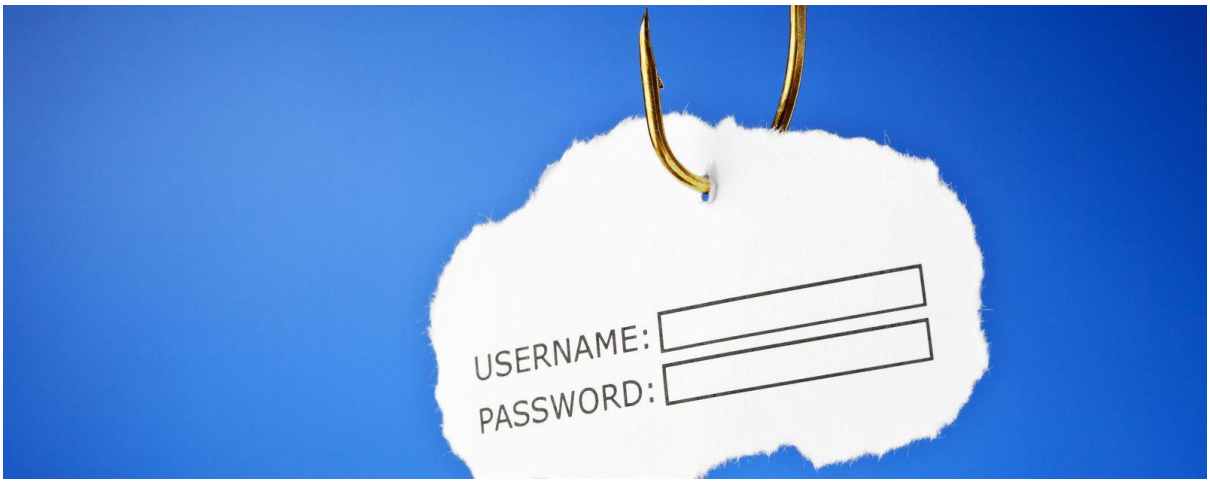
A systematic method to execute simulated phishing tests

By Demetris Antoniou



Deloitte.

TU Delft

Delft, The Netherlands, 2015

Acknowledgements.

This thesis represents the end of my studies as a master student at the Technical University of Delft. The past two years have been an eye-opening experience; both in terms of formal education and in all the experiences I have gained by living in the Netherlands. This has been effected by the numerous people I had the opportunity to converse with, work with, laugh with and the amazing times I have had here. More importantly, however, it is experiences like these that prepare one for the future.

The story on how this project came to fruition is a peculiar one. It all started during a lecture on computer security where I asked the professor whether there is any research done on Social Engineering. The professor then instructed me to talk to Dr. Wolter Pieters who is conducting research on the socio-technical aspects of cyber-security within the TREsPASS project. Wolter then got me in touch with Dr. Trajce Dimkov who served as the external supervisor for this project under the auspices of Deloitte. At the time of writing this acknowledgment I am a full-time consultant at Deloitte. My most sincere thanks go to these two men who have been instrumental in both the ending of my academic career (for now?) and the beginning of my professional one. Wolter has been an amazing supervisor who always had meaningful advice on how the thesis should proceed. The project went through several iterations and mishaps yet Wolter was always there to give form to that chaos of information that would come and go. I am thankful for his help and his guidance and even in cases where I would feel lost, Wolter was always finding the way to make sense out of the situation. If Wolter was the Ying Trajce was the Yang. His advice was instrumental in formulating the thesis in a business context and produce something that would provide value to the organization and society in general. Our bi-weekly thesis meetings evolved into mentoring sessions where he, with his pointed remarks, molded me into a professional in way of thinking, understanding and behaving. To both, I owe you my sincerest thanks for being the ones who stepped me up into the next chapter of my life. Many thanks also go to Drs. Marianne Junger and Jan-Willem who provided extensive help in the early stages of the project.

I would also like to thank Dr. Robert Verburg and Dr. Michel van Eeten for being on my committee. They have provided invaluable insight on the project and their advice was instrumental in decisions that had to be taken which eventually gave the project its final form. It is through the difficulties that good things come out of and without their eye-opening remarks, this project would not have been at the quality that it is now. The green-light-turned-to-midterm session will be engraved in my memory as one of the best things that happened to me. For breaking the weak foundation and making me build a stronger one, I owe you many thanks. Thanks also go to the many people who were directly or indirectly involved in the shaping of the project. I would like to especially thank Katrien for being my comrade in our Deloitte internship, Hugo and Peter for their out-of-this-world phishing expertise. The entire Deloitte hacking team is truly an exceptional group of people; the trainings I received as part of it have been some of the most memorable learning experiences of my life. I also thank Anamaria for helping me with the translations.

Last but not least, come all the people who have put their brick in shaping me as a person. My physics teacher, who taught me to be better than expected; my history teacher, who vividly described the world and motivated me to discover it; the Greek military officer, who taught me that how you say something is more important than what you say. To my parents, who never doubted me and will always be there for me. And to Georgia, my mainstay, the joy of my life and the world's most beautiful smile (and eyes!)

Management Summary.

This research project was done for the Management of Technology master program at TU Delft in cooperation with Deloitte Nederland. The purpose of this project was to systematize the creation of simulated phishing test e-mails and to enable their measurement so as to facilitate the statistical analysis of their results. This project is focused on simulated phishing tests which are distinct from regular phishing e-mails both in their background and form.

Phishing is a scalable act of deception whereby impersonation is used to obtain information from a target. It poses a risk to both individuals and corporations due to the loss value that is incurred as a result of a successful phishing attack. Organizations have started conducting Simulated Phishing Tests (SPTs) to measure the susceptibility of their employees to actual phishing attacks, compare to others in their industry and to improve awareness through training. Deloitte is offering SPTs to other companies who would like to do so. This has resulted in a unique dataset with the results of previous SPTs. The dataset is unique because every e-mail is accompanied by its click-rate; something not found on public phishing depositories. Such an expanding dataset can be analyzed so as to gain insight around the *nature* of the e-mail and make statistical inferences. The purpose of this project is to create a coding scheme which will enable this analysis and be used to conduct SPTs in a systematic way. Systematic here means that concrete design decisions will be formulated which the maker of the SPT can follow and make the e-mail according to. These design decisions will result in an e-mail which will also have a set of enumerated variables accompanying it. For example, an e-mail can purposively be designed to use a happy event in its message whereas another can make its focus on corporate policies. These variables can then be correlated to the actual click-rate of the e-mail which will can result in predictive models regarding the click-rates of SPTs in the future. These predictive models can then be used to create phishing campaigns of increasing difficulty; where difficulty could be numerically expressed. Currently (2015), phishing has evolved into a more personalized form also known as *spear-phishing* where the attacker will modify their phishing e-mail to make it more relatable to the recipient. For example, a phishing attack is only sent to employees of a particular company using that company's e-mail structure and logo. This introduces several variations of the regular phishing attack which is sent to thousands or millions of people without any form of personalization or target segmentation. SPTs conducted by industry are similar to spear phishing attacks and are usually sent to various departments of a client company and purport to be addressed solely to employees of the company.

The first step in fulfilling the purpose of the project was in performing an extensive literature review on phishing, how it evolved and the current state of affairs. The literature review set the definitions and investigated current research on phishing. It has been identified that most research attempts to investigate differences between individuals and their susceptibility to phishing. However, little research has investigated the effect of the e-mail itself to the click-rate. This can be partly explained by the lack of data on click-rates over several different e-mails; something offset by our dataset. The literature review established the definition of phishing as a deceptive piece of text. This definition then led to an exploration of deception in general and focused on deception in asynchronous computer-mediated communications. The result was a set of computer readable indicators on deception which were the first part of the coding scheme. Then, the persuasive elements of phishing attacks were investigated. This resulted in the most populous set of variables for the project which are drawn from various established theories such as Maslow's theory on motivation and Cialdini's principles of influence. The enumeration of these variables needs human coders.

The final set of variables which describe a simulated test e-mail were derived from the existing dataset of previous SPTs and from the literature. This resulted in variables which are unique to the dataset or were viewed as important and were not found in literature. These variables also require rating by coders. No conclusions can be drawn on the effect of each variable to the click-rate but informed assumptions on the expected effect are provided. The number of tests available is too small to make any inferential analysis at the moment; this project provides a measurement scheme which will enable it in the future. Moreover, guidelines are provided on how to conduct experiments to discover the actual impact of each variable on the click-rate. By using the results of these experiments e-mails can then be designed around their perceived *difficulty.* More difficult e-mails are harder to detect as phishing e-mails, are much more persuasive and thus are expected to have a higher click-rate than less difficult ones.

The results of the literature review and the dataset were then combined and the coding scheme was operationalized. A total of thirty-five variables were included in the coding scheme which were drawn from literature and the dataset. They can be found in Table 18 in Chapter 5. Questions were formulated for the human coded variables with brief explanations on the nature of the variables and the importance of coder training is discussed. Using human coders introduces considerable subjectivity in the measurement of e-mails therefore training is of paramount importance in maintaining the replicability and reliability of the coding scheme.

Recommendations are also provided on what can be done with the coding scheme for future SPT engagements. The main benefits of using the coding scheme are outlined below:

- SPAs can now be constructed in a systematic way where the characteristics of the e-mail can be manipulated along scientifically backed principles.
- The resulting e-mails from SPTs will have a measurement accompanying them which can be correlated to the resulting click-rate.
- The e-mails sent to the same organization over time can be created with ascending difficulty which can be a better indicator on the efficacy of anti-phishing awareness training.

A methodological approach is proposed which can enable the conduct of experiments which can find the effect of each variable on the click-rate of the e-mail. This experimentation can enable the finding of truth in what the impact is of each variable to the click-rate. Organizations who apply this approach will be able to gain more information about their phishing tests. The use of A/B testing is recommended where the recipients are split into two groups where there is equal representation of all subject specific factors. This ensures that two different e-mails are sent to two groups which are similar to each other as possible. With such an approach the effects of different organizations have on the impact of the click-rate are diminished and the characteristics of the e-mail can be investigated with minimal bias. A/B testing can provide concrete numerical results on the impact that individual variables have on the click-rate of the e-mail.

In addition to the coding scheme and experimental guidelines recommendations are made on how statistical analysis of the results should be done. Simpler linear regression models are recommended for analysis when the number of samples is small and the adoption of more complex statistics is recommended as the dataset expands. The approach best suited for analysis when there is a sufficiently large dataset is Neuro-Fuzzy logic since it can incorporate the experimental results from A/B testing

whilst shaping the membership functions. In addition, fuzzy-based models can handle linguistic input well which is what the answers to the coding questions are. The approach is then expected to produce an accurate predictive model which will also be understandable by researchers as it provides a good trade-off between explanatory and predictive power.

The implementation of the project's recommendations will enable future conduct of SPTs in a scientific, enumerable and structured way which will enable the creation of scientifically backed phishing e-mails that will be ready for statistical analysis in the future.

This project has been conducted in cooperation with Deloitte Nederland B.V. who have graciously provided the infrastructure, data and expertise in the conduct of simulated phishing tests.

