since it is the only one that takes inspiration from the work of Gambetta, but even then it does not get to the core of the matter. The claimed goal of many of these models is to provide a structure in which to understand the findings of other research, and to enable quicker integration of new results.

A *graph-based model* very similar to a generic attack tree (e.g. see [48]) has been proposed as a way of reasoning about phishing attacks [60]. Pieces of information and equipment are represented as vertices, and the means to get at them as edges. An attacker starts at one vertex, and needs to traverse the graph to another specified vertex to succeed. The contribution of the model is the identification of linked graphs between multiple people, which allow contextualised attacks, as a result of the connectivity of credentials. The model was used to derive and test a phishing attack on eBay. The limitation of the model are the absence of any dynamics in the graph, and lack of an overarching theory. Network theory [87] may be a way to address these points.

Most trust research focusses on making people more trusting, guided by marketing interest, not enabling better trust decisions. Previous papers focus on trust-warranting properties, and do not take into account the possibility of mimicry. A *signalling theoretic model* has been described that explains trust as characterised by signals mediating underlying variable and malicious entities that seek to manipulate these [70]. Signalling theory is seen as a framework that can provide helpful insights into situations marred by incomplete information, although the authors do not support the rationality assumptions of signalling theory. The focus of the associated research is the difference between experts and non-experts. Attackers are entities that can influence signals. Signals are the information available to users, which can be informative about the state of the world or not. Experts are expected to have better perception of signals with more meaningful signals and less misleading and missed signals. A problem with the application of the framework is that it focusses too much on the a binary notion of expertise. Interviews only give a perspective of the practices experts preach, not those that they engage in. Mistakes made by users with more expertise may be just as intriguing as those made by unexperienced users. Another limitation is that the spoofing of signals is misrepresented, since it does not consist of control of signals by an attacker, but the sending of different ones. Interestingly enough this model does not build on the theories of Gambetta, which would enrich the framework.

A user-phishing *interaction model* has been described that looks at the problem of phishing from the perspective of decision making [28]. It is used in analysing how users detect attack and where they fail. The observations that the authors make are: more focus should be placed on designing the initial communications since this is where users pay most attention and are most suspicious: designers should identify the chance of the information they display being spoofed and think of avenues for dealing with these vulnerabilities, and user education should emphasise perception over detection, since this is where most problems appear to be, such that users should be educated in selecting the appropriate information in different contexts, interpreting the selected information, and developing a better feeling of what to expect from the interaction with the computer. Other observations are that the mismatch between users' mental models and actual implementations needs to be decomposed further to enable useful analysis. The construction of perception suggests two mismatches: the perceived participant not being the actual participant, and the perceived consequences not being the actual consequences. The inability to discover mismatches often has it's roots in the formation of false perception. often the problem is not that the victim's actions are not rational, but that they build on a wrong view of the situation and the solving of the wrong problem. It is noted that: 'the victim's response is flawlessly rational according to the perception'. the formation of a false perception is identified to be the result of: selecting insufficient (meta) information to construct an actual perception, incorrect interpretation of the selected information, and inaccurate expectations leading to biased information selection. The assumption is that attackers can only directly affect the information displayed by the user interface. This may be a very dangerous assumption, and shows the importance of keeping good track of the assumptions made in a model, and the need to continuously reevaluate them. Attackers could change the environment of a user.

Subverting the user is not studied as extensively as subverting the system, although they often play a role in security compromises. A *threat model* has been proposed that seeks to offer a systematic analysis of the user's role in authentication [29]. It consists of three steps: describing the nature of credentials, determining the expose level of these credentials, and finding vulnerabilities of entry points. The properties of credentials that are looked at are: mode, factor, assignment, and losability. The effectiveness of passive attacks was found to depend on the exposure of credentials to third parties, while the vulnerability to active attacks depends on how likely users are to be critical when asked to present their credentials. The model also introduced the concept of a security dependency graph, although it is more high-level and less detailed that that in the graph-based model [60]. The security of the system depends on the weakest cluster of machines that allows reaching the target. Targets of impersonation can be external entities with which the user has shared credentials, or actors within the security dependency graph. Finding vulnerabilities of entry points starts with identifying all targets of impersonations, and is followed by determining when credentials are exchanged or modified, these states or transitions are the entry points. The questions to be answered are: where do transitions happen, over what medium, and by whom? Once the entry points have been determined the external entity authentication vulnerabilities are analysed, and may be: no reliable and sufficient authentication information provided (to users), users' lack of knowledge, and security design assumptions concerning users that do not hold in practice. To check the reliability of authentication information look at the protocol for the credentials, identify important agents, determine the protection level of the agents, and determine the likelihood of credential compromise. In all these stages it is important to analyse the plausibility of assumptions made when the system was designed. These assumptions need to be correct to ensure a system that is functional, and since the environment changes such assumptions need to be regularly reevaluated. While the model can provide a good way of identifying problems with credentials, it does not explain the social processes that underlay authentication. An aspect that can be improved is looking at the processes that explain why and how credentials become socially accepted.

A *behavioural model* of phishing has been proposed that seeks to explain trust behaviour as resulting from non-associative model of learning (habituation and sensitisation) and locates this behavioural activity in a broader psychological model [119]. In the model the establishment of trust happens through habituation, while maintenance of trust is based on sensitisation to malevolent events. The gradual reduction of distrust is an example of habituation, which results from repeated harmless or beneficial exposure to a stimulus. The distrust curve depends on contextual factors such as possible monetary loss, but starts of at the maxima of distrust during the initial exchange. When trust is broken through direct or indirect experiences the behavioural response can be modelled using the mechanisms of sensitisation, which allows quick elimination of habituation through the presence of an aversive stimulus in place of the expected stimulus. Sensitisation predicts that victims will redevelop trust when provided with positive interaction. This does not work in all situations however, since cognitive processes may take over, potentially leading to a general aversion of the system. The challenge in security systems is encouraging processing at the cognitive level while preventing user aversion. Humans engage in several levels of processing, with increased extraction of meaning as processing gets deeper. If users have become habituated, and if shallow message features appear correct, a message is more likely to be processed at the behavioural level than at the cognitive level, and sufficient depth will not be attained. This explains why even experienced users can become victims. To encourage processing at sufficient depth the system should focus users on perceptual cues before engaging users with behavioural responses. (On a side note it may be observed that the focus of advertising can make this difficult.) To make sensitisation events more effective multiple sensory inputs can be used. The mechanics of habituation and sensitisation have been accurately fitted to real-world data in the natural world. Several aspects of these mechanisms may be relevant to human problems. To determine the relevance of biological principles to security decisions further experiments are necessary, so that situation where habituation and sensitisation are relevant can be determined. Also, the model needs to be extended to include multiple parties. Because the behavioural level is just one within the psychological stack, future work needs to look at the interplay with other layers, such as the interaction between cognitive and behavioural processing.

The *trust-warranting properties model* builds on Gambetta's as a good source of inspiration, but arguably does the exact opposite of what Gambetta sought to do. It describes what the trust-warranting properties are from a global perspective, while Gambetta sought to stress that trust problems are often not the result of the primary trust dilemma, but instead result from the secondary trust dilemma. It is noted that all trust is mediated by signals not much more than this is covered except the note that mimicry *may* happen, but it is stressed that a focus on trust-warranting properties will be likely to help designers. The argument is that the focus should not be on identifying trusted entities, but trying to design trustworthy behaviour into the system. The attained design may not necessarily be robust against well-equipped attackers, but it should enable more trusting online communities. This approach is helpful in online communities, but should be done in combination with looking at attacker strategies and opportunities for impersonation.

