Social Engineering Defence Mechanisms and counteracting Training Strategies

Peter Schaab[*], Kristian Beckers[*], Sebastian Pape[+]

[*]Technische Universität München (TUM)
peter.schaab@in.tum.de, beckersk@in.tum.de
[+]Goethe Universität Frankfurt
Sebastian.Pape@m-chair.de

Abstract

Purpose – The paper aims to outline strategies for defence against social engineering that are missing in current best practices of IT security. Reason for the incomplete training techniques in IT security is the interdisciplinary of the field. Social engineering is focusing on exploiting human behaviour and this is not sufficiently addressed in IT security. Instead most defence strategies are devised by IT security experts with a background in information systems rather than human behaviour. We aim to outline this gap and point out strategies to fill the gaps.
Design/methodology/approach – We conducted a literature review from viewpoint IT security and viewpoint social psychology. In addition, we mapped the results to outline gaps and analysed how these gaps could be filled using established methods from social psychology and discussed our findings.
Findings – We analysed gaps in social engineering defences and mapped them to underlying psychological principles of social engineering attacks e.g. social proof. Furthermore, we discuss which type of countermeasure proposed in social psychology should be applied to counteract which principle. We derived two training strategies from these results that go beyond the state of the art trainings in IT security and allow security professional to raise companies' bars against social engineering attacks.
Originality/value – Our training strategies outline how interdisciplinary research between computer science and social psychology can lead to a more complete defence against social engineering by providing reference points for researchers and IT security professional with advice on how to improve training.

Keywords

social engineering, security management, persuasion, human-centred defence mechanisms

Paper type

Literature review

1. Introduction

Although security technology improves, the human user remains the weakest link in system security. Therefore, it is widely accepted that the people of an organization are the main vulnerability of any organization's security, as well as the most challenging aspect of system security (Barrett, 2003; Mitnick and Simon, 2011). This is emphasized by many security consultants, as well as from genuine attackers, which accessed critical information via social engineering (Gragg, 2003; Warkentin and Willison, 2009). Early on Gulati (2003) reported that cyber attacks cost U.S. companies $266 million every year and that 80% of the attacks are a form of social engineering. A study in 2011 showed that nearly half of the considered large companies and a third of small companies fell victim of 25 or more social engineering attacks in the two years before (Dimensional Research, 2011). The study further shows that costs per incident usually vary between $25 000 and over $100 000. Furthermore, surveys, like Verizon's 'Data Breach Investigation Report' (2012; 2013), show the impact of social engineering. Even though the awareness about the phenomenon of social engineering has increased, at least in literature, the impact has grown from 7% of breaches in 2012 to 29% of breaches in 2013 according to these studies. In addition, current security awareness programs are apparently ineffective (Pfleeger et al., 2014). These alarming numbers question whether the existing approaches towards awareness and defence of social engineering are fundamentally incomplete.

Frangopoulos et al. (2010) consider the psychological aspects of social engineering and relate them to persuasion techniques in their 2010 publication. In contrast to our work their work is not based on a literature review of behaviour psychology, but based on the expertise of the authors. Moreover, the scope of the authors is broader and consider physical measures, as well as security standards in their work. Our results classify existing research in IT security and persuasion in literature and contribute a structured gap analysis. In addition, Frangopoulos et al. (2012) transfer the knowledge of psychosocial risks, e.g. influence of headaches and colds on decisions, from a managerial and organisational point of view to the information security view.

Our hypothesis is that the psychological aspects behind social engineering and user psychology are not considered to their full extend. For instance, Ferreira et al. (2015) constitute psychological principles in social engineering and relate these principles to previous research of Cialdini (2009), Gragg (2003) and Stajano and Wilson (2011). Thus, as starting point we analysed the psychological explanations of these social engineering principles by relating the insights of Cialdini, Gragg and Stajano and Wilson. As these principles have to be the fundamental concern of any security defence mechanism against social engineering, we contribute a list of concepts that address social engineering defence mechanisms. In particular, we analyse recommendations from IT in comparison to recommendations given by social psychology. The results of our analysis are twofold. On one side we provide a mapping between the influence of the identified defence mechanism to mitigate social engineering attacks based on the individual psychological principles. On the other side the analysis reveals fundamental gaps in today's security awareness approach. We provide a road map that shows how to address these gaps in the future. Our road map

is an instrumental vision towards reducing the social engineering threat by addressing all relevant psychological aspects in its defence.

2. Methodology

Our research was guided by the methodology outlined in Fig. 1. We initialized the work with a working definition of social engineering (Sect. 3) and surveyed the state of the art from the viewpoint of computer science in particular with regard to IT security (Sect. 4) and separately from the viewpoint of social psychology (Sect. 6). We used the meta search engines Google Scholar and Scopus, which include the main libraries of IEEE, ACM, Springer, Elsevier and numerous further publishers. Based on the findings of our literature survey and a review of psychological principles behind social engineering (Sect. 5), we identified requirements and techniques from social sciences for defending against social engineering (Sect. 6) and map these to underlying psychological principles of the attacks (Sect. 7). Next, we map these to the defence mechanisms used in IT security today (Sect. 8). We outline the resulting gap and present a vision for overcoming these shortcomings of current IT security defences and derive missing training strategies (Sect. 9). Finally, we discuss our results and provide directions for future research.