Table of Contents

Acknowledgements. 1

Management Summary. 3

1. Introduction. 8

1.1. A brief overview of Deloitte's SPT approach. 10

1.2. Problem Identification. 10

1.3. The business benefits of the solution. 11

1.4. The research opportunities of the solution. 11

1.5. Research Objective and Research Questions 12

1.6. Methodology and Thesis Structure 13

1.7. Expected Results 14

2. Scientific Foundation 15

2.1. Social Engineering 15

2.2. Phishing 17

2.3. Spear Phishing and Targeted Attacks 18

2.4. Trends in spear phishing 19

2.5. Existing research on phishing: the lack of focus on the e-mail 21

2.5.1. The e-mail and the individual: the bigger picture 23

3. Deception and Persuasion 25

3.1. Deceptive elements of phishing 25

3.2. Persuasive elements of phishing 28

3.3. Extraction of phishing characteristics from literature 32

4. Dataset Inspection 37

4.1. Ordering of the current dataset 37

4.2. Exploration of representative examples 41

4.3. Identification of common themes 48

4.4. Extraction of phishing characteristics from the dataset 49

4.5. Limitations of the dataset 50

4.6. Possibilities of the dataset 50

5. The coding scheme and its application. 51

5.1. The coding scheme 51

5.1.1. On the method used in analyzing the e-mails. 53
5.2. Informed Assumptions on the impact of variables. 53
5.3. The importance of coders and practical considerations. 55
5.3.1. Training of coders. 58
5.4. Quantitative attempts at the dataset and recommendations on analysis. 60
5.5. Towards prediction: Experimentation using the coding scheme 66
6. Evaluation of the coding scheme. 70
6.1. Evaluation of the coding scheme: a first test 70
6.1.1. The objectivity of our coding scheme 70
6.1.2. The reliability of our coding scheme 71
6.1.3. The replicability of our coding scheme 71
6.1.4. The systematicity of our coding scheme 71
6.2. Using the coding scheme for future SPT engagements 72
6.1. Limitations of the coding scheme. 72
7. Conclusion 73
7.1. The deliverable 73
7.2. Conclusion 75
7.3. Recommendations 75
7.4. Reflection 76
Contribution to the Management of Technology Program. 77
List of Figures. 78
Bibliography. 79

1. Introduction.

Ever since their inception, computers were always operated by humans. As stand-alone computers evolved into networked computer systems so did their complexity, usefulness and accessibility. This resulted however, to the growth of malicious users who are willing to use the Internet for fraud, illegal profit, vandalism and other forms of illegal activity. These people are commonly (and often incorrectly) known as hackers who actively try to gain unauthorized access to computer systems. Although most people see hackers as highly gifted individuals with curiosity the current situation (2015) is different. Organized Crime groups use cyber-space more and more to gain illicit financial income. [1] One of the methods that malicious hackers use is the so-called 'Social Engineering'(SE). One of the earliest definitions of social engineering within the context of cyber-security came in 1995 when it was stated that SE is "the process of using social interactions to obtain information about a victim's computer system".[2] Although often overlooked[3], SE is a real and damaging threat to Information Based organizations. For example, many Cyber-security consultants claimed 100% success rate whenever they employed SE to penetration testing.[1] Academics have realized the imbalance that exists between the research conducted in the technical domain of Cyber-Security and the social domain. The conclusions of Bjorck and Yngstrom state that the human element of security is very impactful on information security. In their review of existing publications they have identified that 80% of research conducted on Cyber-Security is about technical issues whereas only 20% deals with the social aspects. [4] This makes research into SE an interesting and necessary research topic. Phishing is a subset of social engineering whose latest exact definition will be discussed later.

Phishing was first mentioned as such on January 2, 1996 on a Usenet newsgroup.[2] The origins of phishing can be traced back to the time when America Online (AOL) was the prominent provider of internet in the U.S. At the time, attackers were sending false e-mails to AOL users and managed to extract their login credentials which in turn lead to the conduction of spam attacks using the hijacked accounts.[5] Attackers then moved on to trying to elicit financial information from their victims, especially credit card numbers. At the time however, the amateur nature of phishing e-mails was a good enough giveaway for to-be victims who quickly learned to recognize phishing attempts usually through the bad spelling and grammar. Nowadays, most e-mail accounts will have their spam inboxes filled with various offerings of cheap drugs, lottery winnings and account suspensions. However, phishing has evolved into something more dangerous and professionally made.

The widespread use of Social Media has provided malicious social engineers with large amounts of information that can be used against a target. [6] This leads to the phenomenon of *spear* phishing or *context-aware* phishing where the e-mail is targeted. This means the e-mail is tailored to the recipient based on their employment, family status or anything that distinguishes them from others. [7] In contrast, regular phishing is one e-mail sent to many recipients (sometimes hundreds of thousands)

[1] http://www.techrepublic.com/blog/it-security/securitys-weakest-link-technology-no-match-for-social-engineering/

[2] http://www.phishing.org/history-of-phishing/

without any personalization. The type of phishing under study here is the former spear-phishing type. *Simulated Phishing Tests* (SPTs) conducted by industry are also of the spear-phishing type. Existing research revolves around mass phishing where there are well studied elements of the e-mail which are promptly filtered by existing software. [8] However, there seems to be little research specific to simulated phishing tests (SPTs) conducted by industry with the goal of raising awareness toward phishing.

There are numerous instances where companies have been victimized by means of phishing. One of the most striking examples, is the one done on RSA in 2011. RSA is a computer security organization mostly known for their SecurID two-factor authentication service. In 2011 they were successfully infiltrated by means of a spear-phishing attack.[3] Over a two day period an attacker (presumably state-backed) sent two e-mails to a four RSA employees. Although the e-mails were filtered into the spam folder, one employee actually took the e-mail out of the spam folder and opened the malicious attachment. The targeted employees were not even critical members of the firm such as high management or system administrators. Nevertheless, the SecurID service was severely compromised which resulted in the recall of most security tokens since users of SecurID such as Lockheed-Martin reported that attacks were being carried out using the compromised SecurID's. Technology news website, Wired.com, also obtained a screenshot of the actual spear phishing e-mail (EMC is the parent company of RSA):

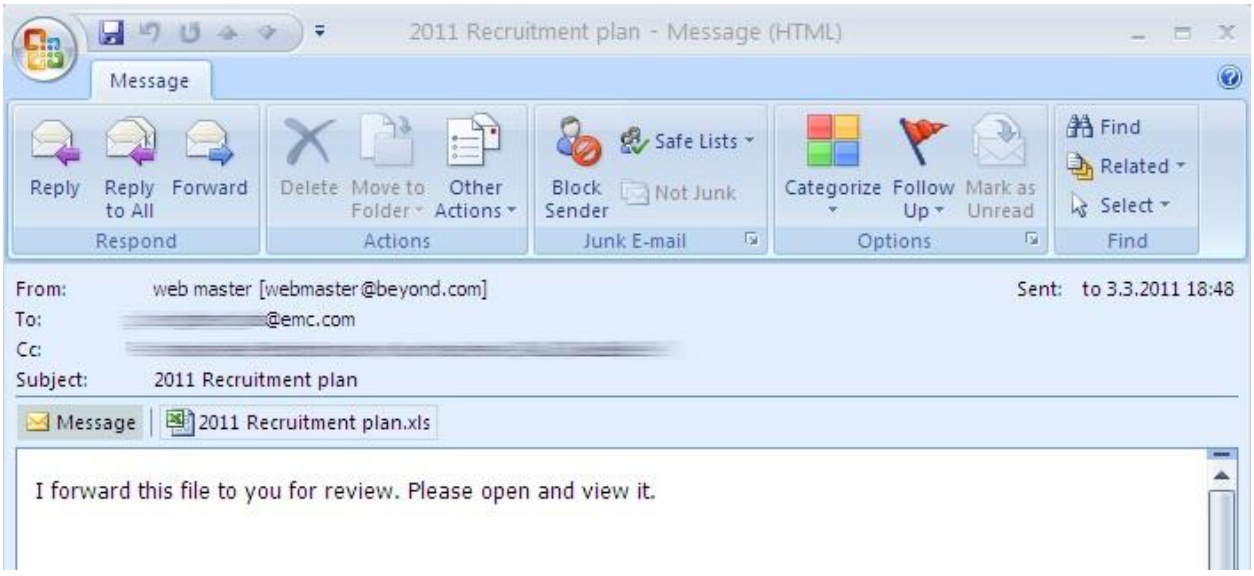


Figure 1: The spear-phishing e-mail which compromised RSA. Source: wired.com

As can be seen from above the spear-phishing e-mail was not extremely lengthy or elaborate. However, the title 'Recruitment Plan' seemed to be a good enough trigger for an employee to open it. Someone who was unsure of their job security or an intern could have opened it due to their need to know whether they were included. The above example demonstrates that even companies which are in the business of computer security can still be infiltrated using spear phishing tactics. Organizations have

[3] https://blogs.rsa.com/anatomy-of-an-attack/

recognized the risk that such attacks present to the continuity of their business. Therefore, industry has begun to incorporate phishing awareness training as part of their overall cyber-security strategy.

Gartner has claimed in a recent report that organizations which employ what it calls, Anti-Phishing Behavior Management (APBM), "will experience up to 30% less breaches than organizations which rely solely on security awareness."[9] The report recommends that chief information security officers 'should leverage APBM solutions aggressively to establish, reinforce and maintain appropriate security behavior within all employee groups.' Statements like these from organizations like Gartner point to the fact that SPTs will become, in the future, a standard in an organization's cyber security strategy.

1.1. A brief overview of Deloitte's SPT approach.

Simulated phishing attacks are conducted by industry towards organizations who want to gage the susceptibility of their employees to spear phishing and to better educate employees on how to detect and thwart phishing attacks. Deloitte in the Netherlands has been conducting such engagements since 2014 with more engagements planned in 2015. The engagements are conducted as follows:

1. The client agrees to the execution of a phishing exercise. The client's departments that will take part are decided upon and timeframes are agreed.
2. A number of phishing scenarios are created according to the client's specifications. The scenarios are presented to the client who will choose and may suggest changes to the e-mails.
3. Using a specialized infrastructure the e-mails are sent out within the agreed timeframe and responses to them are recorded. Each employee has a unique but anonymous ID with which their actions can be traced.
4. After some time has passed, the results are collected and a report is sent to the client. The results include the percentage of recipients who clicked as well as to what extent they interacted with the phishing website. They could have just opened the link on the phishing e-mail or went all the way through and divulged their username/password information. This project will deal only with the e-mail and whether there was a click or not.
5. The report is either sent as-is or follow-up training is provided. There is also the option to debrief the employees at the time they are phished and they can be re-directed to an e-learning platform.

The service offered by Deloitte has yielded an interesting and expanding dataset which not only has the e-mail itself and the phishing website but also the results of each exercise. This makes the data generated unique in the sense that is accompanied with data pertaining to its click rate. Publicly available phishing databases such as the Anti-Phishing Working Group (APWG) have an enormous amount of phishing e-mails but not accompanying data on their targets or their click rates. The data generation of SPTs present a massive opportunity to conduct statistical analysis in the future which will be useful for both the business and science in general.

1.2. Problem Identification.

The experience obtained from the current SPT engagements has generated a number of questions whose answer is not clear at first glance. For example, a particular client signed up for a phishing campaign where a total of four phishing e-mails would be sent over several month to their employees

with training sessions conducted in-between. After the first e-mail there was a marked decrease in the percentage of employees who fell for the phish but in the last e-mail the response rate jumped back to pre-test levels. This presents the interesting question on whether the e-mail itself can have an impact on the click rate of the engagement. The question that each e-mail can be measured according to its intrinsic features rate can serve as a starting point in finding the answer to this question.

Industry's current SPTs have yielded a number of different 'scenarios' which have sometimes been used several times. However, distinctions are made only regarding the context of the e-mails' scenarios. For example, there is the ethical guidelines scenario, there is the work from home scenario and so forth. This classification, however, is rather crude and dependent entirely on the specific context of the e-mail. At present, there can only be comparisons of the click rate between scenarios without any significance testing between the results. If the e-mails were viewed in a more context-independent way then more meaningful comparisons could be made. For example, the e-mails could be rated on the nature of their language along the same scale which would allow for comparisons to be made regardless of the exact nature of the scenario. This will allow for scientific analysis of the e-mails as well as offer new avenues for SPT engagements.

In summary, the main problems which arise from the current SPT approach are that 1) the creation of the phishing e-mails happens in an ad-hoc way and results in unstructured e-mails and 2) the resulting e-mails do not contain any quantifiable information and do not pave the way for the use of predictive analytics in the future.

1.3. The business benefits of the solution.

An alternative approach to the creation of SPT e-mails will yield significant benefits for the business. For one, it will enable the creation of e-mails which may seem to be different context-wise but have similar properties which contribute to their click rate. Moreover, the ability to use a structured approach to generate the e-mails will allow for non-experts to make them with the same quality of an expert. In addition, it can help new practitioners understand the underlying principles of a phishing e-mail.

SPT engagements are sometimes conducted in a campaign fashion. For example, over the course of a year an organization may sign up for several e-mails to be sent. With the use of a structured approach, the e-mails can be adjusted so as their deceitfulness and persuasiveness can be manipulated. At the moment no claim can be made on whether they would influence the click rate positively or negatively but their mere manipulation will enable discussions to occur along their lines. The e-mails can be made progressively more persuasive and/or deceptive which will have the result of educating employees in recognizing phishing e-mails of various types. Training sessions can be designed to train employees to understand the more subtle elements of e-mails which were manipulated by the sender. In the end, evidence-based phishing can be conducted where there can be a very reliable prediction on how well a particular e-mail will perform.

1.4. The research opportunities of the solution.

Science can also benefit from a structured approach in the making of phishing e-mails. By coding the generated e-mails with a predefined scheme will allow for their analysis in the future. Every e-mail sent as part of a SPT will be accompanied by a number of features which can be studied or manipulated for scientific reasons. It will also enable the design and execution of experiments in the A/B testing

methodology. For example, two e-mails can be made with identical properties except one. Combined with proper sampling, real scientific data can be produced on the impact of individual properties on the click rate.