The main insights from the models that are discussed above can be summarised as follows:

- credentials can be represented on a graph, illustrating their exposure;
- analysing the nature, exposure, and entry points related to credentials can provide a good idea of the threat;
- signalling theory can serve as a framework, while neglecting assumptions;
- user behaviour can be rational in the face of wrong perceptions;
- processing at the behavioural level can prevent more accurate analysis of potential phishing at the cognitive level; and
- Gambetta's framework has not yet been used in the information security field to study the threat of phishing attacks from a mimicry perspective.

These findings are useful when constructing other models of phishing.

The analysis of phishers, surrounding processes, and underlying mechanisms provide a necessary foundation from which to approach the problem. A word of warning is that understanding through models and descriptions is only useful if it enables more efficient countermeasures. Results need to be put to work if they are to have any impact.

3.9 Looking Forward

After studying the literature for a wide range of responses to phishing attacks it was found that one is not more useful that the other, but that a combination of countermeasures is needed if the threat of phishing is to be addressed successfully. The findings can be used to identify the relevance of Gambetta's theories surrounding mimicry. This is the focus of the next section.

4 Revisiting Gambetta

The phishing literature provides a lot of results that can form the basis of a reevaluation of Gambetta's framework. In the previous section there is ample evidence that most of the assumptions that Gambetta makes do not hold true. The finding that phishing statistics are hard to come by invalidates the assumption that users know the likelihood of running into a phishing and the potential consequences. Usability evaluations showed that many users do not spot trust indicators, invalidating the assumption that users know the signals that are available. The need for systems that do not rely on the user was observed, since they tend to make errors and are unmotivated about security. Also, users did not always process information at the cognitive level, and instead made behavioural decisions.

These findings are rather ravaging for the framework. Although it works in situations where the players need to be extremely rational because of environmental pressures, it does not seem to be applicable to the character of the Internet. However, we feel that one of the cornerstones of the model still stands, namely that trustworthiness is not directly observed, but instead mediated by signs. While it has been observed that frameworks with non-functional assumptions can still be used in reasoning about a problem [70], we feel it is appropriate to upgrade the framework in the face of the research into phishing, and some other sources.

4.1 Revisiting Rationality

The security behaviour of users has been both called rational and irrational. Rational because if users spend a significant amount of time checking trust indicators this would lead to more productivity loss than the yearly damage due to phishing [54]. Irrational because the people often lack the appropriate information and tend to trade long-term privacy for small short term gain [2]. How rationality is defined will lead to different views of the problem.

Research into privacy decisions of users has shown that users are decidedly non-rational when weighed against objective measures. Behavioural biases, ambiguity, and uncertainty play a role in causing irrational user behaviour [3]. Another list of causes is incomplete information, bounded rationality, and psychological deviations from rationality [2]. When privacy concerns are made salient to users, they are less likely to divulge personal information, even if the risk would be low, while suppressing privacy concerns leads user to divulge information even if it is not in their best interest [66]. In a test of the perceived trustworthiness of various forms of media representations of experts it was found that users have a media bias, preferring audio and video over text advice [99].

There has been progress in economics away from the purely rational-choice models that are appealing but descriptively incomplete, towards more thorough conceptions such as procedural (or bounded) rationality [15]. This has been accompanied with a lot of scientific evidence that refutes various of the assumptions made in rational choice theory. The theory of bounded rationality seems to offer a variable alternative to rational-choice theory. Bounded rationality finds that individuals are limited in their rationality though intuitive actions and and the impact of a specific cognitive focus [68]. Humans do not tend to optimise, and instead satisfice or seek outcomes that are good enough [104]. Rationality is bounded by limited information and limited processing capacity. Instead of the assumptions of rational-choice theory it seems that models of phishing must work with theories such as bounded rationality.

4.2 A Tentative Model

Apart from the mediating function of signs in trust decisions, the framework of Gambetta needs a lot of corrections. The adjustments presented here are based on insights gleamed in the previous sections of this paper.

The ability to display a credential depends on the individual's link to the credential. Those credentials that have links to many people have an increased risk of misuse. Users may traverse the credential graph through authentication in order to create links to other systems and credentials. In order to choose secure credentials an analysis can be made into the properties and entry points of the available credentials.

The effectiveness of signals mediating trustee characteristics is determined by the design of the system One aspect is the ability of a system to make its own decisions so that it may filter out mimics before they can communicate with the user, while the other aspect is correctly conveying credentials to the user. The system thus serves as a filter that can remove both communication from mimics as well as valuable evidence that a mimic is on the loose. Another filter is the perception of the user, which can often be wrong. When the user has the wrong perceptions they may employ phantom credentials, that do not provide any evidence pertaining to the trustworthiness of the the trustee.

Users process phishing messages on multiple levels. The behavioural level depends on conditioned responses, while the cognitive level involves the application of bounded rationality. Displaying the wrong behavioural cues can lead the user to process the message at too shallow a level. Even when processing happens at the cognitive level users are liable to make mistakes. As users are limited in the amount of information they can process, their ability to process credentials is expected to drop as the number of credentials grows. Equation 7 shows such behaviour for positive values of x, although the function of the equation is purely illustrative. Such descriptions can also be constructed for the other processing layers through which credentials pass.

$$\text{percentage of credentials observed} \propto \text{nr of credentials displayed}^{-x} \quad (7)$$

The tentative model presented here shows that credentials are not objectively analysed, but instead they go through several filter layers, namely the security layer, the usability layer, the perception layer, the behavioural layer, and the cognitive layer, before a decision can be made as to the trustworthiness of the trustee. By focussing on the signals that mediate trust-warranting properties the model looks at the authentication behaviour of users in the face of mimicry. This allows more of a focus on the human element of the authentication, which may provide more secure systems where humans are the weakest link. The next section will apply the framework to the problem of browser interact design in the hope of finding potential solutions.

5 Finding Solutions

Based on the tentative model in the previous section it is possible to evaluate systems to look for possible improvements. This section provides a look at the design of some parts of web browser interfaces, as well a quick rundown of some promising avenues for further research.

5.1 Cognitive Walkthroughs

The analysis process is explained with the help of cognitive walkthroughs. These are a way of imagining the thoughts and actions of users when they use a system for the first time [73]. It works from a specification of the prototype and the intended users, a description of the tasks to be carried out, and a list of actions needed to do so. When doing the cognitive walkthrough a plausible story is generated about each of the user's tasks, and all of the user's actions are motivated. The questions have been proposed [73] for evaluating the user's actions are:

- Will users be trying to produce whatever effect the action has?
- Will users see the control (button, menu, switch, etc.) for the action?
- Once users find the control, will they recognise that it produces the effect they want?
- After the action is taken, will users understand the feedback they get, so they can go on to the next action with confidence?

A hypothetical user is walked down the list of actions to be completed, and their likely feedback is evaluated. When they are expected to get stuck this is recorded and the tester continues as if the user had completed the task correctly to evaluate the other actions on the list. The method is regarded as a good way of quickly identifying usability problems with a product.