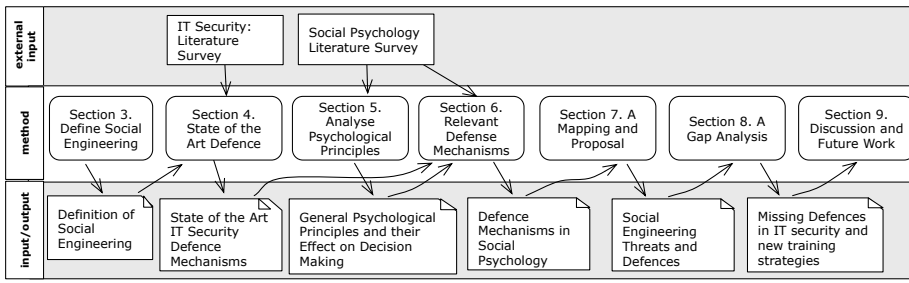

Figure 1: Methodology

3. Definition of Social Engineering

Although there is no agreed upon definition of social engineering, the common idea arising from the available definitions is that social engineering is the acquisition of confidential, private or privileged information by methods including both technical and non-technical means (Manske, 2009). This common idea is quite general, as it includes means of gaining information access such as shoulder surfing, dumpster diving, etc. However, it especially refers to social interaction as psychological process of manipulating or persuading people into disclosing such information (Thornburgh, 2004). Other than the former methods of accessing information, the latter are more complex and more difficult to resist, as persuasion is based on psychology. In this context, persuasion can be viewed as "any instance in which an active attempt is made to change a person's mind" (Petty and Cacioppo, 1996, p.4). The concept of 'optimism

bias' states that people believe that others fall victim to misfortune, not themselves (Weinstein, 1980). Additionally, they tend to overestimate their possibilities to influence an event's outcome. Hence people think that they (i) will not be targeted by social engineering and (ii) are more likely to resist than their peers.

To actually raise resistance, we analyse how information security awareness can be increased. In alignment with Kruger and Kearny (2006) we define information security awareness as the degree to which employees understand the need for security measures and adjust their behaviour to prevent security incidents. Furthermore, in accordance with Veseli (2011) we focus on the information security dimensions attitude (how does a person feel about the topic) and behaviour (what does a person do) as they are an expression of conscious and unconscious knowledge (what does a person know).

4. Analysis of Social Engineering Defence Mechanisms in IT Security

After having established the concept of social engineering, we analyse how the threat of social engineering is met in IT security. As the main vulnerability exploited by social engineering is inherent in human nature, it is the human element in systems that needs to be addressed. Thus, we concentrate on human based defence mechanisms. Predominantly three human based mitigation methods are proposed: Policies, audits and security awareness programs, as indicated in Table 1. User awareness and security policies dominate the recommendations to defend social engineering (Scheeres, 2008).

Dimension		**Defence Mechanism**	**Description**
Knowledge	**Attitude**	Policy Compliance	- Foundation of information security - System standards and security levels - Guidelines for user behaviour
		Security Awareness Program	- Familiarity with security policy - Knowledge about sensitive, valuable information - Basic indicators, suspicious behaviour connected to social engineering attacks - (Recognition of being manipulated)
	Behaviour	Audit	- Test employee susceptibility to social engineering - Identify weaknesses of policy and security

			awareness program

Table 1: Defence mechanisms used in IT security

Security Policies. Any information security is founded on its policy (Mitnick and Simon, 2011). Furthermore, policies provide instructions and guidelines how users should behave. It is especially hard to address social engineering in security policies, since people need to know how to respond to ambiguous requests (Gragg, 2003). By safe-guarding information, users should not come into uncertainty to decide whether certain information is sensitive or not. Necessarily these policies need to be enforced consistently throughout the system.

Security Awareness Programs. Upon establishment of a security policy all users need to be trained in security awareness programs to follow the policy, practices and procedures (Mitnick and Simon, 2011; Thornburgh, 2004). In general, the literature agrees upon the cornerstones of an awareness program. First of all, familiarity with the security policy needs to be established. It is important that everyone in the organization knows what kind of information is sensitive, hence particularly valuable for an attacker. Secondly, knowledge about social engineering is to be conveyed. This includes basics of social engineering, and how attacks work in detail. This should help employees to understand the reasons for related security policies that simply contains rules and usually not the reasoning behind it. The idea is that the understanding of why these polices were defined, will increase compliant behaviour among employees. In addition, the thought knowledge should reach beyond the rules in the policies and contain in particular indicators of social engineering attacks and what behaviour could be suspicious, such as requesting confidential information or to refuse provision of personal or contact information. Gragg (2003) demands the inclusion of additional training for key personnel to include inoculation, forewarning and reality check, see Section 6.

Audit. The conduction of audits is complementary to the above approaches (Thornburgh, 2004). It serves the purpose to test the susceptibility to social engineering attacks (Mitnick and Simon, 2011). Hence, it tests the effectiveness and identifies weaknesses of the other conducted methods (Winkler and Dealy, 1995). In this particular case, classic audits or penetration tests need an extension to social engineering penetration testing as done by Bakhshi et al. (2008). This extension is not trivial since it tests humans who can get upset and the work council needs to be involved.

5. Analysis of Psychological Principles underlying Social Engineering

According to Rusch (1999) two ways of persuading an individual exist:

1. A central route to persuasion based on sound analytical reasoning of facts;

2. or a peripheral route to persuasion relying on acceptance without deeply reasoning about the facts by triggering mental shortcuts or eliciting emotions.[1]

Thornburgh (2004) and Ivaturi and Janczewski (2011) state that the central route (1) is not a real option for a social engineer as his entire approach is based on "misrepresentation and dissembling" (Thornburgh, 2004) meaning the employment of deception and manipulation. Gragg (2003), analysed literature on persuasion, influence and social engineering and suggests seven psychological triggers which are explicitly referred to as being applicable to social engineering. Scheeres (2008) has deduced that Gragg's triggers are in line with the principles of Cialdini (2009). This means Cialdini's result is also applicable to social engineering. Additionally, it is generally accepted that the same psychological techniques are applied in social engineering as in traditional fraud (Rusch, 1999). Therefore, Stajano and Wilson (2011) identified seven principles of scam applicable to social engineering. Based on these findings Ferreira et al. (2015) enhance the principles of Cialdini, Gragg, and Stajano and Wilson by constituting a complete set of psychological principles of persuasion in social engineering. Ferreira et al. related the already existing principles to each other and identified five principles of persuasion in social engineering that account for all available principles. We investigate why people are prone for them to get insight into the prevalent threats of social engineering. We do not analyse studies that suggest or validate the behaviours described.[2] Instead we focus on the triggered behaviours and try to find insights into their functioning. This is done to gain further valuable understanding of the triggers to find valid countermeasures in a next step. A summary of these principles is provided in Table 2.