The results of SPT engagements can be used in other research and can also be corroborated with research conducted on the susceptibility of individuals to phishing. In the future, a complete mapping of all the factors which influence the likelihood to respond to a phishing can be made which come from both the recipient and the e-mail.

1.5. Research Objective and Research Questions

Following the problem identification the main research objective of the project will be defined. The research objective will be broken down into specific research questions whose answer will stem from the sections of the thesis.

The research objective comes from the identified problems. In order to formulate our research objective it has to fully fulfill the requirements in order to solve the problems. To generate the research objectives, a problem-centered approach will be used as stipulated in the work of Peffers et al.[10] Although Peffer's work encompasses the entirety of a design process we will only use the first step of the process; where objectives are to be matched with the requirements and problems. Requirements stem from the identified problems. The summary of section 1.2 identified two main problems. The table below shows the problems, their requirements and the objectives of the project:

Table 1: Problem identification, requirements and objectives.

Problem	Requirement	Objective
Phishing e-mails are generated in an ad-hoc way. Result is dependent on the skill of the specialist.	Enable the structured generation of phishing e-mails based on predefined characteristics.	Develop a coding scheme which will allow the generation of e-mails in a structured way and enable their measurement for future analysis
E-mails generated in the current SPT architecture are difficult to measure based on their characteristics.	Enable the coding of phishing e-mails based on predefined characteristics.	

Thus, the research objective of this project is:

To create a coding scheme which will enable the generation of SPT e-mails in a structured way and which will enable the analysis of the results of SPT engagements.

In terms of deliverables, the research objective is to create a coding scheme for SPT generated e-mails. The deliverable can then be used to both rate e-mails and guide their creation.

To fulfill our research objective a number of research questions need to be answered. The questions begin with the more fundamental elements and move on to the specifics of the objective:

Research Question 1: What is the state-of-the-art in phishing research?

Sub-Question 1.1: What is spear phishing?

Sub-Question 1.2: What research has been conducted on phishing?

Research Question 2: What characterizes a spear-phishing e-mail?

Sub-Question 2.1: What are the deceptive variables of a spear-phishing e-mail?

Sub-Question 2.2: What are the persuasive variables of a spear-phishing e-mail?

Research Question 3: What are the characteristics of SPT e-mails?

Sub-Question 3.1: What are the common themes found in the existing set of previous SPT engagements?

Research Question 4: How can the e-mails be coded based on the discovered characteristics?

Sub-Question 4.1: How can the coding scheme be used to enable future measurement of e-mails and estimate their click rate?

Sub-Question 4.2: How can the coding scheme be used to guide the generation of SPT e-mails?

Research Question 5: How can the coding scheme be evaluated and implemented?

Sub-Question 5.1: How good is the coding scheme based on coding scheme evaluative literature?

Sub-Question 5.2: How can the coding scheme be implemented?

The above questions and sub-questions will be answered in the next chapters of the thesis.

1.6. Methodology and Thesis Structure

In this section the way with which the research questions will be answered and the research objective achieved will be described. Research Question 1(RQ1) will be answered by reviewing the relevant literature. The literature will explain what phishing is, what spear-phishing is and how it evolved from regular phishing and round it up with a review on existing research regarding phishing. This will give us first a definition of phishing and spear-phishing and then by the literature review we will identify what current research has found on the subject. The focus of RQ1 will be first to find the proper definitions whose wording will guide further literature research. The knowledge gained from RQ1 will then enable us to dive deeper into the more semantic characteristics of phishing e-mails namely the deceptive and persuasive elements of phishing e-mails and answer RQ2. The answer to RQ2 will come from literature drawn from general papers of deception and persuasion and focusing in on more phishing related publications. The answering of RQ2 will provide the phishing characteristics derived from literature. After the literature-derived properties of phishing e-mails are defined, we will move on in answering RQ3 by inspecting the current dataset generated by SPTs. This will be done by qualitatively inspecting the dataset and *coding* the e-mails. Coding is done by reading the data and segmenting it according to the goal of the study.[11] This inspection will attempt to identify elements of the SPTs dataset which may not have been explored in the literature and/or are specific to SPTs. Anything identified in RQ2, however, will not be included in this inspection nor will it affect or guide the inspection in any way. We

will then move on to bring together the characteristics identified from both the literature and the dataset and develop the overall coding scheme of phishing e-mails. The project will then explore the possibilities, limitations and propose applications of the coding scheme.

The methodology for achieving the research objective has two main parts. First is the literature review. In this first part we will attempt to bring together research that has been conducted regarding phishing in general with emphasis on the textual elements of the e-mail. It is important to note that we will not deal with standard phishing cues which have already seen research. This includes the inclusion of links or weird URLs and how computers can identify them. Our goal here is to study the e-mail as understood by a human and not to develop ways to enable computers to detect phishing e-mails.

The thesis will follow a structure which will answer the research questions and, ultimately, fulfill the research objective. Section 2 will be the scientific foundation of the project i.e. the literature review. It will define and explore social engineering, phishing and spear phishing and discuss recent work done on the subject. This will answer the first research question. Section 3 will further explore the theories of deception and persuasion and how they connect to spear phishing. Section 3 will result in the literature backed characteristics of phishing e-mails and consequently answer the second research question. In section 4 we will take a closer look at the current SPT dataset and extract the relevant features, thus answering the third research question. All the work will converge in section 5 which will coalesce the knowledge gained and create the coding scheme. The final and sixth section will summarize the results, crystallize the deliverable, suggest opportunities for future work and reflect on the thesis as a whole. The structure of the thesis is shown in tabular form below:

Table 2: The thesis structure.

Section Number	Section Name	Purpose
1	Introduction	Problem Statement, Research Objective
2	Scientific Foundation	Answer Research Question(RQ) 1
3	Deception and Persuasion	Answer RQ2
4	Dataset Inspection	Answer RQ3
5	The coding scheme and its application	Answer RQ4
6	Evaluation of the coding scheme	Answer RQ5
7	Conclusion	Conclusion, Deliverable, Recommendations, Reflection

1.7. Expected Results

The main deliverable of the thesis will be a coding scheme. What is meant is that there will be a method with which the e-mail will be measured according to its characteristics. This coding scheme will have two purposes. One, is to enable the generation of phishing e-mails according to scientifically backed principles. Second, is the measurement of the e-mails. For the existing dataset, the coding scheme will measure them. In future engagements, the coding scheme will both guide their design and measure them.

As stated earlier, the generation of new phishing e-mails is done in a pretty much ad-hoc way. The specialist will have a discussion with the client regarding the scenario and the theme and will proceed to create a phishing e-mail. During the creation of the e-mail there are no decision points to be taken such as which principles of influence or reasoning scheme is to be used for the e-mail. The specialist will make what they think is a believable scenario and may or may not include several cues which would point to the reader that the e-mail is phishing. This process is done by experience of the specialist. The coding scheme will enable to create these e-mails based on scientifically-backed principles which opens up a number of business and scientific opportunities. Moreover, it can enable the generation of 'data-driven' phishing e-mails instead of intuition or experience based e-mails.

2. Scientific Foundation

This section will establish the definitions and notions which are to be used for this project. Furthermore, it will investigate current research that has been conducted on phishing. Ultimately, this section will set the theoretical foundation for the project.

2.1. Social Engineering

Phishing is a subset of social engineering[12], therefore before delving in the specifics of phishing a step back needs to be taken and look at what social engineering is and what it entails.

The definitive authority on social engineering is Kevin Mitnick whose seminal book, *The Art of Deception: Controlling the Human Element in Security,* has essentially established the basic tenets of social engineering.[13] Mitnick's book, although not a scientific text, discusses what techniques and tactics a person can use in order to obtain the information they want, with the use of technology or not. Pretexting is a hugely important concept which has been introduced by Mitnick which describes the need for a social engineer to pretend to be someone they are not. Phishing essentially always uses pretexting since the sender will purport to be someone else, usually a person who would make the target follow through with their demands.

Mouton et al. have developed the first ontology of social engineering with clear definitions based on extensive review of existing taxonomies. An ontology is 'the explicit specification of a conceptualization' as stated by Gruber in his seminal work on ontology.[14] An ontological model attempts to represent knowledge in way that it maintains its declarative essence and is independent of representation. Essentially, an ontology is trying to create a conceptualization which is independent of any specific viewpoint or modelling language. Therefore, their ontological model presents a good starting point for defining social engineering for the requirements of the project. The definitions are shown below:

- *'Social Engineering:* The science of using social interaction as a means to persuade an individual or an organization to comply with a specific request from an attacker where either the social interaction, the persuasion or the request involves a computer-related entity.'
- *'Social Engineering Attack:* A Social Engineering attack employs either direct communication or indirect communication, and has a social engineer, a target, a medium, a goal, one or more compliance principles and one or more techniques.'[15]

Figure 2 shows a more complete depiction of a social engineering attack:

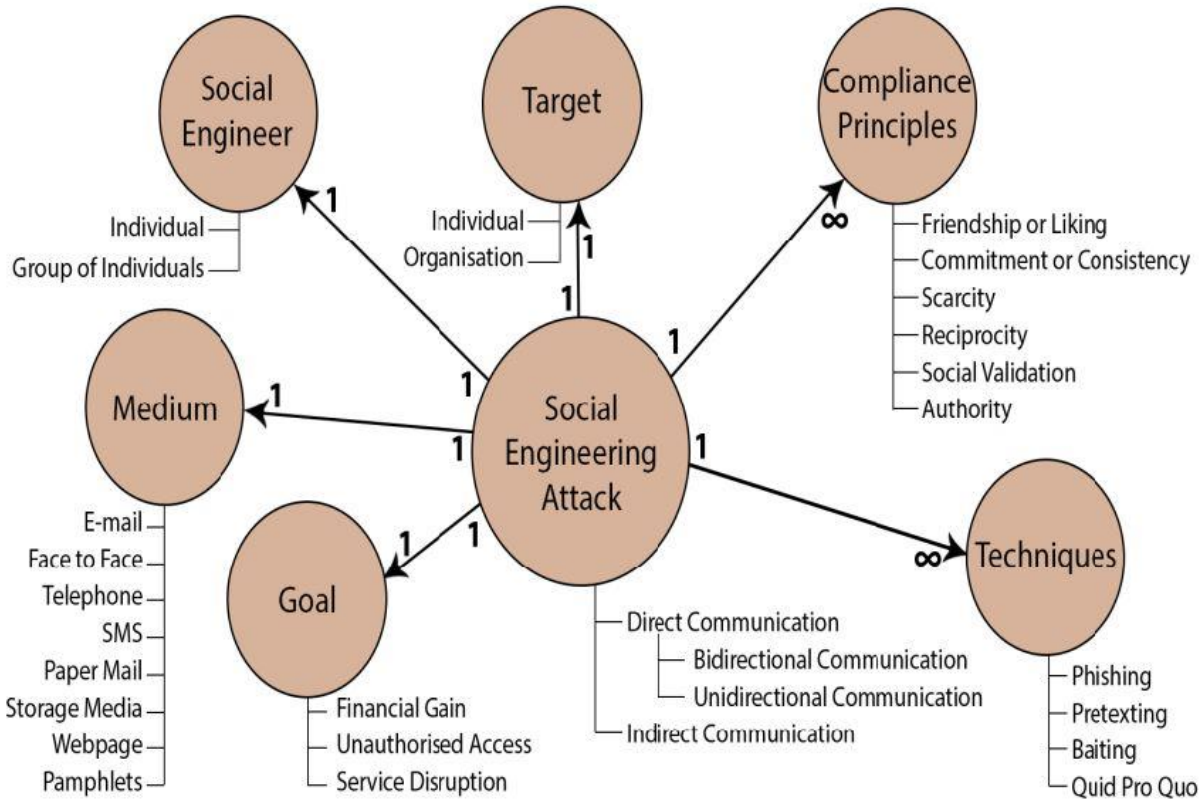


Figure 2: An ontological model for social engineering attacks by Mouton et al. [12]

Mouton's ontology describes phishing as a technique of social engineering which aims to achieve a goal. This description essentially states that phishing can be considered as a form of Social Engineering Attack. Note that Mouton states that there can only be one goal, medium, social engineer (individual or group), target but an unlimited number of compliance principles and techniques. Looking more closely, Mouton could be wrong in the sense that there is only one target which can be either an individual or an organization. This is false in the case of mass phishing campaigns where the targets are individuals who are usually not related. However, in the context of this project, Mouton's approach encompasses all the necessary elements for SPT.