5.2 Web Browser Interfaces

The security behaviour that is investigated here is the simple act of verifying the website publisher's identity. The *product* is the Firefox 6 web browser, *users* are taken to behave as has been found in previous sections, the *task* to be carried out is verifying the authenticity of a webpage, which consist of the actions of checking whether SSL is enabled, parsing the URL, verifying whether it matches with the user's expectation. In Firefox 6, which was released this summer, the variably highlighted URL bar was introduced. The base of the URL is now black,

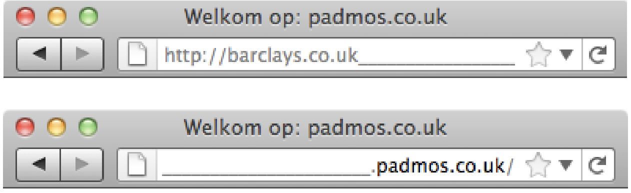
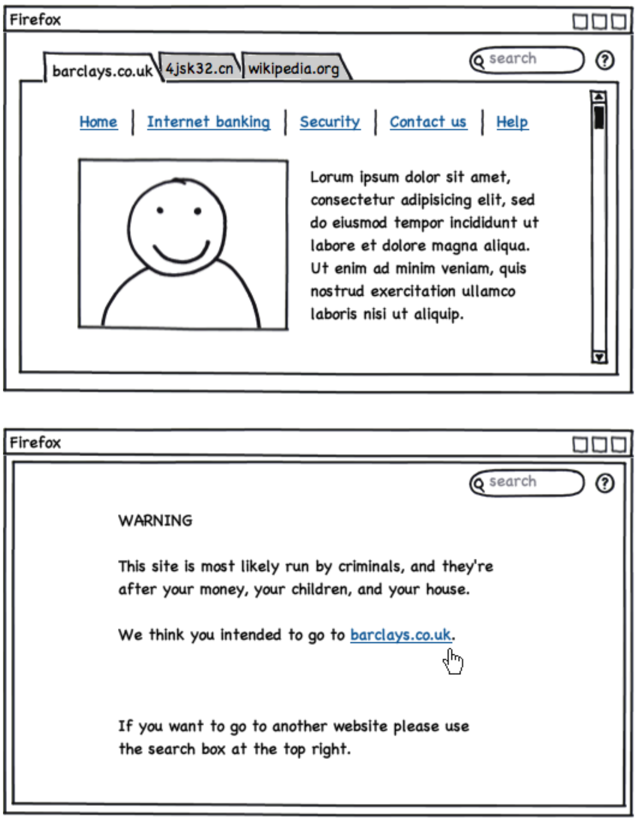


Figure 2: Example of how Firefox 6 handle URLs

with the rest of the URL greyed out, as shown in figure 2. The latest versions of Internet Explorer and Chrome do a similar thing, although Chrome does not grey out subdomains. Whether these measures help prevent phishing attacks is investigated through a cognitive walkthrough of the Firefox 6 interface.

The first action of the user is checking whether SSL is enabled on the webpage. To do so the user has to observe the URL bar and check whether Firefox displays the base address bar in blue or green dedicated box. Various papers discussed in the section on phishing found that many users do not notice when the SSL notifications are absent. An explanation of this is that the credentials are filtered by the perception layer since they are at the periphery of the user's vision. Also, users may mistake a favicon for security indicators, not be interested in checking the validity of the website, or not understand the concept of certificates. If users do notice that SSL is enabled by seeing the blue or green box in the URL bar they need to parse the URL. Users may not parse the URL properly because they don't give any or enough attention to it or because they don't understand the concept of a URL. This can be explained by users being stuck at the behavioural layer, and merely reacting to erroneous impulses, or as a result of cognitive processing that relies on incorrect information. The last action that the user needs to do is to verify whether the parsed URL corresponds to their expectations. Many users have been found to accept URLs that are semantically similar to their expectation as correct. Such users do not produce the intended effect of the action. Another problem can be that users neglect to check whether the URL matches at all. Such user responses can be the result of the bounded rationality when processing at the cognitive level, or compulsive behaviour at the behavioural level.

From the cognitive walkthrough above it becomes clear that domain highlighting may not offer much solace for users of Firefox 6. While it may make the action of URL parsing a little bit easier for the user the task of identifying websites is marred by difficulties. A design for a browser interface is proposed that makes the action of URL parsing close to irrelevant. Figure 3 shows the interface where the URL bar is removed and tabs are labeled with the base domain name. Entering a web address happens in the search box, which helps prevent



Lorum ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip.

WARNING

This site is most likely run by criminals, and they're after your money, your children, and your house.

We think you intended to go to barclays.co.uk.

If you want to go to another website please use the search box at the top right.

Figure 3: Mockup of the proposed browser interface

landing on a wrong website due to typos. When information is presented to the user that can be controlled by an attacker extra care needs to be taken [20]. The website title is not displayed in the browser windows in order to take away the ability of phishers to display information in any area besides the web page. This was also the reason for replacing the tab title with the domain name. The reasoning behind this is that the model described in the previous section suggest that number of credentials that users pay attention to decreases as the total number of credentials increase. While the title of the webpage is a very unreliable credentials there may be users that incorrectly rely on it. Additionally it detracts from the most relevant credential, which is the presence of SSL along with the correct domain. The reasoning behind the choice of not displaying the full URL is that many users do not have sufficient information to make proper choices at the cognitive level, or may default to compulsive behavioural responses because of the cost of parsing URLs.

The actions that are not addressed by the mockup, and that remain problematic, are determining SSL status and domain matching. Many users focus solely

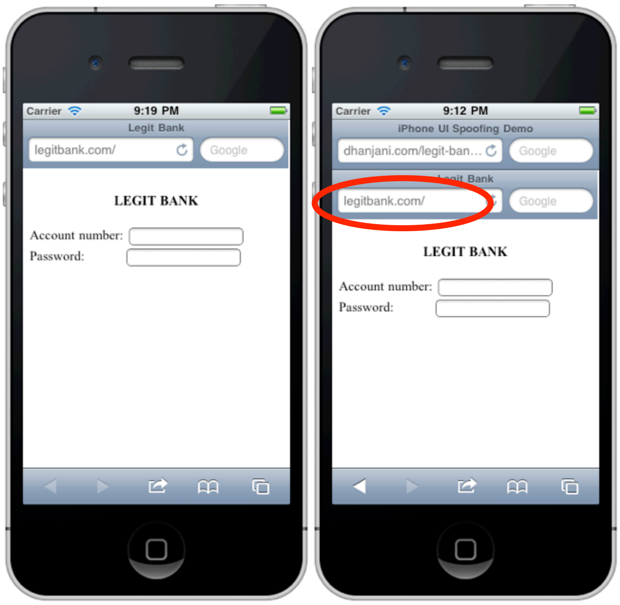


Figure 4: An example of a spoofed URL bar on the iPhone [26]

on the appearance of a website [74], which is understandable since it takes up more than 90% of screen real estate. To deal with this issue more research is necessary into creating signals that are likely to be perceived by the user. The problem of domain matching could be addressed through a more draconian framework for domain registration, such that semantically similar domain names cannot be registered. The impact on freedom of expression and associated costs will need to be investigated. A minor issue in the mockup is deciding how to deal with extremely long domain names that do not fit within the tab width. Also note that the security indicators are not specified. These could be a green tab colour, or whatever other signals are found to be effective. Lastly, the privacy issues that come up due to the use of search engines as intermediaries for URL entry needs to be addressed. One solution may be comparing URL to the local history, and only checking with the search engine when necessary.

Based on the findings of this example we feel that cognitive walkthroughs, coupled with insights from the framework described in previous sections, allows for a thorough analysis of system security and the finding of countermeasures. Many systems seem to be able to benefit from such an analysis. One example is mobile browsers, where the URL bar can be hidden using simple JavaScript [26]. This vulnerability was presented back in 2010 [25], but is still present in many browsers [117]. An example of a spoofed URL bar is presented in figure 4, which shows the danger of mimicry in the face of problematic design.

5.3 Other Avenues

Besides cognitive walkthroughs building on a conception of the user as proposed in the tentative model, solutions to the problem of phishing might be found in the analysis of human protocols, the creation of signals that discriminate between computers and humans, and evaluating why human are successful at the use of biometrics.