Psychological Principle	**Description**
Authority	- Conditioning to respond to authority - Beneficial to unconditionally conform to authority - Authority indicated by abstract rank
Social Proof	- Reliance on majority's apparent behaviour in determining appropriate behaviour in uncertainty - Confidence when seemingly not solely responsible

[1] These two routes are referred to as System 1 and System 2 in cognitive psychology.

[2] Valid examples, studies and a variety of scenarios, which principle is applicable when, can be found aplenty in Cialdini (2009), Gragg (2003) and Stajano and Wilson (2011).

Liking, Similarity and Deception	- Tendency to react positively to whom some kind of 'relationship' has been established - Relationship sources: attractiveness, compliments, familiarity, liking - Satisfaction of expectations through manipulation
Commitment, Reciprocation and Consistency	- Urge to consistency with commitment - Societal obligation to future repayments of received concessions
Distraction	- Limited attention is focused on seemingly important facts or actions - Directing attention in desired direction by manipulation of focus

Table 2: Psychological principles of social engineering

Authority. "Society trains people not to question authority so they are conditioned to respond to it" (Cialdini, 2009).As Milgram (1974) puts it, conforming to authority figures' wishes and commands has always proved to be beneficial for us. As long as we can think these people (e.g. parents, teachers) knew more than us, and for us taking advice had advantages — partly due to greater wisdom, partly due to the control of rewards and punishments (Milgram, 1974). This pattern persists up to adulthood, only authority figures change, now appearing as e.g. employers or judges. But it continuously might be wise to comply with the dictate of constituted authorities, independently of how this authority constitutes itself. In modern society responses to authority are made to abstract rank, even in the absence of any substance of authority, as long as it is indicated by an insignia, uniform or title (Cialdini, 2009; Milgram, 1974). Due to this societal trained behaviour of unconditioned response to authority, people without questioning adhere to the dictate of authoritative figures as demonstrated by the famous Milgram (Milgram, 1963) experiment.[3] "People usually follow an expert or pretence of authority and do a great deal for someone they think represents authority" (Cialdini, 2009).

Social Proof. People rely on others in determining what is appropriate in any given

[3] In an experiment individuals were instructed to supervise electric shocks of increasing strength to other individuals when those made mistakes. The victims were accomplices who did not in fact receive the shocks. The individuals complied with shocking extent. They continued to apply electrical shocks of up to 450 V. Even when victims pretended screaming and fainting they did not spare the experimental subjects.

situation. According to Cialdini (2009) experience tells us to act according to social evidence rather than to its contrary. Especially when in uncertainty of correct behaviour, the behaviour of the majority of people tends to be correct and therefore constitutes correct behaviour for ourselves or at least provides a feeling of confidence and safety to conduct an otherwise doubtable action or an action against our self-interest (Cialdini, 2009; Rusch, 1999; Stajano and Wilson, 2011). Furthermore, the behaviour of people similar to us, more powerfully establishes what is considered correct. As pointed out by Stajano and Wilson (2011) and Gragg (2003) this principle also accounts for people's will to take risks in an action, especially if not being held solely responsible. "People let their guard and suspicion down when everyone else appears to share the same behaviours and risks. In this way, they will not be held solely responsible for their actions." (Cialdini, 2009)

Liking, Similarity and Deception. Humans have a tendency to abide and comply or at least react positively to whom some kind of 'relationship' exists or is established. This relationship can take a variety of manifestations. Cialdini (2009) describes the major mechanisms of deceiving an individual into one of these relationships:

Attractiveness. Physical attractiveness is a characteristic that is associated with a 'halo effect'.[4] And therefore people assign favourable traits such as kindness, honesty and trustworthiness to attractive persons and therefore treat these persons favourable.

Similarity. To have identical or similar characteristics with an individual incentivizes people to favour this individual. This similarity can be accomplished in a wide range of attributes, such as opinions, personality traits, personal interests, background, appearance, etc.

Compliments. People tend to react positively to praise, affinity or general compliments to such an extent as for liking and compliance.

Contact and Cooperation. Attitude, especially the favour, towards an individual is influenced by the exposition to it. Therefore, familiarity evoked by contact usually leads to a more favourable mindset. This can even be increased through mutual cooperation or the attempt to establish a 'we' or 'us' as Gragg (2003) points out as well.

Conditioning and Association. Simple association with bad or good things influences how people feel about someone, it is enough to stimulate either like or dislike (Lott

---

[4] A halo effect occurs when one characteristic of an individual dominates how this individual is perceived by others.

and Lott, 1965).

Besides deceiving an individual into one of the above relations, Stajano and Wilson (2011) indicate that by knowing people's expectations, an individual can be deceived into authenticating a person and therefore it can be manipulated into moving along within any situation as long as the individual's expectations are satisfied.

Commitment, Reciprocation and Consistency. People feel induced to be consistent once having committed (publically) to a specific action. This tendency is neither influenced by the commitment not being very wise, nor by recognizing it to be foolish or in contrast to our own interests (Cacioppo et al., 1986; Cialdini, 2009). As Stajano and Wilson (2011) and Gragg (2003) emphasize this also accounts for requests that may not have been legitimated or are even illegal. According to Cialdini (2009) people encounter personal and interpersonal pressures to stay consistent with an earlier commitment causing them to act accordingly to their previous commitment. People tend to take considerable pains to stay consistent (Rusch, 1999). Staying consistent is in fact considered as central motivator for human's behaviour as it is highly rewarded in our culture. It is associated with integrity, personal and intellectual strength, whereas inconsistency is viewed as untrustworthy and therefore an undesired personal characteristic. Consistency provides reasonable orientation to our lives. This is accompanied by the tendency to believe that others express their true feelings and attitudes when making a statement (Gragg, 2003).