Mouton's definition attempts to include social engineering in its entirety. However, for this project, we are only interested in phishing. Therefore, Mouton's ontological model is adapted to describe the phishing service conducted by industry:

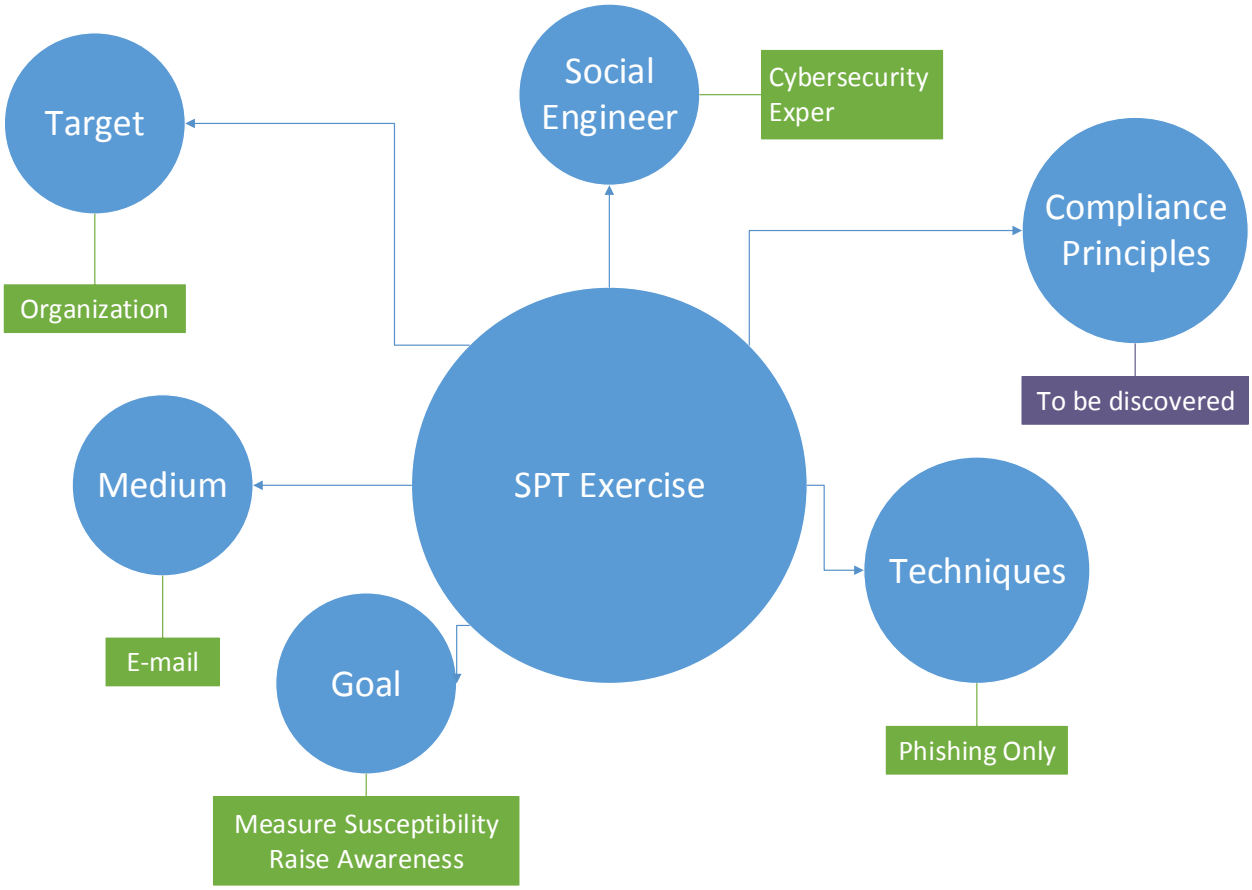


Figure 3: The adapted ontological model for industry's SPTs.

The above model is the adaption of Mouton's ontological model in relation to the SPTs offered by industry. Social Engineering Attack is replaced by SPT engagement. In SPT engagements the target is always an organization ranging from major banks to smaller retailers. The medium is always e-mail (no phone based social engineering or SMS phishing) which always leads to a spoofed website. Malicious attachments or Trojans are not used. The social engineer is industry's specialists on cyber-security and penetration testing. The goal of SPT engagements is to measure the susceptibility of employees to phishing and to enable to raising of awareness with follow-up training sessions and e-learning modules. The technique used is only Phishing. The compliance principles that are used will be explored later when the dataset is investigated.

2.2. Phishing

Now that phishing is clearly defined as a subset of social engineering in general, it is time to look more closely to phishing. The most exhaustive work that has been done so far in the quest for a consensual definition of phishing comes from Lastdrager[16]. Lastdrager has collected 2458 publications which included 113 distinct definitions on phishing. By combining all these mentions he has created a consensual definition which is inclusive as it is minimal:

‘Phishing is a scalable act of deception whereby impersonation is used to obtain information from a target’, quoting Lastdrager.

Before discussing phishing it is interesting to see how Lastdrager’s definition coincides with the conceptual model of Mouton. Phishing is a *scalable act of deception*; this scalability means that the phishing attack can be easily scaled, quoting, ‘from a single occurrence to hundreds, thousands or millions’. Mouton does not include this in his more generic ontology of social engineering where he claims that there is one target. However, Mouton talks of social engineering whereas Lastdrager for phishing. For the purposes of this project, phishing can be considered scalable in the sense that it can be repeated across many organizations with each repetition having a scalable number of individuals. We also hinge on the word deception which opens up the study of a large body of science published pertaining to deception, which will be discussed later. The next part of the definition states that ‘*impersonation is used to obtain information from a target’.* Impersonation can be considered as a form of pretexting; the attacker tries to appear to be someone who they are not. Pretexting is also included in Mouton’s definitions as one of the techniques used for a social engineering attack. The goal that is described by Lastdrager is the ‘act […] to obtain information from a target’. To obtain information from a target is an inclusive statement which includes all three of Mouton’s goals: unauthorized access, financial gain and service disruption. In the case of SPTs the goal is to measure susceptibility and to improve awareness which also falls under Lastdrager’s definition of obtaining information.

2.3. Spear Phishing and Targeted Attacks

A more practical description of phishing comes from a working paper published from FireEye, a leading cyber-security vendor, in 2012.[17] They describe regular phishing as an ‘exploratory attack’ where an e-mail is sent to thousands or millions of people without prejudice on what to expect. A success could result in a compromised machine or the theft of credit card details. The attacker does not know much about their targets (except the e-mail address of course) and does not have a specific result or desired asset in mind. The report then goes on to describe *spear-phishing* which is, quoting, ‘a more targeted version of phishing attacks that combines tactics such as victim segmentation, email personalization, sender impersonation, and other techniques to bypass email filters and trick targets into clicking a link or opening an attachment.’ The report also describes the key differences between phishing and spear-phishing, namely:

- Spear Phishing lacks the characteristics of spam. This means that whereas regular phishing e-mail is very similar to spam in terms of form and offering, spear phishing is different. For example, a spam mail selling fake drugs will be very similar to a phishing e-mail offering lottery winnings. This is especially true when it comes to spelling, grammar and other e-mail characteristics. In addition, spear-phishing uses multiple vectors of attack such as zero-day exploits, well-spoofed websites and dynamic URLs which can bypass traditional spam filters.
- Spear phishing is usually perpetrated by well-organized actors such as Advanced Persistent Threats who may have the backing of a state. Regular phishing is usually done by criminals whose goal is monetary gain. Spear phishing can be used to obtain access to a network or extract valuable and not-always-monetary data.

- Spear phishing is usually part of a multi-staged attack. It can be the first step to an organized campaign to infiltrate an organization. In contrast, regular phishing is an isolated event with a single goal.
- Most importantly, spear-phishing is tailored and targeted to the recipient. The level of personalization can vary from 'some' to 'very high', according to a report from Cisco.[18] E-mails with very high personalization will be discussed but they are not within the scope of this project.

Spear-phishing is what SPTs try to emulate. SPTs do not generate mass phishing e-mails and send them to the organization. It does not include any malicious attachments but in all cases it prompts the recipient in opening a link to a spoofed website and ask for their credentials. This project will deal solely with the e-mail and the act of clicking the link originating from e-mails with 'some' personalization, which is what SPTs entail. The distinction between 'some' and 'very high' personalization is an important one.

Some personalization means that the e-mail is tailored to a defined yet loose group of individuals. This includes a phishing e-mail sent to employees of a specific company or customers of a specific bank. On the other hand, very high personalization means that the e-mail is tailored to an individual or a small group of people. This would include conducting extensive research on a high-value individual within a company and sending them an extremely tailored phishing e-mail which will not only be impeccably written and designed, it will also have the latest technical advances and may even use a zero-day exploit. This particular type of phishing of highly personalized e-mails to high worth individuals is called 'whaling' by the cyber-security community.[19] In the case of a campaign, a successful whaling attack can have wide ranging implications for the organization. For example, if a CEO's computer is compromised then the attacker could use the CEO's eminence in the company to send e-mails to network administrators who, given that they got an e-mail from the CEO, they would gladly perform what was requested from them. However, SPTs conducted by industry do not use that level of personalization since that would heavily infringe the privacy of the targets as well as cause massive amount of stress to recipients. A malicious state backed threat does not have such limitations but a profit-making organization does. Therefore, we can say that SPTs use some personalization during the engagements, namely, they target employees of a specific company and/or department.

2.4. Trends in spear phishing

In the past few years social media have seen a huge increase in their user base. For example, Facebook passed the one billion active users mark in mid-2011 and Instagram had around 150 million users in 2013.[4] LinkedIn is a professional social networking platform on which professionals post their current and past employment, qualifications and other work related information. Social media and especially LinkedIn are considered goldmines in the eyes of malicious users who intend on performing social engineering attacks.[20] The result of this adoption of social media has put a large chunk of people's lives on the Internet with varying degrees of accessibility to third parties. This in turn makes the word of a social engineer much easier.

[4] http://www.sociallystacked.com/2014/01/the-growth-of-social-media-in-2014-40-surprising-stats-infographic/

The most current research which has explored social networks and social engineering has been done by Argani et al. [6] [21] [22] In their research two models have been proposed which try to describe how social networking sites are useful to social engineers. In their source-based model, the authors present a list of different 'sources' that a social engineer can use to achieve their goal. Figure 4 shows the source-based model by Argani et al.:

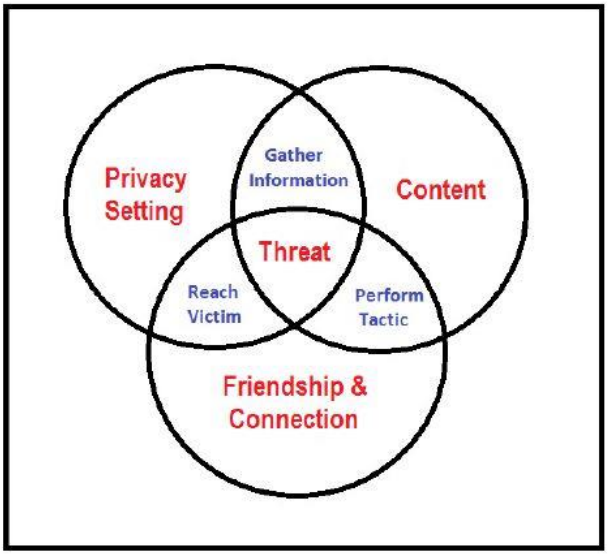


Figure 4: Source-based model by Algarni et al.

As can be seen in the above figure, there are three main sources. The content of the target's profile, the privacy setting of the target and the friendship & connection between the attacker and the victim. The combination of these three sources will help the attacker in gathering information, reaching the victim and performing the tactic. Furthermore, they have also proposed a phase-based model which denotes the steps of a social engineering. Figure 5 shows the phase-based model:

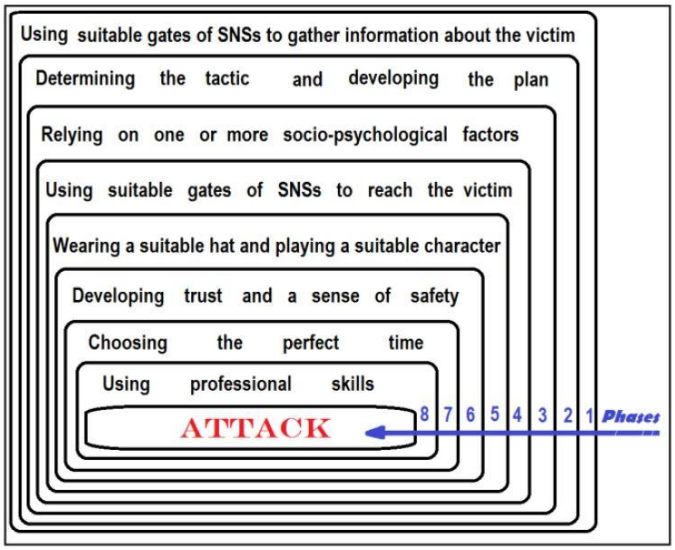


Figure 5: Phase-based model developed by Algarni et al.