Currently information security is quite focussed on dealing with bit and bytes, with society often located outside the research focus. The problem with this is that the traditional focus of security is too narrow, since the larger world has been dealing with security questions long before computers and computer networks appeared [13]. Human-scale security protocols (HSSP) are seen as a viable research area that would allow designers to take input from non-traditional protocols to strengthen information security protocols. The high-level security protocols of human-to-human interaction have evolved over many centuries to deal with human problems. HSSP are relevant because they are often optimised for efficiency and performance, even though they tend to operate in an ad-hod fashion. While traditional and HSSP studies may pollinate one-another, HSSP are said to have the greatest prospects since they are human systems that have evolved over time and are likely to be well optimised and tuned to risk.

By discriminating between human and computer users the cost of an attack could be increased. Research into human interaction proofs (HIP) has sought to find ways of distinguishing between humans and computers. Traditionally the focus was filtering bots, but a new class of HIP allows human to distinguish computers [22]. It was found that HIP can be used to secure human-machine interaction where an untrusted intermediary is present [77]. If this untrusted intermediary is a machine it will not be able to read the exchanges information; if the untrusted intermediary is a human then listening in on this channel will be more expensive than automated wiretapping. These tools are most relevant in cases where transactions are completed since humans may later learn the values that were exchanged. Nonetheless, it is an interesting approach for some problems. An issue that needs to be kept in mind is how disabled users can make use of the system [78]. Also, like CAPTCHA generation can be provided as a web service [21] (e.g. reCAPTCHA), solving CAPTCHA and other HIP may be turned into a service built on outsourcing to low-wage labour countries.

Biometrics is sometimes referred to as the odd one out when it comes to authentication, since it is much more fuzzy than cryptographic algorithms and protocols. However, humans have used biometrics since the start of time for doing authentication. Maybe inspiration can be taken from the general successfulness of biometrics in human protocols. We propose the term *technometrics* for signals that represent properties of a technical system, and that are easy for humans to perceive and verify. Random Art [92] is one such technometric that makes use of the visual cortex, but possible alternatives that come to mind are soundscapes, electrical currents, and smell creators (for the deduction of cheesy terminology relating to phishing the reader is left to their own device). Computers have various sources of distinctiveness that can serve as the basis to technometrics. A recent example is the possibility of extracting device signatures from NAND flash storage [96].

The opportunities for making mimicry difficult by increasing the discriminatory potential are vast. Of the research that has been done in this area, most is dispersed across subfields. In order to deal problems of mimicry in information security such as phishing it may be advisable to tackle the problem in a more structured manner. A possible banner under which to place such an effort is 'mimicry during authentication'. Because the goal of such research would be looking for discriminatory credentials the ethical issues associated with this cannot be ignored, and are thus the topic of the next section.

6 Ethics

It has been established that credentials enable discrimination between authentic entities and mimics to various degrees: credentials are discriminatory. This has inherent ethical implications, which is raised by Gambetta and Hamill but not addressed in much detail [44]. They argue that discrimination by their sample of taxi drivers is ethical because of the grave threat that the drivers face and the effectiveness of discrimination. There is the potential for those who look suspicious to adapt their behaviour, but there are limits to this. In information security the threat is also there, but ethical issues become much more pronounced because the threat to life is often not there and the issue may be more visible. Whether such dilemmas are a result of a difference between the online and offline world is not too clear, but most likely this is not the main differentiator. Instead, the use of information technology has allowed far more automation with respect to profiling, categorisation, and sharing as a result of more processing power and greater connectivity [114].

6.1 Biometrics

A class of credentials that seems especially problematic is biometrics. While cryptography may link bits to individuals, these bits have much less baggage associated with them than biometric credentials. Technometrics seem less problematic because computers are not granted rights to privacy. Whether this will change when artificial intelligences approach human capabilities is an interesting question, but for now not relevant. A more pressing question is whether users are able to use systems that deploy technometrics. Disabled individuals may not have the cognitive capabilities to register signals sent by a machine: e.g. blind users cannot make use of visual hashes.

One of the main issues with the use of biometrics is the informatisation of the body [115]. The camera captures someone's appearance, the phonograph their voice, and DNA sequencing their genes. When computers entered the scene, and things became digitised, traditional boundaries no longer worked, and people's informatisated bodies could escape and roam cyberspace, separated from their worldly shells. Before long digital body parts started being traded like commodities, which leads to questions of human dignity: a core concern with the use biometrics [86].

Providing a biometric means giving over 'a piece of ourselves' [6]. As control of biometric information is relinquished to varying degrees, privacy problems start to pop up. Currently, many users are not fully aware of the risks and vendors are too dismissive of privacy concerns, which has led to calls for the need of free and informed consent [6]. However, this has been argued to be impossible, due not only to the limited understanding of individuals, but also due to the fundamental opaqueness of future scientific developments [115]. Consequently, for the informatisation of bodies to provide individual and communal benefit the discussion about biometrics would have to move to the level of technical infrastructures and intermediate institutions.

Mouse movement based authentication systems have been proposed that have decent detection rates [97]. However, the effectiveness of mouse based authentication has been questioned, noting that the low error rates are likely due to loosely controlled environmental variables [67]. Even so, there are measurable individual differences, and mouse movements can give information about the user's psychological state, and properties of the software and surface texture. Although mouse based authentication may not be terribly effective at authenticating individuals, it could provide a technique for companies to filter out potential clients that pose higher risks, such as those with Parkinson's disease.

Research has shown that potentially sensitive information about a woman's sex life can be derived from recordings of their gait [88]. Note that this is preliminary research, but the mechanism behind it is supported by findings in other studies. Similar research based on the shape of the upper lip has also been done [14]. These results raise interesting questions about data collection and retention, and shows that the use of biometrics may leak more information that typically assumed. The use of biometric information, even when it does not at first sight pose any ethical dilemmas, needs to be extensively investigated if privacy is to be safeguarded. Where consequences are unclear it may be better to rely on the precautionary principle [93].

It's safe to say that not even half the kinds of information that can be gleamed from people's behaviour and bodies are known. In light of this there needs to be a serious discussion about what biometrics can potentially disclose about individuals, and whether society is really better off when this information is floating around in the ether. Will governments, companies, and other actors choose to make use of this information in a socially and individually desirable perspective? What if biometrics allow with high likelihood the observation of traits that are undesirable in the current political system? When technologies have the potential to enable a witch hunt they should not be introduced haphazardly.

6.2 Other Credentials

Discrimination based on credentials does not restrict itself to biometrics. Credentials that are not directly related to a person's body or behaviour can be the basis of selective treatment. Whether such selective treatment is ethical depends on the social and individual impact. Consider the requirement that a job candidate display a degree from a reputed university. Although it may serve as an effective signal for weeding out promising candidates, the use of this signal brings up many questions concerning access to education, appropriate spending of public money, the institutionalisation of expertise, the role of gatekeepers, and opportunities for those with alternative backgrounds. These questions are too deep to delve in here, but they are one illustration of the problematisation of the use of credentials. Less ethically ambiguous uses of credentials are biased hiring-selection based on gender cues and access control on mobile phone networks based on running through a challenge-response protocol correctly.

On the Internet a big issue when it comes to credentials is automated indexing, slicing, bundling, and reselling of personal information and related practices such as targeted advertisements, all based on identification and reidentification. The largest threat to people's online privacy and self-determination is probably the advertising industry. Credentials allow the observation of an individual's characteristics and may compromise their privacy. This is especially true when they allow the unique identification of an individuals, and even more so when this is possible over multiple encounters. By identifying and reidentifying individuals, and collection as much information as possible about their preferences, advertisers seek to provide more targeted advertisements. Going by the techniques used by advertisers they seem to be a rather unscrupulous bunch. Advertisers use many of the same techniques as con artists, which raises questions of ethics and legality [106]. Going by the tactics and half-truths of advertisers it seems that their use of credentials should be prevented.