The desire to appear consistent in our actions has formed another strongly connected behaviour or well-established rule in social interactions — reciprocation. This rule obligates an individual to future repayment for favours or generally any- thing given or promised to us (Cialdini, 2009; Rusch, 1999). According to Gouldner (1960) this rule is ingrained into any human society. As Cialdini (2009) puts it, a society wide shared feeling of future obligation is necessary to make social interaction in today's form possible, as it lowers natural inhibitions against transactions and instead allows an individual to provide resources with confidence that the given is not being lost but returned in the future. As this brings immense advantages, people are trained to comply and not question the rule of reciprocation. Again, society considers individuals that take and do not return anything with negativity and therefore it is inherent in human's desire to try and avoid this.

The above comprises certain implications, that distinguish the rule of reciprocation from the other principles (Cialdini, 2009):

– By imposing a favour on us a disliked or unwelcomed person enhances his chance of our compliance significantly.

– An uninvited favour causes a feeling of indebtedness, as receiving the favour obligates to repay. This enables others to choose who is indebted to them, not oneself.

– Although generally the rule encourages equal exchanges, it enables an individual

to choose both the kind of initial indebting favour, e.g. a small one, as well as the kind of compensating return favour, e.g. a significantly larger one.

– Furthermore, the rule implies the obligation of a concession, if someone has made an initial concession. Mutual concession promotes compromise in social interactions, as requirements of interacting persons often are unacceptable to one another.

Distraction People focus their limited attention on what is perceived to be most interesting or most important for a variety of reasons, and ignore seemingly uninteresting and unimportant facts or actions that may happen simultaneously (Stajano and Wilson, 2011). Due to this limited attention, it is possible to direct an individual in any desired direction, the individual is distracted. Basically these distractions heighten people's emotional state, which interferes with their ability to evaluate facts or actions by logical reasoning (Ferreira et al., 2015; Gragg, 2003). This can be achieved in a number of ways:

Human's Needs. Knowing a person's needs, desires and fears provides an understanding what drives him and how he behaves. This makes him vulnerable to emotional manipulation (Gragg, 2003; Stajano and Wilson, 2011). The phenomena is called counterfactual thinking and describes how the anticipation of future possibilities, caused by aiming at a person's needs, impedes reasoning (Landman and Petty, 2000).

Time. Depending on the urgency of a request the caused response may be different as it hinders evaluation (Stajano and Wilson, 2011). The same accounts for an information or sensory overload (Gragg, 2003). This is due to time not being available to process all information or implications of a request.

Scarcity. Potential loss highly influences decision making. By considering the availability of something people may often come to a decision about quality or worthiness without actually reasoning about e.g. their need (Lynn, 1989). Additionally, humans have a need to retain their freedom, thus in case a choice is limited or threatened the desire to preserve their freedom decidedly raises, as personal control is reduced (Brehm, 1966).

When people's attention is focused, directed or influenced by any of the above factors, they are distracted from proper evaluation and protection of their true intentions (Stajano and Wilson, 2011).

The analysed psychological principles share one special characteristic. They all describe how an individual or humans in general are induced to use a specific, automated decision mechanism, often called heuristic or mental shortcut, rather than rational reasoning. This is achieved by making use of the described concepts and human tendencies. After having analysed these tendencies and triggers, it is necessary to understand the different mechanisms in decision making. As Kahneman (2003) explains, humans cognitive functioning is distinguished into two separate cognitive systems. One system intuits (System 1) and the other reasons (System 2):

> "The operations of System 1 are typically fast, automatic, effortless, associative, implicit (not available to introspection), and often emotionally charged; they are also governed by habit and therefore difficult to control or modify. The operations of System 2 are slower, serial, effortful, more likely to be consciously monitored and deliberately controlled; they are also relatively flexible and potentially rule governed." (Kahneman, 2003)

Kahneman (2003) furthermore describes the differentiating aspects of the two systems. System 1 generates impressions of perceptions and thought, which are involuntarily and not necessarily verbally explicit. In comparison, judgments are intentional and explicit even when not verbally expressed. This means, when judging System 2 is usually involved, whether the judgment originates from impressions or reasoning. If a judgment directly reflects impressions and was not modified by System 2 then it is an intuitive judgment. Normally many intuitive judgements are expressed, even though System 2 is set to monitor mental operations (Gilbert, 2002; Stanovich and West, 2002). The competing behaviour of the two systems is summarized in Figure 2. As self-monitoring as well as reasoning are effortful operations, System 2 is affected by dual-task interference (Kahneman, 2003). Due to operations of System 2 being effortful, the monitoring of intuitive judgments usually is not very strict and therefore erroneous ones are not hindered because plausible judgments that are readily made are trusted (Cialdini, 2009; Kahneman, 2003). Being lax in monitoring is not only laziness or the attempt to avoid hard thinking, it is also a mechanism to reduce cognitive load. Besides by leaving System 1 in autopilot and not thinking straight troubling realizations can and will be avoided (Kahneman, 2003).

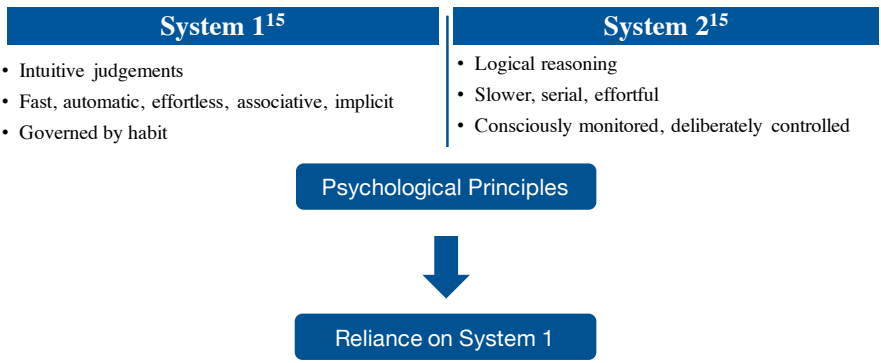

Figure 2: Psychological Triggers

Figure 2 illustrates that by using any of the psychological triggers, a social engineer tries to push the person opposite to rely on System 1, as there exists an evolutionary built heuristic that delivers an intuitive judgment, which is usually not monitored by

System 2.[5]