Although Algarni et al. have researched the role of social networks in social engineering there does not seem to be enough research focused on the information gathering phase. The information gathering phase is vital for a number of reasons. It will help the attacker to choose the weakest possible target to execute their attack. An interesting master thesis published in 2012 by Scheelen and Wagenaar provides some interesting insights on the methods can be used in the information gathering phase. [23] They managed to create a hierarchy of an organization by cross-linking information between social networks (namely Facebook and LinkedIn). Their research is a scientifically sound method of collecting personal information from publicly available sources. The existence of such research shows that methods that were once in the realm of hackers in some dark corner of the Internet are now entering the mainstream.

2.5. Existing research on phishing: the lack of focus on the e-mail

There is a sizeable amount of research pertaining phishing. However, almost all of it revolves around the investigation of individual susceptibility to phishing. Common approaches include field studies [7], surveys[24] and experiments[25]. This section will discuss selected publications that have researched phishing in the past.

As stated earlier, Kevin Mitnick is considered the father of social engineering. In a book he published with Simon in 2013 they outlined a number of tactics used by social engineers; these include trust building, reciprocation, sympathy, guilt and intimidation. Since then, other researchers have tried to formalize these common tactics employed by social engineers. Workman built on the works of Mitnick and others and compiled a list of marketing-based communication methods which can be used by an attacker. [26]Workman's work is unique in that it has been empirically tested. Figure 2 shows the impact that the various variables of a person on their susceptibility to social engineering:

Variable	Subjective Security Behaviors	Objective Security Behaviors
Normative commitment	.13***	.11**
Continuance commitment	.17***	.20***
Affective commitment	.16***	.20***
Trust	.24***	.16**
Obedience	.10**	.11**
Reactance	.04	.06
Employee age	.13***	.10*
Employee gender	.01	.01
Employee education	.07*	.12**
Previous victimization	−.03	−.01
Model adjusted R^2	.46	.40

Figure 6: The various metrics which could predict the success of a social engineering attack by Workman (2008)

As seen from the table above, a number of personality traits correlate significantly with susceptibility to social engineering attacks. Workman's methodology, which investigated both subjective (self-reported) and objective security behaviors, has provided some useful insights. There is a very small difference between self-reported and actual behavior which leads to the conclusion that people are aware of their own susceptibility to a social engineering attack. Therefore, people who have been trained to resist phishing attacks will be themselves aware of their actual inoculation. In other words, people are aware of how vulnerable they are and if they think that they are not very vulnerable, then they most likely are not. Note that previous victimization does not seem to correlate with susceptibility to a social engineering attack, according to Workman. This contrasts the training programs offered by some consultancies who follow a system of attacking first and educating later (embedded training). [27]

Vishwanath has investigated the attention that e-mail recipients pay to the various elements of an e-mail within the context of phishing susceptibility. In addition, he has studied the impact of knowledge about scams and computer self-efficacy.[28] He concluded that recipients who were involved (i.e. they found the subject matter interesting), paid attention to the sender identity, attention to spelling and grammar and attention to urgency. Moreover, he investigated the impact of 'e-mail load' on the likelihood to respond to phishing. The e-mail load construct is an interesting one; he discovered that people who have many incoming mails daily will develop a habitual media consumption approach which results in a more automatic processing of the e-mail. This habitual approach to e-mail reading contributes significantly to the likelihood to respond to phishing.

Jakobsson is one of the more prominent researchers on phishing who has done a very influential study which essentially compared phishing to spear-phishing.[7] In their study they found that a regular phishing e-mail had an average click rate of 17% whereas a personalized phishing e-mail was over 70%. They made the distinction of phishing and spear-phishing by inspecting the subjects' social networks and generated e-mails which purported to come from their friends. This is a striking finding and represents a rigorous approach in evaluating the e-mail itself. However, they only manipulated whether the e-mail was sent from a friend or not. E-mails have numerous elements which characterize them beyond the sender/receiver familiarity.

Sheng et al. have investigated the impact of demographic factors on phishing susceptibility.[29] They concluded that women are more susceptible than men and also that education on phishing markedly decreases the likelihood of getting phished. As with all research, however, their conclusions should be taken with a grain of salt. In most experiments the subjects are university students; a population where generally there are more women than men. So concluding that a specific gender is more susceptible than another is dangerous since the properties of the population come into play. It is important to understand the impact that the population characteristics have on such experiments and this is something that we will discuss later on for our project.

The above review demonstrates that most research conducted on phishing has focused on the demographic elements of the subjects rather than the content matter of the e-mail itself. Jakobsson investigated the effect of sender familiarity but not what the e-mail's specific subject. This limited amount of research on the e-mail specifically can be partly explained by the difficulty in designing experiments which could measure the e-mail. It is necessary that as much as possible all the other

demographic factors are kept equal which is oftentimes impossible. This situation motivates this thesis; we will create a coding scheme which will enable to study the e-mail itself and measure the impact of e-mail specific effects on the overall click-rate of the phishing exercise. In section 5.2. we will also provide a method for splitting the subjects in such a way so as to minimize the impact of demographics when comparing phishing e-mails.

2.5.1. The e-mail and the individual: the bigger picture

As we have shown in the preceding section, there is a considerable amount of research which explored the susceptibility of individuals to social engineering and phishing. Our project will deal with the properties of the phishing e-mail. The nexus between the individual and the phishing e-mail lies in the study of communication. We will define communication within the work of Berlo who created the Sender-Message-Channel-Receiver (SMCR) model.[30] In Berlo's model, communication is governed by the properties of the sender, the message being sent, the channel over which the message is sent and the properties of the receiver. As we will explore in later chapters, we essentially cover all aspects of communication except the properties of the receiver; we did a brief exploration on the receiver in section 2.5. The individual's susceptibility to phishing has been widely explored as it seems that is also the more convenient thing to measure. For example, sending the same e-mail to different people who have had their personalities analyzed using tried-and-tested methods is a much safer path to scientific knowledge instead of sending different e-mails to different people and trying to discern the differences between the e-mails. In this section, we will take Berlo's model and adapt it to our own subject. This will help us frame this research in the context of wider research conducted which will, in turn, make the recommendations for future research clearer in Chapter 7.

The image below shows Berlo's model in a pictorial way:

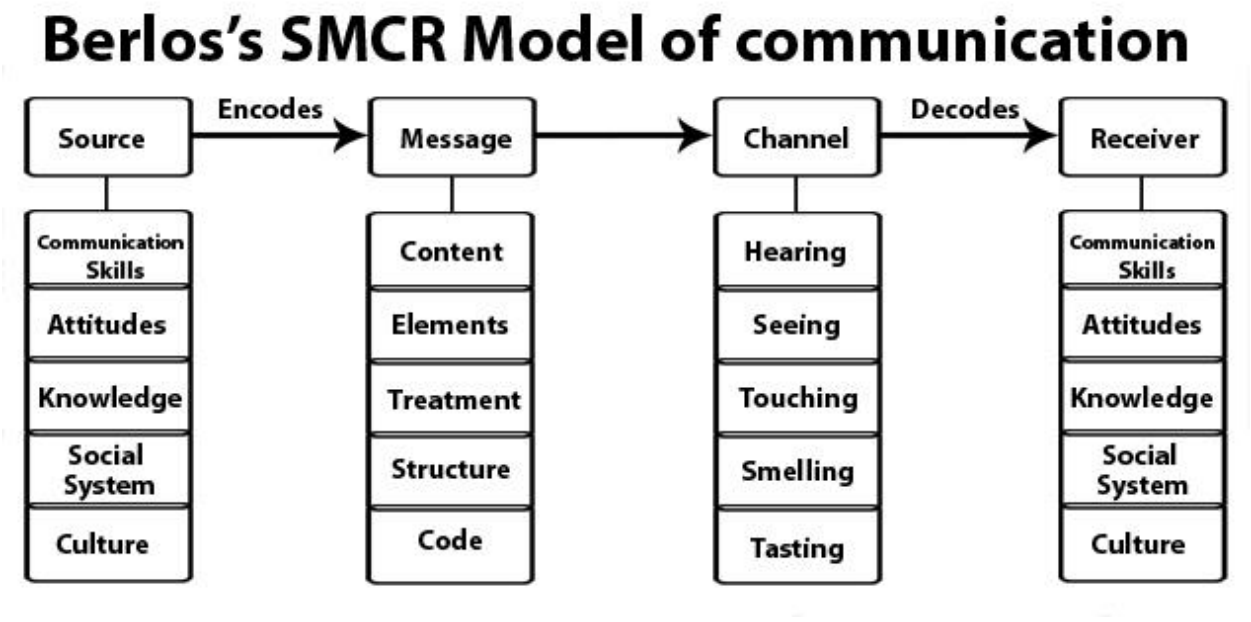


Figure 7: Berlo's Model of Communication. From http://communicationtheory.org/berlos-smcr-model-of-communication

As we can see above, there are four major pillars in Berlo's communication model. Framing it in the context of phishing we can relate each pillar to the people and processes involved in a Simulated Phishing Test. The source represents the sender of the phishing e-mail. In our case is the maker of the phishing e-mail or, more broadly, the organization who sends the phishing e-mail. Using the properties described in the original model seems a little unfitting. For example, culture and social system do not define the writing of an SPT e-mail as these e-mails are the result of a logically-driven approach and not a spontaneous expression of cultural norms or attitudes. Moreover, communication skills also do not come into play by virtue of the phishing not being a medium over which 'true' communication happens. However, in this project we discuss the *apparent* sender. Since impersonation is always employed in SPTs, the sender tries to impersonate a person or organization. So, from the viewpoint of the receiver, the sender identity and status can be emulated to achieve the desired result. We will explore this further in Section 3.2. where we look deeper into the concept of *apparent authority.* The message is the focus of this project. We will investigate the content and the form of the phishing e-mail in our quest to identify what makes a phishing e-mail obtain more clicks. The channel of SPTs is e-mail only. In Berlo's view, it can be considered as being wholly seeing. Finally we have the receiver. As we have explored, most phishing literature is focused on the properties of the receiver. Although for generalized SPTs where e-mails are sent to departments there is no prior investigation of the individuals receiving it, targeted attacks do. For example, in cases where a company would like to measure the susceptibility of computer-network administrators (the Holy Grail for a malicious user) a more personalized form of phishing occurs. The phishers would conduct an extensive investigation of the recipient and identify what would seem as likely to be more successful for them. In computer security jargon this is known as Open Source Intelligence or OSINT. For this project, we focus on spear-phishing and not targeted attacks. However, it is always a good idea to at least take into account what the overall interests of the recipients would be. For example, if a phishing test is to be done to the marketing department of a company then formulating the e-mail along some marketing-related theme should serve the purpose better instead of sending an e-mail with a seemingly irrelevant topic.

The schematic below demonstrates the adaptation of the SMCR model to Simulated Phishing Tests:

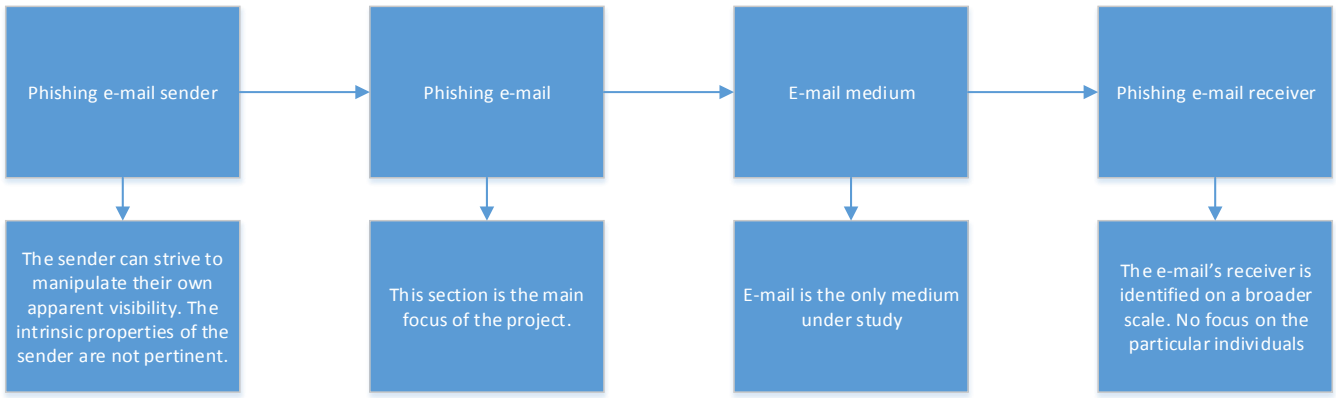


Figure 8: The adapted SMCR model for Simulated Phishing Tests

As can be seen above, a phishing engagement essentially encompasses all the elements of communication with some element being less relevant. For example, the channel is e-mail always so studying it would not make sense. What would make sense would be to try and study different forms of

social engineering such as vishing (voice phishing) or social media attacks. It is important to take into account, however, that the e-mail sender, the e-mail itself and the recipient can all affect the eventual click-rate of an e-mail. In the future, a combination of studies into these SMCR inspired is expected to provide a much clearer picture of the click-rate of an e-mail as it would take into account many more defining variables of the entire phishing sending and receiving chain. This project, will focus on the content of the message.