When credentials are presented to multiple organisation (such as an IP address) trying to block their misuse is rather difficult. The question to be answered is: is there a case for the use of the credential of should it be engineered away? Taking the IP address example, using an anonymisation approach will prevent the display of the credential, but it will also prevent legitimate uses, such as the authentication check during Facebook login. If a system requires the use a class of credentials, but their abuse should be prevented, an approach is needed to prevent the display of credentials in specific scenarios. One way is the use of filters based on whitelists or blacklists coupled with an anonymising proxy. An alternative mechanism is the institution of a legal framework. However, this has enforcement issues, especially on a global network such as the Internet. Note that not displaying a certain credential can be problematic when everyone else is displaying the credential. In such a case the signals that people display need to be standardised.

It was found that the use of certain credentials is ethically charged. This depends upon the context in which they are displayed. These contexts come about from the actions of the participants as well as the designers and constructors of the system, and their interactions. How these contexts will evolve over time is unclear, and the discriminating potential of credentials could drastically change as a result of future developments, such as the discovery of better data mining algorithms. With the unclarity over the future impact of credentials it is difficult to decide whether an application is safe or not. Still, current uses of credentials can be analysed and problematised, thus providing more visibility of the issue. Also, the incentives and power relations of players can be evaluated to detect possible asymmetries that may build on the discriminating effects of credentials.

6.3 Control Structures

The discussion of the ethical nature of credential choice points towards the broader question of the ethical problems that come up in systems that are designed or operated by a minority group. Some of the issues that arise are power, racial, ethnic, gender, and expertise asymmetries. In the information security sector such asymmetries are also present. Although information security specialists should be labouring on their own demise, it seems that structures have been put in place that require continuous maintenance by a technocratic elite. Currently, there is an obsession with risk management [95] and auditing [94]. These structures result in a centralisation of control and expertise, which is symptomatic of the current state of information security.

One important point to remember is that control systems built for one purpose can become control systems for another purpose if the controller has the incentives to do so. This point becomes particularly problematic when looking at the privatisation of security companies and the supply of services and goods to suppressive regimes. Gatekeepers cannot be expected to make objective decisions, when those decisions are against their personal incentives (e.g. the social comparison bias [45]). The role of gatekeepers is worthy of a serious discussion. Power relations build, and build on, the credentials that discriminate between individuals.

6.4 Guidance

Mimicry preventing technologies are not alone in their potential to lead to discrimination and segregation of society: information security research has shown how to both cloak and fingerprint individuals, how to both hide and recover data, how to both fix and hack computer systems. The choices that researchers make have implications far beyond the lab, and they can hardly look to governments or companies to ensure that ethical principles are followed. If ethical guidelines are to be followed they need to be engineered into the system. This requirement goes further than *privacy by design* [52]: researchers should seek *ethics by design*. However, we still have legacy systems, and like the need for *privacy by redesign* [17] we should be seeking *ethics by redesign*. In order to do this the reevaluation of discriminatory potential seems a good place to start.

When deciding on credentials to use it is paramount to think of the ethical implications of one's choices. Whenever credentials are created or used that can be used for societal discrimination there is the potential for society to be worse off. It appears that something similar to a *privacy impact assessment (PIA)* [57] may be needed for credentials, which looks at the potential for credentials to ostracise individuals from society, which we refer to as a *discriminatory impact assessment (DIA)*. The illustrative equation 8 shows a way to go about weighing security against social impact, whereby the provided security is clearly shown to be adequate and the potential for distinguishing mimics from genuine users exceeds the potential for adverse societal impact. The weight and quantification of these factors are left for further debate, as is the general form of the equation.

$$\frac{\text{security provided}}{\text{security demanded}} \times \frac{\text{mimic discrimination}}{\text{societal discrimination}} \tag{8}$$

Designers have a moral obligation to design systems that do not allow cheap and easy prosecution of minority groups or control by minority groups. Looking to the market to provide imperatives to do so is foolish, keeping in mind that markets are amoral. Governments have rather spotty track records, even those exalting the virtues of human rights. Asking individuals to protect themselves is problematic when the issues are shrouded in gobbledygook. What is needed is a clear and enlightened discourse, unmuddled by financial and political interests, to determine guiding principles for the design, development, and deployment of technologies. This section on the discriminatory potential of credentials seeks to be one part of such a discourse.

7 Conclusion

The main contribution of this paper is thought to be putting the problem of mimicry on the research agenda, and describing a model that enables the evaluation of a system's susceptibility to mimicry. In the course of this paper most of the objectives that were defined at the start have been attained. The importance of mimicry was shown, Gambetta's theories have been explained, the phishing literature was analysed, a new framework based on the core of Gambetta's theories has been created and applied to the problem of browser interface design, and the ethical dilemmas surrounding credentials have been described. The objectives that remain are presenting testable predications and sketching a roadmap for further research into mimicry.

A theory is only useful if it makes testable predictions. The predictions of the tentative model described in this paper are the importance of signs in building trust in the mind of the user, the layered nature of trust evaluation, the limitations of users with respect to bounded rationality and impulsive behaviours, and the characteristic of a system as one that filters signals before they reach the user and are subject to further filtering. Future research should focus on checking the validity of the predictions of the model, since this was merely an exploratory study. Also, the model can be applied to a wide range of problems in information security where mimicry is present. Such an effort would look a 'mimicry in authentication' in its broadest sense.

System designers should use the results of this research to pick, grow, and nurture the right credentials through careful analysis, appropriate system design, and continuous system updates. If this is done information security systems can be made secure in the face of mimicry. Currently, there are many credentials that are not used in authentication scenarios, even though they can help detect mimics. These can be used by designers to build systems that are secure in the face of mimicry. The core message of this paper is that credentials enable a type of discrimination that has the potential to increase system security, but can become unethical if used outside the appropriate context.

Acknowledgements

Many thanks to my supervisor Dr. Lizzie Coles-Kemp for providing invaluable guidance. Also, I would like to thank everyone at the ISG that has made my stay more educative and engaging.

References

[1] S. Abu-Nimeh and S. Nair. Bypassing Security Toolbars and Phishing Filters via DNS Poisoning. In *2008 IEEE Global Telecommunications Conference*, pages 1–6, New Orleans, 2008.

[2] A. Acquisti and J. Grossklags. Privacy and Rationality in Individual Decision Making. *IEEE Security and Privacy*, 3(1):26–33, 2005.

[3] A. Acquisti and J. Grossklags. Digital Privacy: Theory, Technologies and Practices. In *Digital Privacy: Theory, Technologies and Practices*, pages 363–377. Auerbach Publications, Boca Raton, 2007.

[4] A. Adams and M. A. Sasse. Users Are Not the Enemy. *Communications of the ACM*, 42(12):40–46, 199.

[5] B. Alcalde, E. Dubois, S. Mauw, N. Mayer, and S. Radomirovic. Towards a Decision Model Based on Trust and Security Risk Management. In *7th Australasian Information Security Conference*, pages 61–70, Wellington, 2009.

[6] A. Alterman. 'A Piece of Yourself': Ethical Issues in Biometric Identification. *Ethics and Information Technology*, 5(3):139–150, 2003.

[7] American Psychological Association. *Ethical Principles of Psychologists and Code of Conduct 2002*. APA, Washington, 2002.

[8] APWG. Global Phishing Survey. Anti-Phishing Working Group, 2008, 2009, 2010.

[9] M. Bacharach and D. Gambetta. Trust in Signs. In K. S. Cook, editor, *Trust in Society*, pages 148–184. Russell Sage Foundation, New York, 2001.