6. Relevant Defence Mechanisms in Social Psychology

The intentions of security awareness programs are to inform about social engineering and sensitive information. It is assumed that by knowing about the threat of social engineering, users are less likely to be susceptible for such attacks. There is only a few researchers that have found this not to be sufficient, which appears to be ignored by most others. Gragg (2003) considers psychological principles of persuasion behind social engineering. Ferreira et al. (2015) have established a framework of psychological principles. These exhibit the ability to influence and potentially manipulate a person's attitude, believes and behaviour. Gragg therefore recommends techniques to build resistance against persuasion, borrowed from social psychology, to be included into awareness programs. An overview over these methods is given in Table 3. They build on Sagarin et al. (2002):

Dimension		**Defence Mechanism**	**Description**
Knowledge	**Attitude**	Persuasion Knowledge	- Information about tactics used in persuasion attempts and their potential influence on attitude and behaviour - Information about appropriate coping
		Forewarning	- Warning of message content and persuasion attempt
		Attitude Bolstering	- Thought process strengthening security attitude
		Reality Check	- Demonstration of vulnerability to perceive risk of persuasion
	Behaviour	Inoculation	- Exposition to persuasive attempts and arguments of a social engineer - Provision of counter arguments to resist persuasion

[5] Of course there have evolved many more than the above introduced heuristics, allowing people to function effectively but therefore allowing people to bypass System 2 (Gigerenzer and Todd, 1999). We kept to the ones which are directly linked to persuading an individual. Although some of the available heuristics may have further impact on the behavior when attacked by a social engineer. For a collection of these heuristics view Schneier (2008) and Kahneman (2003).

		Decision Making	- Repeated exposition to “similar” decision making situations

Table 3: Defence mechanisms against persuasion borrowed from social psychology

Inoculation. A user gets exposed to persuasive attempts of a social engineer, he is put into a situation a social engineer would put him in. Thereby he is exposed to arguments that a social engineer may use. Also he is given counter arguments that he can use to resist the persuasion. This works the same way as preventing a disease being spread by using inoculation and induces resistance to persuasion.

Forewarning. Forewarnings of message content and the persuasion attempt of the message triggers resistance to a social engineering attack. The intention is to not only warn about the persuasive attempt of a social engineer, but in particular to warn about the arguments being manipulative and deceptive. An example of this technique would be the warning about fraudulent IT support calls asking for user login and password.

Reality Check. As people tend to believe that they are invulnerable due to optimism bias, users need to realize that in fact they are vulnerable. Therefore, it has to be demonstrated to them, that they are vulnerable, to make them perceive the risks and training to be effective. However, any such effort has to be careful not to cause an amount of frustration that leads people to conclude their security efforts are useless. The balance between the demonstration of the vulnerability and the ensurance that people can make a difference in social engineering defence is vital for the success of defences.

Even though it appears that most programs are not extensive or limited in impact, it is unclear how much attention is given to these proposals in security practice. Nevertheless, research in the field of psychology over the past five decades has proven that inoculation is the most consistent and reliable method to induce resistance to persuasion (Miller et al., 2013). We are not aware of any study directly analysing the effects of inoculation to the resistance to social engineering. We are convinced that the principles behind inoculation are sound and we will analyse their effect on people in a future empirical study. In addition, Gragg (2003) has already adopted inoculation as a valuable mechanism for resistance to social engineering. Nevertheless, there exist further techniques in social psychology to train resistance to persuasion:

Persuasion Knowledge. Aim of security awareness programs is for users to experience resistance toward persuasion in case of a social engineering at- tack. This experience is increased if a user is concerned about being deceived (Friestad and Wright, 1994). Persuasion knowledge consists of information about tactics used in persuasive situations, their possible influence on attitudes and behaviour, their effectiveness and appropriateness, the persuasive agent’s motives, and coping strategies (Fransen et al., 2015; Friestad & Wright, 1994). Activated persuasion knowledge usually either elicits suspicion about the persuasive agent’s motives, or scepticism about arguments, and perceptions of manipulation or deception. Furthermore, it directs to options how to respond and selects coping tactics believed to be appropriate (Friestad and Wright, 1994). This positive relationship between persuasion knowledge and resistance to persuasive attempts is demonstrated by

(Briñol et al., 2015): People are aware of persuasive attempts when having knowledge about persuasion and respond appropriately. This means educating users not only about common social engineering attack methods (e.g. phishing) but particularly about psychological principles used in social engineering is an absolute necessity. As people also enhance their persuasion knowledge from experiences in social interactions, inoculation plays a vital role. Knowledge about coping tactics is, as indicated, essential to evaluate response options and to cope with persuasive attempts.

Attitude Bolstering. Awareness and knowledge of security policy, its implications and guidelines about e.g. confidential information are necessary to make use of attitude bolstering. The self or existing believes and attitudes are strengthened and therefore the vulnerability to persuasive attempts can be reduced (Fransen et al., 2015). In this process people generate thoughts that support their attitudes (Lydon et al., 1988). As demonstrated by Xu and Wyer (2012) it is possible to generate a bolstering mind-set that decreases the effectiveness of persuasive attempts. This is even possible when the cognitive behaviour leading to this bolstering mind-set has been performed in an unrelated, earlier situation.

Decision Making. Information is processed by using two different systems as explained by Kahneman (2003): intuition and reasoning. Decisions are made based on either one. Butavicius et al. (2015) found the preference for a decision making style has a link to the susceptibility to persuasion, i.e. phishing. Decisions based on heuristics or mental shortcuts are intuitive, impulsive judgements that are more likely to be influenced by persuasive attempts. But interestingly it seems that the style of decision making can be modified by training. This would imply that recurring exposure to different social engineering approaches helps in establishing effective strategies to cope with social engineering. Furthermore, it demonstrates that solely education about the threats of social engineering is not sufficient.

7. Mapping of Defence Mechanisms against Psychological Principles

In order to get a better understanding how defence mechanisms work, we mapped them against the psychological principles (see Table 4).
Additionally, this mapping provides a structured representation regarding the applicability of a defence mechanism for a particular attack based on any of the psychological principals. Since knowledge is a fundamental requirement which is exerted in the dimensions attitude and behaviour, it is relevant for all principles.