3. Deception and Persuasion

The previous section has established the definition of phishing. We will pick it off from there and investigate more in depth the defining elements of phishing. The definition clearly states that phishing is an act of deception. This opens up the body of study on deception as a major element of phishing. Deception, as we will discuss in section 3.1., almost always involves persuasion. From the definition of phishing we establish it as a deceptive text and from the deceptiveness we establish it as a persuasive text. In this chapter, we will investigate the deceptive and persuasive elements of phishing.

3.1. Deceptive elements of phishing

As we have explored in section 2.2 phishing is a deceptive act. Deception can be defined as "to intentionally cause another person to have or continue to have a false belief that is truly believed to be false by the person intentionally causing the false belief by bringing about evidence on the basis of which the other person has or continues to have that false belief."[31] Using Lastdrager's definition on phishing where he defines it as an act of deception, we can say that phishing falls within the definition shown in Section 2.2. But what really is deception?

Deception is the "deliberate attempt to mislead others".[32] The main words in this definition are "deliberate" and "mislead". Deliberate means that the person who conducts deception is doing so knowingly and willingly. They know the truth, how it is different from falsehood and they present the falsehood so as to achieve their goals which, according to the definition, is to mislead others. Mislead (to) is defined by Webster as the act "to cause (someone) to believe something that is not true".[5] This definition is important as it leads us to assume that deception always includes an element of persuasion as well. Since the attempt is deliberate we could say that the deceiver will try to mislead the recipient using persuasive methods. Persuasion will be further explored in section 3.2. J. K. Burgoon is one of the most prominent researchers to work on deception. She developed a number of theories such as the Interpersonal Deception Theory and her work also compares deception detection between verbal and non-verbal forms of communication[33]. She postulates that deception is feasible because people have the following salient biases:

1. *Truth bias:* People often over overestimate the truthfulness of others.
2. *Visual bias:* The tendency to rely on visual cues when detecting deception instead of verbal or linguistic.
3. *Demeanor bias:* Credibility is afforded to the speaker based on their communication style rather the context of the communication.

[5] http://www.merriam-webster.com/dictionary/mislead

4. *Expectancy violations bias:* The tendency to judge unusual behavior as deceptive.

Burgoon's work compared how well people fare in detecting deception by various means. These include verbal, vocal and visual; either isolated or combined and how they affected the salience of each bias. We are dealing with e-mail which is a wholly verbal communication medium where the receiver cannot see the speaker or listen to their voice. However, this makes the manipulation of some biases easier. Truth bias is ever present regardless of the form of communication and is an independent construct; people will have the bias in most cases. Visual bias occurs when the speaker looks as truthful based on how they look (i.e. wearing a suit). In the case of e-mail, people will be biased when the e-mail contains legitimizing elements such as company logos or formal e-mail structure. Demeanor bias discusses, in the case of e-mail, the writing style. If the e-mail is well written without spelling or grammar errors and follows a logical path of thought then, based on its *demeanor,* people will be biased in understanding it as a truthful text. The expectancy violation bias deals with the detection of what the receiver would understand as unusual. Visual bias is seen by Burgoon as the most important bias in assisting deceivers in achieving their goals. The reasoning for this is lent from the Elaboration Likelihood Model (ELM) which was developed by Petty and Ciacioppo.[34] Although the ELM was focused on deception Burgoon used it to explain her visual bias construct. In the ELM, when a receiver is confronted with a message they will either process it centrally or peripherally. Centrally means that they will carefully analyze the message and make a conclusion based mostly on the argumentation and rationality of the message. On the other hand, the peripheral route is when the receiver will process a message based on various cues that surround the message. In the case of the visual bias and e-mail an e-mail with more elaborate, legitimate-looking and correct visual cues such as logos will be processed through the peripheral route since all these elements point to it being seen as a true message and thus the reader will not afford much attention to the content of the text.

Deception has been a research interest for many decades. In the case of phishing, however, we need to distill our focus to only publications which deal specifically with deception which is conducted through e-mail or other similar means. But what do we mean with other similar means?

Communication has been an integral part of the human race. A general and encompassing definition of communication comes from Shannon and Weaver who have aptly stated that communications is "all the procedures by which one mind may affect another".[35] This definition can be expanded to almost anything. It can be spoken word, writing but even the design of a slide rule to be used by others can be considered a form of communication.[36] Discussing the definitions and philosophical meaning of communication would be an entire thesis in its own, so we need to specify the type of communication we are dealing with.

E-mail is a form of computer-mediated communication defied as "…an electronic communications system that is used to send information from one person/site to another, or from one person to many people as the same time"[37] Moreover, e-mail is an asynchronous form of communication in the sense that the email may be sent at one time and be read later[38]. In contrast, face-to-face communication or instant messaging is synchronous so as when the message is sent it is instantly received and understood by the receiver. This requires that we exclude elements which define deception in a real-time setting

and at the same time include the elements which are unique to, computer mediated asynchronous communication (A-CMC).

Keila and Skillicorn state that "deceptive writing is characterized by reduced frequency of *first-person pronouns* and *exclusive words*, and elevated frequency of *negative emotion words* and *action verbs*".[39] They have used this premise to investigate the Enron e-mail database which included a number of both deceptive and truthful e-mails. They generated a classification scheme based on the four characteristics which were italicized above. They concluded that the above metrics can be indicative of deception or organizational dysfunctions.

Grazioli investigated why some technologically and business –savvy people still fall for online deception.[40] Although their research was focused on the detection of deceptive websites what is of interest is their hypotheses on how do people who are able to detect deception do so. Their finding indicates that people who are able to detect the *presence* of deception cues are more likely to correctly guess a website as deceptive. Moreover, when the subjects were primed to the fact some specific cues were manipulated then the detection rate increased significantly. For example, in their research they presented an online retail website which had only positive reviews for a particular product. Subjects who were able to detect deceptive websites were not able to do so from the context of the manipulated cue but by the understanding that a cue was manipulated. In the product reviews example, people were able to detect deception based on the unnatural number of positive reviews rather than the content of the reviews' text. The impact of this research is that follow-up training after a phishing exercise can effect improvement on the trainees' deception detection capabilities.

Research on deception and its relationship to social engineering was also investigated. Qi claims that there are two main reasons that people can be deceived; one is that there is a truth-bias and the other is that there exist stereotypical cues to deception which cross over cultural differences. In their results they conclude that most people will initially assume a message to be truthful. Whether or not they will elaborate on it to determine depends on the context of the message. They claim that if the receiver perceives the content of the message to be deceptive then they will evaluate the presence or not of deceptive cues.[41] For example, a trespasser in a building may be confronted with a security guard who may want to know what the trespasser is doing there. The security guard will look for common deception cues such as gaze or facial expressions if the story that he is told is unsatisfactory i.e. does not make sense given the environment.

Further work conducted by Zhou et al. has tested the characteristics of A-CMC in an experimental setting.[42] They collected a large number of deception cues which pertain to A-CMC and asked participants to send either truthful or deceptive messages. A "deception cue" is defined as "a piece of manipulated information".[40] They then went on to perform statistical analysis in order to identify which cues are indicative of deception. Their significant findings were that deceptive A-CMC exhibits more quantity, complexity and diversity. The quantity construct was measured based on the overall length of the message, the number of verbs, the number of modifier words and the number of noun phrases. The complexity construct was measured based on the degree of pausality in the text as in$\frac{total\ \#\ of\ punctuation\ marks}{total\ \#\ of\ sentences}$. Diversity was measured on the lexical diversity (percentage of unique

words in the text) and content word diversity which is the ratio of words which express a lexical meaning to the total number of words.

Zhou's work is the most granulated work on deception conducted with focus on A-CMC. Thus, the variables of deception for our coding scheme will be derived by Zhou's work. For our coding scheme, we will only use their empirically supported cues to deception rather than the entire set of their hypothetical cues. We will use the cues which make the distinction between truthful and deceptive *senders.* The table below shows the selected cues to deception:

Table 3: The deception variables and their description.

Deception Cue	Description
Number of Words	More words are a cue to deception
Number of Verbs	More verbs in the text are a cue to deception
Modifier	Modifiers are words which change the meaning of another dependent word. For example, in the sentence "This is a red ball" the word "red" is a modifier. More modifiers are a cue to deception.
Noun Phrase	Noun phrases are phrases which include a noun and are one of the most prominent elements of writing. For example, "He went to the cinema" is a noun phrase since *he* and *cinema* are nouns. More noun phrases are a cue to deception
Pausality	This construct is about the number of periods, commas and semicolons present in the text.
Lexical Diversity	The ratio of unique words to total words
Content word diversity	The ratio of different content words to the total of content words. Content words are those which describe an object or an action and are not functional words of the sentence.

In summary, deception is prevalent in both society and online. Our focus here is phishing, therefore we will use the results of phishing-specific deception cues from literature in order to analyze the phishing e-mail corpus provided by industry.

3.2. Persuasive elements of phishing

Persuasion is 'the strategic behavior in presenting information designed to influence someone's beliefs, attitudes, decisions, or actions'.[43] Phishing e-mails always ask the recipient to perform an action, either clicking on a bogus link or downloading an infected file. Therefore, we can define phishing e-mails as persuasive messages. This categorization allows to apply theoretical models which deal with the effectiveness of persuasive messages and the cognitive process that recipients follow when they receive such a message.

When trying to search into persuasion, the oldest work on it comes from Aristotle. In his classical work, Rhetoric, Aristotle discussed persuasion[44]. His claims on persuasion revolve around three major axes. First, is the *virtue* or character of the speaker. If the speaker seems credible to the audience then the audience is more likely to be persuaded. Second, is the *pathos* of the audience as in the emotional state of the audience. If the speaker manages to arouse a certain emotion such as anger, then the audience is more likely to take in the speaker's arguments. Finally, is *logos* or the rationality of the argument itself. A speaker whose arguments are perceived as rational by the audience will more likely agree with the statements of the speaker. A complete analysis of Aristotle's rhetoric would be an entire thesis in itself yet the take out from this small mention is that, as we will see later on, his ideas have endured and are still relevant up to this day even within the context of such a technologically dependent issue such as phishing. As we will prove later on, Aristotle's ideas have been reformulated by more contemporary research but still Aristotle's overarching classification is still there, as we will show later on.

An obvious starting point is Cialdini's persuasion principles (or principles of influence).[45] According to his seminal work, people try to influence or persuade others using 6 basic principles:

Table 4: Cialdini's principles.

Name	Description
Reciprocity	The return of a favor. Getting something makes you giving something back.
Commitment and Consistency	If there has been a previous commitment to do something then people will likely stick to it even if they normally would not.
Social Proof	Doing things based on seeing others doing it.
Authority	A request from an authority figure will warrant an action even though it may be questionable.
Liking	People are more likely to be persuaded by people they like. People may like someone based on their looks, behavior or both.
Scarcity	The presence of perceived scarcity will create more demand. A typical example is the 'limited time offer' method used by advertisers.

Connecting Cialdini to Aristotle we can see that Cialdini has essentially granulated two of the three principles of Aristotle's description of persuasion. Authority and Liking fall within the virtue of the speaker, or in the case of phishing the sender of the e-mail. A person of authority or someone who is liked will have an elevated virtue in the eyes of the recipient, therefore they are more likely to fall for the phish. In addition, reciprocity, commitment, social proof and scarcity are all emotionally based principles; people are motivated by their emotions to fulfill the needs generated by Cialdini's principles.

Akbar has conducted research where the relationship between influence principle *use* in phishing was investigated.[46] In their conclusions, the authors have identified that phishing e-mails make extensive use of Cialdini's principles and the use of the authority and scarcity principles is especially prevalent.

Therefore, the inclusion of persuasion principles in the characteristics of phishing e-mails is a justified course of action.