[10] R. Basnet, S. Mukkamala, and A. H. Sung. Detection of Phishing Attacks: A Machine Learning Approach. In *Soft Computing Applications in Industry*, pages 373–383. Springer-Verlag, Berlin, 2008.

[11] E. Bekkering, D. Hutchison, and L. A. Werner. A Follow-up Study of Detecting Phishing Emails. In *Conference on Information Systems Applied Research 2009*, Washington DC, 2009.

[12] H. Berghel, J. Carpinter, and J.-Y. Jo. Phish Phactors: Offensive and Defensive Strategies. In *Advances in Computers, volume 70*, pages 223–268. Academic Press, London, 2007.

[13] M. Blaze. Toward a Broader View of Security Protocols. In *12th International Workshop on Security Protocols*, pages 106–120, Cambridge, 2004.

[14] S. Brody and R. M. Costa. Vaginal Orgasm is More Prevalent Among Women with a Prominent Tubercle of the Upper Lip. *The Journal of Sexual Medicine*, Online publication before print, 2011.

[15] C. Camerer. Bounded Rationality in Individual Decision Making. *Experimental Economics*, 1(2):163–183, 1998.

[16] C. Castelfranchi and Y.-H. Tan. The Role of Trust and Deception in Virtual Societies. *International Journal of Electronic Commerce*, 6(3):55–70, 2002.

[17] A. Cavoukian and M. Prosch. *Privacy by ReDesign: Building a Better Legacy*. Information and Privacy Commissioner, Ontario, 2011.

[18] M. Chandrasekaran, R. Chinchani, and S. Upadhyaya. PHONEY: Mimicking User Response to Detect Phishing Attacks. In *2006 International Symposium on on World of Wireless, Mobile and Multimedia Networks*, pages 668–672, Niagara-Falls, 2006.

[19] R. Clayton. Insecure Real-World Authentication Protocols (or Why Phishing Is So Profitable). In *13th international conference on Security protocols*, pages 89–96, Cambridge, 2005.

[20] G. Conti, M. Ahamad, and J. Stasko. Attacking Information Visualization System Usability Overloading and Deceiving the Human. In *Symposium On Usable Privacy and Security 2005*, pages 89–100, Pittsburgh, 2005.

[21] T. Converse. CAPTCHA Generation as a Web Service. In *2nd International Workshop on Human Interactive Proofs*, pages 82–96, Bethlehem, 2005.

[22] R. Dhamija and J. Tygar. Phish and HIPS: Human Interactive Proofs to Detect Phishing Attacks. In *2nd International Workshop of Human Interactive Proofs*, pages 127–141, Bethlehem, 2005.

[23] R. Dhamija and J. Tygar. The Battle Against Phishing: Dynamic Security Skins. In *Symposium On Usable Privacy and Security 2005*, pages 77–88, Pittsburgh, 2005.

[24] R. Dhamija, J. D. Tygar, and M. Hearst. Why Phishing Works. In *CHI 2006 Conference on Human Factors in Computing Systems*, pages 581–590, Montreal, 2006.

[25] N. Dhanjani. UI Spoofing Safari on the iPhone. *http://www.dhanjani.com/blog/2010/11/ui-spoofing-safari-on-the-iphone-.html*, November 2010.

[26] N. Dhanjani. New Age Application Attacks Against Apple's iOS and Countermeasures. In *Black Hat Europe 2011*, Barcelona, 2011.

[27] A. Dixit. *A Game-Theoretic Perspective on Diego Gambetta's Codes of the Underworld*. Paper presented at the book-launch conference for Codes of the Underworld, Oxford, 2010.

[28] X. Dong, J. A. Clark, and J. Jacob. Modelling User-Phishing Interaction. In *2008 Conference on Human System Interaction*, pages 627–632, Krakow, 2008.

[29] X. Dong, J. A. Clark, and J. Jacob. Threat Modelling in User Performed Authentication. In *10th International Conference on Information and Communications Security*, pages 49–64, Birmingham, 2008.

[30] S. Egelman, J. King, R. C. Miller, N. Ragouzis, and E. Shehan. Security User Studies: Methodologies and Best Practices. In *CHI 2007 Conference on Human Factors in Computing Systems*, pages 2833–2836, San Jose, 2007.

[31] A. Eriksson and P. Wretling. How Flexible is the Human Voice? A Case Study of Mimicry. In *5th European Conference on Speech Technology*, pages 1043–1046, Rhodes, 1997.

[32] P. Finn and M. Jakobsson. Designing Ethical Phishing Experiments. *IEEE Technology and Society Magazine*, 26(1):46–58, 2007.

[33] FIPS 190. *Guideline for the Use of Advanced Authentication Technology Alternatives*. National Institute of Standards and Technology, 1994.

[34] FIPS 200. *Minimum Security Requirements for Federal Information and Information Systems*. National Institute of Standards and Technology, 2006.

[35] FIPS 201. *Personal Identity Verification of Federal Employees and Contractors.* National Institute of Standards and Technology, 2005.

[36] I. Flechais, J. Riegelsberger, and M. A. Sasse. Divide and Conquer: The Role of Trust and Assurance in the Design of Secure Socio-Technical Systems. In *2005 Workshop on New Security Paradigms*, pages 33–41, Lake Arrowhead, 2005.

[37] D. Florencio and C. Herley. Analysis and Improvement of Anti-Phishing Schemes. In *IFIP TC-11 21st International Information Security Conference*, pages 148–157, Karlstad, 2006.

[38] B. Friedman, D. Hurley, D. C. Howe, E. Felten, and H. Nissenbaum. Users' Conceptions of Web Security: A Comparative Study. In *CHI 2001 Conference on Human Factors in Computing Systems*, pages 746–747, Minneapolis, 2002.

[39] D. Gafurov, E. Snekkenes, and P. Bours. Spoof Attacks on Gait Authentication System. *IEEE Transactions on Information Forensics and Security*, 2(3):491–502, 2007.

[40] S. Gajek, M. Manulis, and J. Schwenk. Enforcing User-Aware Browser-Based Mutual Authentication with Strong Locked Same Origin Policy. In *13th Australian Conference on Information Security and Privacy*, pages 6–20, Wollongong, 2008.

[41] S. Gajek and A.-R. Sadeghi. A Forensic Framework for Tracing Phishers. In *International Federation for Information Processing, volume 262*, pages 23–35, Karlstad, 2008.

[42] D. Gambetta. Deceptive Mimicry in Humans. In S. Hurley and N. Chater, editors, *Perspectives on Imitation: From Neuroscience to Social Science*, volume 2, pages 221–241. MIT Press, Cambridge, 2005.

[43] D. Gambetta. *Codes of the Underworld: How Criminals Communicate.* Princeton University Press, Woodstock, 2009.

[44] D. Gambetta and H. Hamill. *Streetwise: How Taxi Drivers Establish Their Customers' Trustworthiness.* Russel Sage Foundation, New York, 2005.

[45] S. M. Garcia, H. Song, and A. Tesser. Tainted Recommendations: The Social Comparison Bias. *Organisational Behaviour and Human Decision Processes*, 113(2):97–101, 2010.

[46] S. Garera, N. Provos, M. Chew, and A. D. Rubin. A Framework for Detection and Measurement of Phishing Attacks. In *2007 ACM Workshop on Recurring Malcode*, pages 1–8, Alexandria, 2007.

[47] D. Geer. Security Technologies Go Phishing. *Computer*, 38(6):18–21, 2005.

[48] D. Gollmann. *Computer Security*. Wiley Publishing, Indianapolis, second edition, 2006.

[49] J. Grossklags, N. Christin, and J. Chuang. Predicted and Observed User Behaviour in the Weakest-Link Security Game. In *1st Conference on Usability, Psychology, and Security*, pages 8:1–8:6, San Francisco, 2008.