Dimension		Psychological Principle/ Defence Mechanism	Authority	Social Proof	Liking, Similarity, Deception	Commitment, Reciprocation, Consistency	Distraction
Knowledge	**Attitude**	Persuasion Knowledge					
		Forewarning					
		Attitude Bolstering					
	Behaviour	Reality Check					
		Inoculation					
		Decision Making					

Table 4: Mapping of defence mechanisms against attacks based on psychological principals. Grey illustrates applicability of a defence mechanism, while black indicates non-applicability

As visualized above, there seem to exist two kinds of attacks based on the psychological principles. Firstly, attacks that are mainly defendable through the dimension of attitude, namely authority, social proof and distraction. Secondly, attacks that require a training of both dimensions, attitude and behaviour, in particular attacks based on liking, similarity, deception and commitment, reciprocation, consistency.

We first consider the dimension of attitude. Persuasion knowledge increases the likelihood of perceiving manipulation or deception attempts. Hence, it is relevant against all attack principles. In particular, since the main goal of social engineering attacks is to manipulate and influence the victims. In the same manner, forewarning is also relevant against all attacks based on the named psychological principles. Due to the fact that an attacker generally attempts to induce a pressure situation to his victim this mechanism generates awareness towards the malicious intentions. Especially, if forewarnings are combined with a precedent training of persuasion knowledge, the forewarning might trigger the recognition of attacks. Attitude bolstering is suitable for attacks based on Social Proof, Liking, Similarity, Deception, and Commitment, Reciprocation, Consistency if the security policies of the organisation are setup well. Mainly, because attacks based on this principles try to exploit a positive relationship built before the attack and/or try to assemble pressure due to societal obligations. However, the aim is to provoke the victim to practice a noncompliant security behaviour. By strengthening the attitude and improving awareness that not following the security policies can be harmful, attacks may be prevented. Given that the security policies are setup in a proper way. Attitude bolstering does not work well for attacks based on distraction and authority principles, since the main reason those attacks succeed is not that the victim is intentionally violating security policies. If those attacks

succeed, the user is either not aware that he is violating a policy or he is acting in good faith, obeying orders. The latter, is a fundamental principle of most organisational structures and therefore it would be risky to challenge this behaviour in a large scale. We discuss this idea in more detail at the end of this section.

Moving on to the dimension of behaviour, reality checks and inoculation are not applicable to attacks based on authority, because our societal system is based on and functions through authority. If a social proof is coherent, it just underlines how we function as a social being. And attacks based on distraction may not be countered, because limited attention is a characteristic that is not changeable. However, reality checks and inoculation are relevant to attacks based on liking, similarity, deception and commitment, reciprocation, consistency due to common unawareness that "naivety" in relationships and societal obligations is misused for this kind of attacks. Decision making is relevant to all kind of attacks since all attacks aim to influence the way the victim is making it's decisions, e.g. by not letting the victim think and getting him to rely on a heuristic. The defence mechanism helps in improving the victim's decision style or at a minimum evoke an awareness that decisions do not have to follow the first intuitive, impulsive reaction.

Another dimension we need to briefly mention is that the company should still preserve some kind of cooperative environment. Users should not overreact because they are afraid of being attacked. They still need to trust their colleagues to allow collaboration. Thus, another challenge to the user is to permanently do trade-offs between collaborating with his/her colleagues and avoiding/countering attacks. This often involves not following a policy for practical purposes, especially if they are contradictory and/or badly designed. As this kind of trade-off is very challenging to users and bears the risk that users are overburdened and simply give up in either of the two directions, as shown by Adams and Sasse (1999) in regards to user passwords.

8. A Gap Analysis of Missing Defence Mechanisms in IT Security against Social Engineering

As indicated above, the available defence mechanisms can be classified into the dimensions attitude and behaviour, which in turn exert knowledge. Table 5 presents a mapping of defence mechanisms comparing suggestions in IT security against techniques known in social psychology. When comparing the dimension attitude, the limited scope of IT security becomes evident. As established in Section 4, in the dimension attitude IT security considers establishment of policy and security awareness programs. The purpose of security awareness programs is twofold. Firstly, it is concerned with getting users to know and adhere to the established policy. Secondly, security awareness program's scope is usually limited to the provision of basic knowledge about social engineering. In comparison social psychology offers distinctively more. Although some approaches may be at least partly covered. Forewarning can be seen as included in the education of social engineering basics, as malicious intention of social engineers certainly belongs to basic knowledge about social engineering. But persuasion knowledge goes beyond social engineering basics as it includes knowledge about persuasion strategies as well as counter tactics to rely

on in any persuasive situation. For reliance on attitude bolstering good knowledge about security policy is necessary. Again IT security does the first step in user education, but fails in the second step, the enhancement of this knowledge. The use of attitude bolstering, implies not only the knowledge about policy but its implications and a thought process initiated by each user that strengthens his attitude to e.g. keep sensitive information private. The necessity to perform a reality check can directly be deduced from the concept of 'optimism bias', as illustrated in Section 3. It might partially be covered in security awareness programs. A reality check might be done for e.g. spam mails. But as this particular reality check has a technical background and people tend to dismiss their possible failure by it being a technical detail and in the same time greatly underestimating personal susceptibility, it is important to demonstrate to them their failure in a non-technical environment as well.

Dimension		**IT Defence Mechanisms**	**Psychological Defence Mechanisms**
Knowledge	**Attitude**	Policy Compliance	-
		Security Awareness Program	Forewarning
		-	Persuasion Knowledge
		-	Attitude Bolstering
		-	Reality Check
	Behaviour	Audit	-
		-	Inoculation
		-	Decision Making

Table 5: Comparison of defence mechanisms suggested in IT security and social psychology

Comparing mechanisms in Table 5 presents another crucial finding. The dimension behaviour is under-represented in IT security. The only suggestion made for this dimension is to verify correct behaviour via audits. But IT security fails to actually enhance secure behaviour. Training correct behaviour as part of security awareness programs is, as indicated in Section 4, recommended by only a few authors and is usually at most done for spam mails. Even though this is the application of inoculation, this is only one possible social engineering attack and a particularly technical one as well. Focus should again also be set on the persuasive nature of social engineering attacks. Hence trainings could for example include role plays. Additionally, it has been proven effective to alter the decision making process by conducting decision trainings where users make a "similar" decision in various appearances.