Further work conducted on the persuasive elements of phishing comes from Lee & Song who investigated the information structure of phishing e-mails based on a persuasive communication perspective.[47] In their work, they collected theoretical constructs from persuasion literature, computer specific communications and phishing related literature. Their work's goal was to create a phishing *identification* model based on their variables thus what is of interest is the actual variables they discovered to pertain to phishing. They are shown in the table below:

Table 5: Phishing e-mail variables from Lee & Song.[47]

Variable Category	Variable Name	Description
Credibility	Expertise, Trustworthiness	Whether the e-mail was sent from someone the receiver considers as an expert or trustworthy
Rational Appeals	Rationality of argument,	Whether the receiver understands the argument of the e-mail as rational
	Plausibility of evidence	Whether the provided evidence is plausible in the eyes of the receiver
Emotional Appeals	Degree of fear	To what extend feelings of fear are induced in the reader
Phishing-specific variables	Email address discrepancy	Different sender and reply address
	Quick response	A quick response is required
	Collecting Info	The e-mail or the website ask for personal information
	Link text discrepancy	The included links' actual address is different than the one shown
	Confusing destination address	Use of the @ symbol to confuse the reader

The above table contains elements of phishing which come from persuasion literature as well as some phishing specific variables. The first three categories, credibility, rational appeals and emotional appeals are directly connected to Aristotle's persuasion axes, virtue, logos and pathos respectively. The phishing-specific variables will be included in the final coding scheme and they will be evaluated against the dataset. They do not fall under any specific category but they represent a set of literature identified variables which will be used in the coding scheme. We reached Lee & Song when investigating

persuasion in phishing and these variables were included in their publication. Thus, we can include them in our coding scheme.

A more tangible exploration of persuasive characteristics of e-mails come from Kim & Kim who conducted a categorical and semantic network analysis on a large, publicly available database of phishing e-mails. A categorical analysis means that they had trained coders manually review the texts and rate them whereas a semantic analysis is one conducted by software which is programmed to detect recurring patterns based on keywords, sentence structure and other elements. [48] This particular work is interesting as it represents the most granulated approach to the characteristics of phishing e-mails and clearly explains the way that these characteristics were found. They made three overarching categories, 'Source Credibility', 'Argument Quality' and 'Message Structure'. The table below shows their findings:

Table 6: Kim & Kim's persuasive elements of phishing

Grouping	Variable Name	Description
Source Credibility	Legitimate e-mail sender prefix	The sender uses a proper e-mail prefix such as RE:, FW: or brackets
	Legitimate domain name	The e-mail purports to be sent from a legitimate domain(like a well-known bank)
	Clearly Specified Sender	The sender's name or identity is clear in the e-mail
	Company logo	The e-mail uses a legitimate company logo
	Contact telephone number	A contact phone number is included
	Contact e-mail address	A reply e-mail address is included
Argument Quality	Rational Appeals	The reasoning of the e-mail in three types: cause, sign and analogy
	Emotional Appeals	Induction of feelings of fear, affection or happiness
	Motivational Appeals	Appealing to motivations such as safety, self-esteem and belongingness needs
	Time Pressure	Whether action is asked to be done within a time frame; ranging from Immediately to Over 72 hours in 24 hour increments

Message Structure	Conclusion	The e-mail may have an explicit or implicit conclusion
	Repetition	There may be repetition within the e-mail
	Order	The e-mail's main message can be in the beginning, middle and end of the text.

Comparing the above tables we can see a number of recurring themes between the different ways developed to evaluate phishing e-mails. For example, Kim & Kim, Lee & Song, Cialdini and Aristotle speak of the credibility of the source (with Cialdini calling it Authority and Aristotle calling it virtue). The same recurrence can be seen in the rational and emotional appeals that all four authors mention. The only author not to deal with the rationality of the argument is Cialdini who focused his work on influence techniques independent of the context of the message.

In section 3.3 we will bring all the deception and persuasion variables together and generate a phishing e-mail codebook which will be used to both measure and help in the design of phishing e-mails. We will also look at the recurring themes more closely and unify them into something more practical for the thesis.

3.3. Extraction of phishing characteristics from literature

In this section we will combine the variables that were identified earlier ensuring that in cases where the authors were talking about the same characteristics are one. Beginning with the deceptive cues the following table is produced:

Table 7: Operationalization of Deception variables.

Variable Name(abbreviation)	Measurement scale	Operationalization
Number of Words(WRD)	Absolute number of words	These variables can all be automatically measured using text mining software (as in the work they are derived from).
Number of Verbs(VRB)	Absolute number of verbs	
Modifier(MOD)	Absolute number of modifiers	
Noun Phrase(NPHR)	Absolute number of noun phrases	
Pausality(PAUS)	Ratio of punctuation marks over number of sentences	
Lexical Diversity(LDIV)	Ratio of unique words over total words	
Content Word Diversity(CWD)	Ratio of unique content words over total content words	

The above table shows the deception variables that will be used for the coding scheme. This set of variables can be automatically measured. The author recommends that the programming language Python due to the fact that it already has extensive libraries which were designed for natural language

processing. It is also one of the easier languages to program in in general thus implementing the abovementioned variables is a simple task.

Moving onto persuasion, we will now operationalize the persuasive variables of phishing e-mails. As stated in the beginning of this section, we need to ensure that there is no redundancy of variables in cases where the authors were essentially discussing the same phenomenon. Aristotle's three axes will serve as a starting point and we will begin with virtue. The table below shows how virtue is expressed by the more contemporary authors:

Table 8: Virtue and its breakdown

Author Name	Construct
Cialdini	Authority
Lee & Song	Credibility
Kim & Kim	Source Credibility

Aristotle's notion of virtue is perhaps the most general one and would not be prudent to try and measure it in some way. Moving down the table, Cialdini's concept of authority is also quite general and it would be difficult for a layperson to identify its presence. However the last two authors have granulated virtue/authority into more tangible and phishing-specific elements. Lee & Song talk about the *expertise* and *trustworthiness* of the sender; more specifically the apparent expertise and/or trustworthiness of the sender. Kim & Kim took it one step further where they talked about source credibility being expressed in a number of the e-mail's elements. These are explained in Table 4. It is important to note that Kim & Kim describe very tangible elements of the phishing e-mail whereas the previous authors do not. Thus it is necessary that we make the distinction between those easily identifiable variables and those who involve the understanding of the reader. Lee & Song's description of credibility uses the words expertise and trustworthiness. This is a more human-oriented understanding as these are relative terms. For example, if a manager sends an e-mail then, in the eyes of interns, he will be much more trustworthy and expert-looking to interns whereas in the eyes of senior managers he would not be as much. This construct is highly dependent on the context of the e-mail but we can still distill some variables which can measure it. We could use the inclusion of titles as a way to convey authority. Especially in countries like the Netherlands the use of titles such as Dr., ir. or Prof. are legally protected and can only be used by people who have earned them from accredited institutions. If an e-mail includes such honorifics then it can be said that it attempts to express authority. Authoritative language can also be detected in the language of the text. If the text is clearly making use of the authority of the sender then it can be said to project credibility. Note that authority does not solely stem from a particular individual. It can be a department or even a higher authority within an organization. For example, asking employees to accept new ethical guidelines as expected by the Governance Board then the appeal to the governance board means that the text is making an appeal to authority. Thus the following table is compiled:

Table 9: Operationalization of virtue

Variable Name	Operationalization
Uses Honorifics	"Does the e-mail contain titles which denote the sender's position or expertise?"
Appeals to Authority	"Does the e-mail make any appeals to authority?"
Legitimate e-mail sender prefix	"Does the e-mail use e-mail prefixes characteristic of e-mails?"
Legitimate domain name	"Does the e-mail come from a legitimate domain name?"
Clearly Specified Sender	"Does the e-mail come from a clearly specified sender such as a person or a group?"
Company logo	"Does the e-mail use the logo of the company?"
Contact telephone number	"Does the e-mail include a telephone number?"
Contact e-mail address	"Does the e-mail contain a contact e-mail address? I.e. a reply address?"

The above operationalizations are essentially questions which are to be answered by the coder or designer of the e-mail. The coder will have to identify whether the variables are present or not and the designer whether to include them in the e-mail. All the variables above are dichotomous; they can be either present or not present. In a computer-understandable way they would be 1 or 0 with 1 being true (present) and 0 false (not present).

The second Aristotelian group is the logos, or the rationality of the argument. This grouping deals with the actual rationality of the text and not the emotions that may be expressed by reading it. Cialdini's principles do not encompass rationality as his theories are mostly related to influence. Therefore, we will only use Lee & Song's and Kim & Kim's variables. The table below shows the breakdown of logos:

Table 10: Logos and its breakdown

Author Name	**Construct**
Lee & Song	Rational Appeals
Kim & Kim	Rational Appeals

Both authors use the term "Rational Appeals" in their text. In Lee & Song, rational appeals are broken down into two concepts: rationality of the argument and plausibility of evidence. This is a direct representation of formal logic where rationality of the argument means that the text follows the rules of formal logic (identity, non-contradiction and excluded middle). The plausibility of evidence is whether the premises that are presented to the reader are likely to be understood as true. A piece of text may violate one or both of these concepts. For example, the premises "the moon is made of cheese" and "cheese is edible" can be combined to say that "the moon is edible". This argument has correct rationality in that it deductively concludes that the moon is edible. The plausibility of evidence, however, is clearly violated since the most people would not agree that the moon is made out of cheese (or to some cheese in inedible). In a reverse example, the premises "the sky is blue" and "stars are bright" are plausible in the eyes of most but the statement "the sky is bright blue" would violate logic since the

premises do not mention that the stars are in the sky. For the statement "the sky is bright blue" to be true more premises would have to be added. In reality, it is difficult for most people to write *irrationally* and most phishing e-mails follow a logical argumentation and use believable premises. However, this codebook will enable future measurement of rationality and its impact on phishing. In Kim & Kim, they broke down the rationality of the argument as reasoning by cause, reasoning by sign and reasoning by analogy. In their work they state that over 70% of e-mails use argument by cause which is the most deductive one as, quoting, "gives direct evidence of causality between events". Argument by sign means that indirect evidence that is given for a different event is a sign of another event. Finally, reasoning by analogy uses familiar events to help readers understand a situation. It seems that the breakdown of reasoning done by Kim & Kim is a granulation of the Lee & Song's "rationality of the argument" construct. Thus, the following table is produced for the Logos variables:

Table 11: Operationalization of Logos

Variable Name	Operationalization
Evidence Plausibility	"Is the evidence provided in the text plausible?"
Reasoning by Cause	"Does the text use reasoning by cause?"
Reasoning by Sign	"Does the text use reasoning by sign?"
Reasoning by Analogy	"Does the text use reasoning by analogy?"

Again the answers to these questions is a yes or no. In addition, the three types of reasoning are mutually exclusive as there can only be one type of reasoning in the text or at least a most prominent.

It seems that it may be very difficult to teach coders and designers to understand these concepts without extensive training on logic. Thus, it is necessary to make the questions as understandable and answerable as possible. The importance of training and coders/designers will be further explored in section 5.2.

The most populous part of persuasion in this project is emotion or pathos. Before we lay out the variables a small discussion on the relationship between motivation and emotion is necessary. A number of researchers have identified that emotion and motivation are intertwined.[49], [50]. In the most basic sense, emotions arise when there is a need to satisfy a particular need such as hunger, thirst or sleep. This represents the most basic needs which trigger an emotion which in turn motivates the person to act in a specific way to satisfy their needs. However, in Maslow's seminal work on motivation he moved beyond the basic (or hypothalamic) needs like hunger or thirst and expanded them into safety needs, belongingness needs, needs for esteem and needs for self-actualization.[51] We can say that emotions also trigger these needs or motivations. Thus, we can include motivational appeals in the wider sense of emotional argumentation a.k.a. pathos. The table below is the breakdown of pathos:

Table 12: Pathos and its breakdown

Author Name	**Construct**
Cialdini	Reciprocity Commitment and Consistency Social Proof Liking Scarcity
Lee & Song	Emotional Appeals
Kim & Kim	Emotional Appeals Motivational Appeals

Cialdini's principles are operationalized as they are. However, as stated earlier and be explored later, the coders will require training that will enable them to be able to identify these elements within the e-mail. Lee & Song used only fear as their emotional appeals. Remember that most research stems from analysis of phishing e-mails *in the wild* which are in many ways different from what SPTs generate, as explained in Section 2.3. Thus, Lee & Song probably only encountered fear as the emotion present in their own corpus. Kim & Kim happen to have the more complete set of the various emotions that can be exhibited in a phishing e-mail. The table below shows the operationalization of Pathos:

Table 13: Operationalization of Pathos

Variable Name	**Operationalization**
Reciprocity	"Does the text use the Reciprocity principle?"
Commitment and Consistency	"Does the text appeal to the commitment of the reader?"
Social Proof	"Does the text make use of the Social Proof principle?"
Liking	"Does the text make use of the Liking principle?"
Scarcity	"Does the text make use of the Scarcity principle?"
Fear Appeal	"Does the text induce emotions of fear?"
Affection Appeal	"Does the text make use of affective events?"
Happiness Appeal	"Does the text induce feelings of happiness?"
Safety Needs	"Does the text appeal to the safety needs of the reader?"
Self-Esteem Needs	"Does the text appeal to the self-esteem needs of the reader?"
Belongingness Needs	"Does the text appeal to the belongingness needs of the reader?"