[50] J. Grossklags, N. Christin, and J. Chuang. Secure or Insure? A Game-Theoretic Analysis of Information Security Games. In *17th International World Wide Web Conference*, pages 209–218, Beijing, 2008.

[51] J. Grossklags, B. Johnson, and N. Christin. When Information Improves Information Security. In *14th International Conference on Financial Cryptography and Data Security*, page ens Grossklags and Benjamin Johnson and Nicolas Christin, Tenerife, 2010.

[52] S. Gurses, C. Troncoso, and C. Diaz. Engineering Privacy by Design. In *4th International Conference on Computers, Privacy, and Data Protection*, Brussels, 2011.

[53] D. Harley and A. Lee. Phish Phodder: Is User Education Helping or Hindering? In *Virus Bulletin Conference*, pages 1–7, Vienna, 2007.

[54] C. Herley. So Long, and No Thanks for the Externalities: The Rational Rejection of Security Advice by Users. In *2009 Workshop on New Security Paradigms*, pages 133–144, Oxford, 2009.

[55] C. Herley and D. Florencio. A Profitless Endeavor: Phishing as Tragedy of the Commons. In *2008 Workshop on New Security Paradigms*, pages 59–70, Lake Tahoe, 2008.

[56] S. Hofmeyr, T. Moore, S. Forrest, B. Edwards, and G. Stelle. Modelling Internet-Scale Policies for Cleaning up Malware. In *10th Workshop on the Economics of Information Security*, Fairfax, 2011.

[57] ICO. *Privacy Impact Assessment Handbook, version 2*. Information Commissioner's Office, Wilmslow, 2009.

[58] ISO/IEC 27000:2009. *Information technology – Security techniques – Information security management systems – Overview and vocabulary*. International Organisation for Standardisation, 2009.

[59] T. N. Jagatic, N. A. Johnson, M. Jakobsson, and F. Menczer. Social Phishing. *Communications of the ACM*, 50(10):94–100, 2007.

[60] M. Jakobsson. Modelling and Preventing Phishing Attacks. In *9th International Conference on Financial Cryptography and Data Security*, Roseau, 2005.

[61] M. Jakobsson. The Human Factor in Phishing. In *6th National Forum on Privacy and Security of Consumer Information*, New York, 2007.

[62] M. Jakobsson and S. Myers, editors. *Phishing and Countermeasures: Understanding the Increasing Problem of Electronic Identity Theft*. Wiley-Interscience, Hoboken, 2007.

[63] M. Jakobsson, A. Tsow, A. Shah, E. Blevis, and Y.-K. Lim. What Instills Trust? A Qualitative Study of Phishing. In *11th International Conference on Financial cryptography and 1st International Conference on Usable Security*, pages 356–361, Scarborough, 2007.

[64] M. Jakobsson and A. Young. Distributed Phishing Attacks. In *US Department of Treasury Workshop on Resilient Financial Information Systems*, Washington DC, 2005.

[65] L. James. *Pishing Exposed*. Syngress Publishing, Rockland, 2005.

[66] L. John, A. Acquisti, and G. Loewenstein. The Best of Strangers: Context-dependent Willingness to Divulge Personal Information. In *5th Annual Whitebox Advisors Graduate Student Conference*, New Haven, 2009.

[67] Z. Jorgensen and T. Yu. On Mouse Dynamics as a Behavioral Biometric for Authentication. In *6th ACM Symposium on Information, Computer and Communications Security*, pages 476–482, Hong Kong, 2011.

[68] D. Kahneman. Maps of Bounded Rationality: Psychology for Behavioural Economics. *American Economic Review*, 93(5):1449–1475, 2003.

[69] Y.-G. Kim, S. Cho, J.-S. Lee, M.-S. Lee, I. H. Kim, and S. H. Kim. Method for Evaluating the Security Risk of a Website Against Phishing Attacks. In *IEEE ISI 2008 International Workshops on Intelligence and Security Informatics*, pages 21–31, Taipei, 2008.

[70] P. Kumaraguru, A. Acquisti, and L. F. Cranor. Trust Modelling for Online Transactions: A Phishing Scenario. In *2006 International Conference on Privacy, Security and Trust*, pages 11:1–11:9, Markham, 2006.

[71] P. Kumaraguru, S. Sheng, A. Acquisti, L. F. Cranor, and J. Hong. Teaching Johnny Not to Fall for Phish. *ACM Transactions on Internet Technology*, 10(2):7:1–7:31, 2010.

[72] G. Larcom and A. Elbirt. Gone Phishing. *IEEE Technology and Society Magazine*, 25(3):52–55, 2006.

[73] C. Lewis and J. Rieman. *Task-Centred User Interface Design: A Practical Introduction*. University of Colorado, 1993.

[74] E. Lin, S. Greenberg, E. Trotter, D. Ma, and J. Aycock. Does Domain Highlighting Help People Identify Phishing Sites? In *CHI 2011 Conference on Human Factors in Computing Systems*, pages 2075–2084, Vancouver, 2011.

[75] M. Lomas. Why Are We Authenticating (Transcript of Discussion). In *12th International Workshop on Security Protocols*, pages 291–298, Cambridge, 2004.

[76] D. K. McGrath and M. Gupta. Behind Phishing: An Examination of Phisher Modi Operandi. In *1st Usenix Workshop on Large-Scale Exploits and Emergent Threats*, pages 4:1–4:8, San Francisco, 2008.

[77] C. J. Mitchell. Using Human Interactive Proofs to Secure Human-Machine Interactions via Untrusted Intermediaries. In *14th International Workshop on Security Protocols*, pages 164–170, Cambridge, 2006.

[78] C. J. Mitchell. Using Human Interactive Proofs to Secure Human-Machine Interactions via Untrusted Intermediaries (Transcript of Discussion). In *14th International Workshop on Security Protocols*, pages 171–176, 2006.

[79] D. Miyamoto, H. Hazeyama, and Y. Kadobayashi. SPS: A Simple Filtering Algorithm to Thwart Phishing Attacks. In *1st Asian Internet Engineering Conference*, pages 195–209, Bangkok, 2005.

[80] T. Moore and R. Clayton. How Hard Can It Be to Measure Phishing? Position paper submitted to *Mapping and Measuring Cybercrime*, Oxford, 2010.

[81] T. Moore and R. Clayton. An Empirical Analysis of the Current State of Phishing Attack and Defence. In *6th Workshop on the Economics of Information Security*, Pittsburgh, 2007.

[82] T. Moore and R. Clayton. Examining the Impact of Website Take-down on Phishing. In *2nd APWG eCRIME Researcher's Summit*, pages 1–13, Pittsburgh, 2007.

[83] T. Moore and R. Clayton. Evil Searching: Compromise and Recompromise of Internet Hosts for Phishing. In *13th International Conference on Financial Cryptography and Data Security*, pages 256–272, Barbados, 2009.

[84] T. Moore and R. Clayton. The Impact of Incentives on Notice and Take-down. In *Managing Information Risk and the Economics of Security*, pages 199–223. Springer Science, New York, 2009.

[85] T. Moore, R. Clayton, and R. Anderson. The Economics of Online Crime. *Journal of Economic Perspectives*, 23(3):3–20, 2009.

[86] E. Mordini and C. Petrini. Ethical and Social Implications of Biometric Identification Technology. *Ann Ist Super Sanita*, 43(5):5–11, 2007.

[87] M. E. J. Newman. The Structure and Function of Complex Networks. *SIAM Review*, 45(2):167–259, 2003.

[88] A. Nicholas, S. Brody, P. de Sutter, and F. de Carufel. A Woman's History of Vaginal Orgasm is Discernible from Her Walk. *The Journal of Sexual Medicine*, 5(9):2119–2124, 2008.