9. Discussion and Future Work

Previously, we have discussed (i) a mapping between defence mechanisms against attacks based on psychological principals and we identified (ii) gaps in IT security. Firstly, we want to elaborate on our findings regarding (i). While we provided a complete mapping, we are aware that it may be regarded as subjective. But as far as we are aware, this is the best structured comparison available. Furthermore, it is based on the results of our literature review and bears no experimental validation. To improve the mapping and furtherly validate it, we plan on conducting studies based on e.g. inoculation trainings to measure its influence regarding the psychological principles generally and particularly regarding the principles liking, similarity, deception and commitment, reciprocation, consistency. In a first step, we proposed a serious game (Beckers and Pape, 2016; Beckers et al., 2016) that helps players to understand how social engineering attacks work. The game can be played based on the real scenario in the company/department or based on a generic office scenario with personas that can be attacked. Our game trains people in realizing social engineering attacks in an entertaining way, which shall cause a lasting learning effect. In a next step we want to evaluate the collected data for further validation.

Secondly, we want to discuss the results regarding (ii). As indicated, both dimensions, attitude and behaviour, are represented inadequately in IT security when compared to recommendations from social psychology. To counter this gap, we envision two strategies for available security awareness programs (as shown in Table 5).
In a first strategy persuasion resistance trainings should be conducted.
They should include a broad approach to social engineering including psychological principles and their effects, possible counter strategies, the initiation of attitude bolstering. As optimism bias is a strong enabler of successful social engineering, it would be desirable to demonstrate users their susceptibility. This step is particularly promising, as it is feasible with little monetary effort. The second strategy is persuasive situation role plays. It is conceivable to include experiential exercises in this step as well as repeated decision trainings that force users to re-evaluate their knowledge and attitude by making a "similar" decision multiple times. This step is more effortful and it might suffice to only educate key personnel as it includes "live" training sessions guided by possibly costly trainers, actors or generally personnel capable of creating persuasive situations.

Dimensions	Future defence strategies
Attitude	Persuasion resistance training
Behaviour	Persuasive situation role plays

Table 6: Envisioned training strategies as part of security awareness

Additionally, it is worth to bear in mind, that although it is desirable to educate staff, there possibly exists a fine line to not overwhelm users with rules and knowledge.

10. References

Adams, A. and Sasse, M.A., 1999. Users Are Not the Enemy. *Commun. ACM*, 42(12), pp.40–46. Available at: http://doi.acm.org/10.1145/322796.322806.

Anon, 2011. Dimensional Research Study about Social Engineering. In *Analysis of Social Engineering Threats with Attack Graphs*. Beckers, Kristian Krautsevich, Leanid Yautsiukhin, Artsiom.

Bakhshi, T., Papadaki, M. and Furnell, S., 2008. A Practical Assessment of Social Engineering Vulnerabilities. In N. L. Clarke & S. Furnell, eds. *2nd International Conference on Human Aspects of Information Security and Assurance, {HAISA} 2008, Plymouth, UK, July 8-9, 2008. Proceedings*. University of Plymouth, pp. 12–23. Available at: http://www.cscan.org/openaccess/?paperid=53.

Barrett, N., 2003. Penetration testing and social engineering: hacking the weakest link. *Information Security Technical Report*, 8(4), pp.56–64.

Beckers, K. and Pape, S., 2016. A Serious Game for Eliciting Social Engineering Security Requirements. In *RE*.

Beckers, K., Pape, S. and Fries, V., 2016. HATCH: Hack and Trick Capricious Humans - A Serious Game on Social Engineering. In *BCS HCI*.

Brehm, J.W., 1966. *A theory of psychological reactance*, New York: Academic Press.

Briñol, P., Rucker, D.D. and Petty, R.E., 2015. Naïve theories about persuasion: Implications for information processing and consumer attitude change. *International Journal of Advertising*, 34(1), pp.85–106.

Butavicius, M. et al., 2015. Breaching the Human Firewall : Social engineering in Phishing and Spear-Phishing Emails. *Australasian Conference on Information Systems*, pp.1–11.

Cacioppo, J.T. et al., 1986. Central and peripheral routes to persuasion: An individual difference perspective. *Journal of Personality and Social Psychology*, 51(5), pp.1032–1043.

Cialdini, R.B., 2009. *Influence: the psychology of persuasion* EPub editi., New York: Collins.

Ferreira, A., Coventry, L. and Lenzini, G., 2015. Principles of Persuasion in Social Engineering and Their Use in Phishing. In T. Tryfonas & I. Askoxylakis, eds. *Human Aspects of Information Security, Privacy, and Trust SE - 4*. Lecture Notes in Computer Science. Springer International Publishing, pp. 36–47. Available at: http://dx.doi.org/10.1007/978-3-319-20376-8_4.

Frangopoulos, E.D., Eloff, M.M. and Venter, L.M., 2012. Psychosocial Risks: can their effects

on the Security of Information Systems really be ignored? In N. L. Clarke & S. Furnell, eds. *6th International Symposium on Human Aspects of Information Security and Assurance, {HAISA} 2012, Crete, Greece, June 6-8, 2012. Proceedings*. University of Plymouth, pp. 52–63. Available at: http://www.cscan.org/openaccess/?paperid=35.

Fransen, M.L. et al., 2015. Strategies and motives for resistance to persuasion : an integrative framework. *Frontiers in psychology*, 6(August), pp.1–12.

Friestad, M. and Wright, P., 1994. The Persuasion Knowledge Model: How People Cope with Persuasion Attempts. *Journal of Consumer Research*, 21(1), pp.1–31. Available at: http://www.jstor.org/stable/2489738.

Gigerenzer, G. and Todd, P.M., 1999. *Simple Heuristics that Make us Smart*, Oxford: Oxford University Press.

Gilbert, D.T., 2002. Inferential correction. In T. Gilovich, D. Griffin, & D. Kahneman, eds. *Heuristics and biases*. New York: Cambridge University Press.

Gouldner, A.W., 1960. The Norm of Reciprocity: A Preliminary Statement. *American Sociological Review*, 25(2), pp.161–178. Available at: http://www.jstor.org/stable/2092623.

Gragg, D., 2003. A multi-level defense against social engineering. *SANS Reading Room, March*, 13.

Gulati, R., 2003. The Threat of Social Engineering and your defense against it. *SANS Reading Room*.

Ivaturi, K. and Janczewski, L., 2011. A Taxonomy for Social Engineering attacks. *Proceedings of CONF-IRM*.

Jones, C., 2004. Understanding and Auditing. *SANS Institute Infosec Reading room*.

Kahneman, D., 2003. A perspective on judgment and choice: mapping bounded rationality. *The American psychologist*, 58(9), pp.697–720.

Landman, J. and Petty, R., 2000. "It Could Have Been You": How States Exploit Counterfactual Thought to Market Lotteries. *Psychology & Marketing*, 17(4), pp.299–321.

Lott, A.J. and Lott, B.E., 1965. Group Cohesiveness as Interpersonal Attraction: A Review of Relationships with Antecedent and Consequent Variables. *Psychological Bulletin*, 64(4), pp.259–309.

Lydon, J., Zanna, M.P. and Ross, M., 1988. Bolstering Attitudes by Autobiographical Recall: Attitude Persistence and Selective Memory. *Personality and Social Psychology Bulletin*, 14(1), pp.78–86. Available at: http://psp.sagepub.com/content/14/1/78.abstract.

Lynn, M., 1989. Scarcity effects on desirability: Mediated by assumed expensiveness? *Journal of Economic Psychology*, 10(2), pp.257–274.

Manske, K., 2009. An Introduction to Social Engineering. *Information Security Journal: A Global Perspective*, 9(5), pp.1–7.

Milgram, S., 1963. Behavioral study of obedience. *The Journal of Abnormal and Social Psychology*, 67(4), p.371.

Milgram, S., 1974. *Obedience to authority*, London: Tavistock.

Miller, C.H. et al., 2013. Boosting the Potency of Resistance: Combining the Motivational Forces of Inoculation and Psychological Reactance. *Human Communication Research*, 39(1), pp.127–155.

Mitnick, K.D. and Simon, W.L., 2011. *The art of deception: Controlling the human element of security*, John Wiley & Sons.

Petty, R.E. and Cacioppo, J.T., 1996. *Attitudes and persuasion: Classic and contemporary approaches*, Boulder, CO, US: Westview Press.

Pfleeger, S.L., Sasse, M.A. and Furnham, A., 2014. From Weakest Link to Security Hero: Transforming Staff Security Behavior. *Journal of Homeland Security and Emergency Management*, 11(4), pp.489–510.

Rusch, J.J.J., 1999. The " Social Engineering " of Internet Fraud. *nternet Society's INET'99 conference*, pp.1–12. Available at: http://www.isoc.org/isoc/conferences/inet/99/proceedings/3g/3g_2.htm.

Sagarin, B.J. et al., 2002. Dispelling the illusion of invulnerability: The motivations and mechanisms of resistance to persuasion. *Journal of Personality and Social Psychology*, 83(3), pp.526–541. Available at: http://doi.apa.org/getdoi.cfm?doi=10.1037/0022-3514.83.3.526.

Scheeres, J.W., 2008. *Establishing the human firewall: reducing an individual's vulnerability to social engineering attacks*,

Schneier, B., 2008. The Psychology of Security. In S. Vaudenay, ed. *Progress in Cryptology – AFRICACRYPT 2008 SE - 5*. Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 50–79. Available at: http://dx.doi.org/10.1007/978-3-540-68164-9_5.

Stajano, F. and Wilson, P., 2011. Understanding Scam Victims: Seven Principles for Systems Security. *Commun. ACM*, 54(3), pp.70–75. Available at: http://doi.acm.org/10.1145/1897852.1897872.

Stanovich, K.E. and West, R.F., 2002. Individual differences in reasoning: Implications for the rationality debate. In T. Gilovich, D. Griffin, & D. Kahneman, eds. *Heuristics and biases*. New York: Cambridge University Press.

Thornburgh, T., 2004. Social Engineering: The "Dark Art." In *Proceedings of the 1st Annual Conference on Information Security Curriculum Development*. InfoSecCD '04. New York, NY, USA: ACM, pp. 133–135. Available at: http://doi.acm.org/10.1145/1059524.1059554.

Verizon, 2012. Data Breach Investigations Report. Available at: http://www.verizonenterprise.com/resources/reports/rp_data-breach-investigations-report-2012-ebk_en_xg.pdf [Accessed January 13, 2016].

Verizon, 2013. Data Breach Investigations Report. Available at: http://www.verizonenterprise.com/resources/reports/rp_data-breach-investigations-report-2013_en_xg.pdf [Accessed January 13, 2016].

Veseli, I., 2011. *Measuring the Effectiveness of Information Security Awareness Program*. Gjøvik University College.

Warkentin, M. and Willison, R., 2009. Behavioral and policy issues in information systems security: the insider threat. *European Journal of Information Systems*, 18(2), pp.101–105.

Weinstein, N.D., 1980. Unrealistic Optimism About Future Life events. *Journal of Personality and Social Psychology*, 39(5), pp.806–820.

Winkler, I.S. and Dealy, B., 1995. Information Security Technology?...Don't Rely on It A Case Study in Social Engineering. In *Fifth Usenix Security Symposium*. pp. 1–6.

Xu, A.J. and Wyer, R.S.J., 2012. The Role of Bolstering and Counterarguing Mind-Sets in Persuasion. *Journal of Consumer Research*, 38(5), pp.920–932. Available at: http://www.jstor.org/stable/10.1086/661112.