[89] Oxford Dictionaries. *Credential*. Oxford University Press, April 2010. http://oxforddictionaries.com, accessed 29 July 2011.

[90] H. Pamunuwa, D. Wijesekera, and C. Farkas. An Intrusion Detection System for Detecting Phishing Attacks. In *2007 SIAM International Conference on Data Mining*, pages 181–192, Minneapolis, 2007.

[91] B. Parno, C. Kuo, and A. Perrig. Phoolproof Phishing Prevention. In *10th International Conference on Financial Cryptography and Data Security*, pages 1–19, Anguilla, 2006.

[92] A. Perrig and D. Song. Hash Visualisation: A New Technique to Improve Real-World Security. In *International Workshop on Cryptographic Techniques and E-Commerce*, Hong Kong, 1999.

[93] W. Pieters and A. van Cleef. The Precautionary Principle in a World of Digital Dependencies. *Computer*, 42(6):50–56, 2009.

[94] M. Power. *The Audit Explosion*. Demos, London, 1994.

[95] M. Power. *The Risk Management of Everything: Rethinking the Politics of Uncertainty*. Demos, London, 2004.

[96] P. Prabhu, A. Akel, L. M. Grupp, W.-K. S. Yu, G. E. Suh, E. Kan, and S. Swanson. Extracting Device Fingerprints from Flash Memory by Exploiting Physical Variations. In *4th International Conference on Trust and Trustworthy Computing*, pages 188–201, Pittsburgh, 2011.

[97] M. Pusara and C. E. Brodley. User Re-Authentication via Mouse Movements. In *2004 ACM Workshop on Visualisation and Data Mining for Computer Security*, pages 1–8, Washington DC, 2004.

[98] M. A. Rajab, L. Ballard, N. Jagpal, P. Mavrommatis, D. Nojiri, N. Provos, and L. Schmidt. Trends in Circumventing Web-Malware Detection. Technical report, Google, 2011.

[99] J. Riegelsberger, M. A. Sasse, and J. D. McCarthy. Do People Trust Their Eyes More Than Ears? Media Bias in Detecting Cues of Expertise. In *CHI 2005 Conference on Human Factors in Computing Systems*, pages 1745–1748, Portland, 2005.

[100] C. Roberts. Biometric Attack Vectors and Defences. *Computers and Security*, 26(1):14–25, 2007.

[101] M. A. Sasse. Usability and Trust in Information Systems. In *Trust and Crime in Information Societies*, pages 319–349. Edward Elgar Publishing, Northampton, 2005.

[102] S. E. Schechter, R. Dhamija, A. Ozment, and I. Fischer. The Emperor's New Security Indicators. In *2007 IEEE Symposium on Security and Privacy*, pages 51–65, Oakland, 2007.

[103] J. Shirley and D. Evans. The User is Not the Enemy: Fighting Malware by Tracking User Intentions. In *2008 Workshop on New Security Paradigms*, pages 33–45, Lake Tahoe, 2008.

[104] H. A. Simon. Rational Choice and the Structure of the Environment. *Psychological Review*, 63(2):129–138, 1956.

[105] D. K. Smetters and P. Stewart. Breaking out of the Browser to Defend Against Phishing Attacks. In *5th Conference on Email and Anti-Spam*, Mountain View, 2008.

[106] F. Stajano and P. Wilson. Understanding Scam Victims: Seven Principles for Systems Security. *Communications of the ACM*, 54(3):70–75, 2011.

[107] J. M. Stanton, K. R. Stam, P. Mastrangelo, and J. Jolton. Analysis of End User Security Behaviour. *Computers and Security*, 24(2):124–133, 2005.

[108] D. Stebila. Reinforcing Bad Behaviour: The Misuse of Security Indicators on Popular Websites. In *22nd Conference of the Computer-Human Interaction Special Interest Group of Australia on Computer-Human Interaction*, pages 248–251, Brisbane, 2010.

[109] S. Strid and C. Andreasson. *The Viking Manifesto: The Scandinavian Approach to Business and Blasphemy*. Marshall Cavendish, London, 2008.

[110] J. E. Tapiador and J. A. Clark. Masquerade Mimicry Attack Detection: A Randomised Approach. *Computers & Security*, 30(5):297–310, 2011.

[111] A. Tsow. Phishing with Consumer Electronics: Malicious Home Routers. In *15th International World Wide Web Conference*, pages 22–26, Edinburgh, 2006.

[112] A. Tsow, M. Jakobsson, and F. Menczer. Stopping Distributed Phishing Attacks. In *5th Workshop on the Economics of Information Security*, Cambridge, 2006.

[113] UKCA. Fraud Losses Drop on UK Cards, Cheques and Online Banking. UK Cards Association, Press release dated 9th Match 2011.

[114] I. van der Ploeg. The Politics of Biometric Identification: Normative Aspects of Automated Social Categorisation. Technical report, Institute for Healthcare Management and Policy, University Medical Centre Rotterdam, 2005.

[115] I. van der Ploeg. Genetics, Biometrics and the Informatisation of the Body. *Ann Ist Super Sanita*, 43(1):44–50, 2007.

[116] D. Wagner and R. Dean. Intrusion Detection via Static Analysis. In *2001 IEEE Symposium on Security and Privacy*, pages 156–168, Oakland, 2001.

[117] D. Walsh. Hide the Address Bar within Mobile Web Applications. *http://davidwalsh.name/hide-address-bar*, March 2011.

[118] R. Walton. Cryptography and Trust. *Information Security Technical Report*, 11(2):68–71, 2006.

[119] P. A. Watters. Why Do Users Trust the Wrong Messages? A Behavioural Model of Phishing. In *4th APWG eCRIME Researcher's Summit*, pages 1–7, Tacoma, 2009.

[120] J. Wells, D. Hutchinson, and J. Pierce. Enhanced Security for Preventing Man-in-the- Middle Attacks in Authentication, Data Entry and Transaction Verification. In *6th Australian Information Security Management Conference*, Perth, 2008.

[121] M. Wu, R. C. Miller, and G. Little. Web Wallet: Preventing Phishing Attacks by Revealing User Intentions. In *Symposium On Usable Privacy and Security 2006*, pages 102–113, Pittsburgh, 2006.

[122] G. Xiang, B. A. Pendleton, J. Hong, and C. P. Rose. A Hierarchical Adaptive Probabilistic Approach for Zero Hour Phish Detection. In *15th European Conference on Research in Computer Security*, pages 268–285, Athens, 2010.

[123] R. Yampolskiy. Mimicry Attack on Strategy-Based Behavioural Biometric. In *5th International Conference on Information Technology: New Generations*, pages 916–921, Las Vegas, 2008.

[124] Z. Ye, S. Smith, and D. Anthony. Trusted Paths for Browsers. *ACM Transactions on Information and System Security*, 8(2):153–186, 2005.

[125] K.-P. Yee. User Interaction Design for Secure Systems. Technical Report UCB/CSD-02-1184, University of California, Berkeley, 2002.

[126] K.-P. Yee and K. Sitaker. Passpet: Convenient Password Management and Phishing Protection. In *Symposium On Usable Privacy and Security 2006*, pages 32–43, Pittsburgh, 2006.

[127] Y. Zhang, S. Egelman, L. F. Cranor, and J. Hong. Phinding Phish: Evaluating Anti-Phishing Tools. In *14th Annual Network and Distributed System Security Symposium*, San Diego, 2007.

[128] Y. Zhang, J. Hong, and L. F. Cranor. CANTINA: A Content-Based Approach to Detecting Phishing Web Sites. In *16th International World Wide Web Conference*, pages 639–648, Banff, 2007.