In the able above the Cialdini and Maslow variables have already been explained. The three direct emotions which are derived from the literature are fear, affection and happiness. Fear is quite obvious what it stands for. The same is for happiness. In both cases there is a clear mention that, in the case of fear, something bad will happen or has happened and, in the case of happiness, it references a happy event or induces feelings of happiness in the reader. Affection appeal is a trickier notion. In the original

paper they use as an example the provocation of feelings of pity. For example, there could be a mention of an earthquake and donations are needed. It is not necessarily fear nor happiness but rather an appeal which would make people feel empathy towards others.

The above table is the operationalization of the emotive elements of phishing e-mails. The above tables are to be included in the final coding scheme which is to be developed in Section 5.

4. Dataset Inspection

Now that phishing e-mails have been defined according to their deceptive and persuasive elements it is time to take a closer look at the dataset which contains the previous phishing attempts. A phishing exercise includes the e-mail and a link to a spoofed website where personal details are asked. The resulting database includes also whether recipients interacted with the website and to what degree i.e. if they gave their name only or if they gave their password as well. Investigation of the websites is beyond the scope of this project; we will only look at the e-mail and the percentage of recipients who clicked on the included link. The names of the clients have been anonymized and some parts of the e-mails may be blurred for confidentiality reasons.

4.1. Ordering of the current dataset

As a first step, a table is produced where the e-mails are sorted according to their click rate in an ascending order along with the number of recipients per e-mail. The current dataset includes a total of 27 e-mails. The table is shown below:

Table 14: The current dataset with recipients and click percentage.

#	Name	Recipients(number)	Click (%)
1	GARD2	18	61,11
2	GARD1	17	52,94
3	MAR	64	50
4	PERO	20	50
5	NORDIC1	143	46,15
6	FOREMAN1	11	45,45
7	BANKA3	111	43,24
8	BANKA1	234	42,74
9	NORDIC2	324	41,98
10	BANKA2	238	39,08
11	THUISBANKA	2419	39
12	CONSNL1	5196	30,06
13	TUBE1	25	28
14	SSL1	598	24,25
15	BANKA4	110	23,64
16	CONSBEL1	3262	21,64
17	COFFEEBANKA	2806	21

18	CLOUDBANKA	2419	21
19	CONSBEL2	5197	20,72
20	STARTING1	157	12,74
21	CINEMABANKA	2806	12
22	BANKB2ROM	18	11,11
23	ITCOMP1	27	11,11
24	FRAUDBANKB	28	10,71
25	EMAILBANKB	137	10,34
26	SASBANKB	108	7,35
27	BANKB1ROM	17	5,88

As can be seen in the table above, the dataset includes values ranging from 5.88% up to 61.11%. Moreover, there is a large variability in the number of recipients per e-mail from over five thousand to as little as eleven people. Some data-points are related in the sense that they are part of the same campaign with the e-mails sent to the same people. For example cases 1 and 2 were sent to the same organization with a one month difference. The cases represent also e-mails sent to different countries. At a first glance, the data seems heterogeneous which will be further discussed in section 4.6.

A first test that is done with datasets with a continuous interval response (the Click percentage in our case) is to check whether the responses are normally distributed (or identify what distribution is present, if any). Using R, the following Q-Q plot was generated:

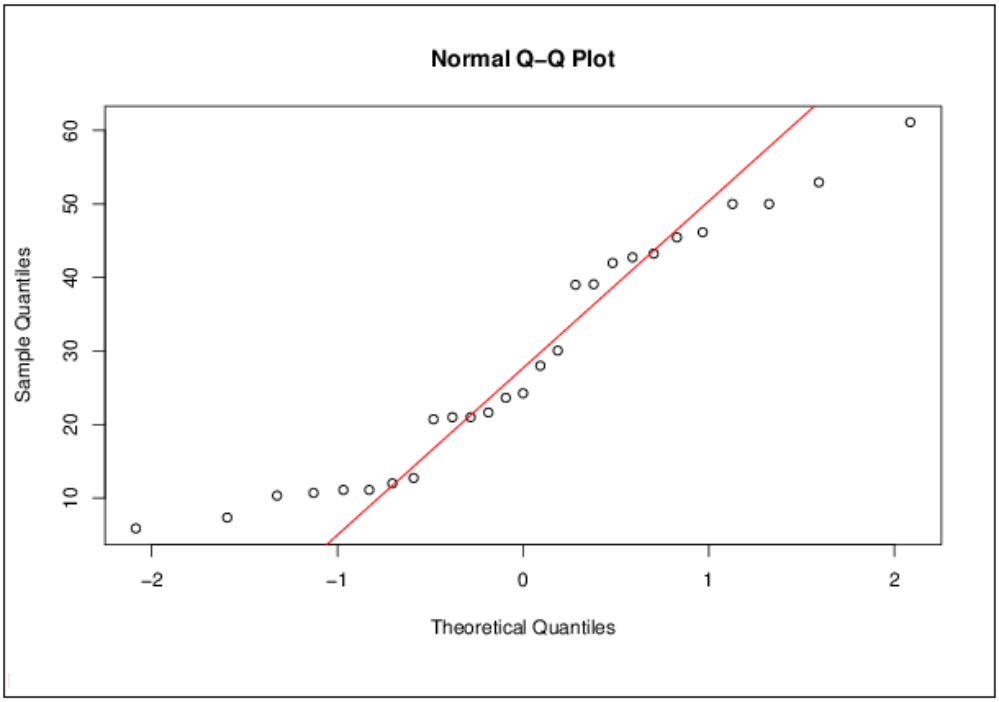


Figure 9: The Q-Q plot of the click percentages of the current dataset.

The Q-Q plot shows three distinct groups. There are eight cases in the lower level, eight in the middle level and eleven in the highest quantile grouping. Also, there are two extreme outliers which go over the 2 standard deviations which are the two cases where there is a %5,88 and a %60,11 click rate. The author performed the Lilliefors test which falsified the hypothesis that the data comes from a normal distribution, yet barely so with p = 0.0532 which is above the 0.05 threshold. From this dataset there can be no assumption of normality which excludes the more classical regression approaches if prediction is the goal. We will explore each group to see what may be responsible for this distribution.

The "low" group comprises of the eight cases which have the lowest click rate. Of these e-mails five of them were sent to the same bank, three in the Netherlands and two in Romania. The other three were another major Dutch bank, a Belgian mortgage company and an IT company. At first glance, it seems that "Bank B" is very resilient to spear-phishing attacks given that *all* their engagements belong in the low group. This could be due to the fact that this particular bank has implored to its employees the dangers of phishing. The IT Company also had a low response rate; this could be due to the fact that most employees are familiar with computer systems and have a baseline understanding of spear-phishing. The Belgian mortgage company's e-mail was the first sent to them as a proof-of-concept engagement therefore we cannot talk about the organization much but the e-mail in that case was unique in the sense that it was purposely riddled with spelling and grammar mistakes; the intrinsic properties of the e-mails will be explored in the next section. The last e-mail of the "low" group was to a major Dutch bank named "Bank A" which has numerous engagements. Therefore, the e-mail itself is probably the reason for the low click rate, given that Bank A yields considerably higher click rates in general.

The "middle" consists of the click rates ranging from 20.72% to 30.06% with a total of nine cases. Note that there is a jump of click-rate from the low group to the middle group with no samples in the range between 12.74% and 20.72%. The middle group has three cases from Bank A, two cases from a Belgian consultancy and one case from a Dutch Oil & Gas Company, one from a small Dutch bank and one from a Dutch consultancy. It is interesting to look at the two cases to Bank A, "COFFEEBANKA" and "CLOUDBANKA" have the exact same click-rate. These e-mails were also part of the same campaign to the same bank but to different departments along with "CINEMABANKA" and "THUISBANKA". Therefore, it would be interesting to investigate the differences between those four e-mails as they were sent to the same organization.

The "high" group is the most populous with eleven cases having a range from 39% to 61.11%. Bank A is represented with four cases along with a Nordic bank with two cases, a Dutch construction company, a gardening company with two cases, a Nordic oil company and a Dutch online retailer. It is interesting to see that the two highest cases occurred with the gardening company. It seems that the gardening company's employees are the most susceptible to spear-phishing; perhaps due to the fact that their line of work is totally independent from the concepts of IT. For example, a banker will probably discuss issues of security, compliance and IT assurance much more often than an employee of a gardening company. Therefore, the line of business could have an impact on the susceptibility of the employees to social engineering attacks.

We could look at the data by looking at the entire recipient list and grouping the number of recipients in groups of click-rate percentage. The result of such a histogram is widely different from the grouping according to engagement. The graph is shown below:

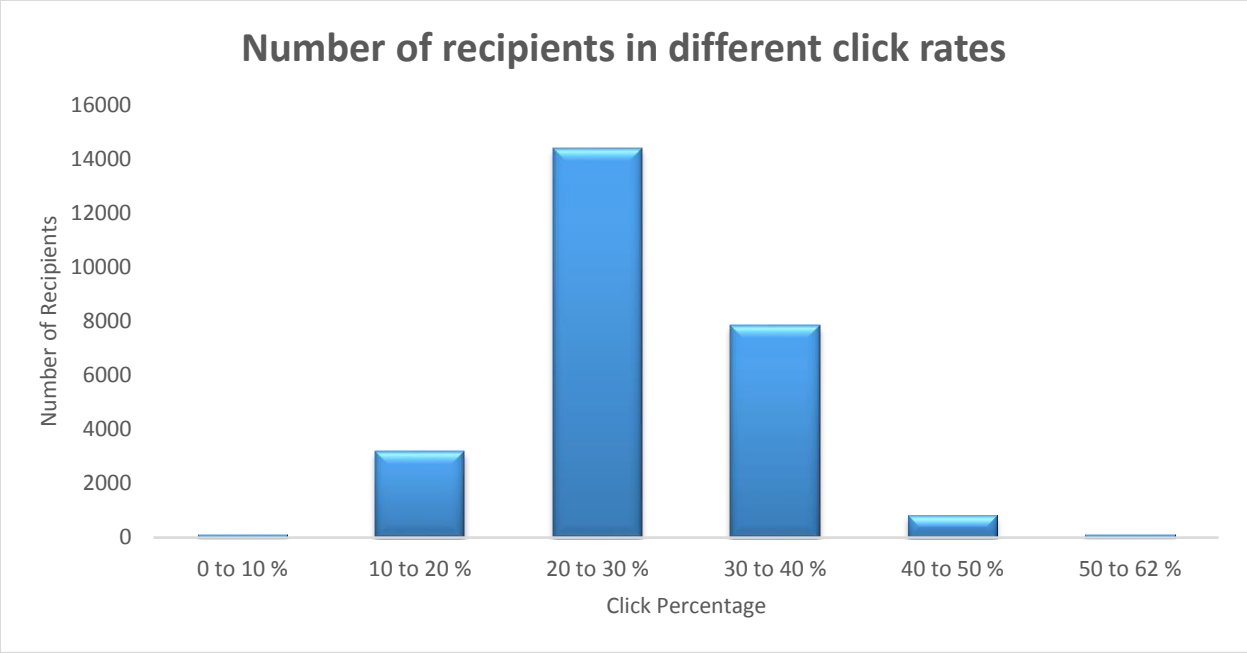


Figure 10: The click rate groupings over the total number of recipients.

The above graph shows a widely different picture than graphs 7 and 8. This is different from the "three-group" distribution we saw earlier with the Q-Q plot. We can see that if we take the total number of recipients and group those over groups of click-rates we can see that the distribution resembles a normal distribution with some skewing towards the higher click rates. If we combine the insights from the above 2 graphs we can see some interesting facts. For one, we see that the engagements which had fewer number of recipients had results which tend to skew towards the extremes. This could be explained by several reasons. First, it could be just a matter of chance that due to the small sample size we end up with very high or very low results. Moreover, we see that smaller sample sizes lead to more extreme values; this could be due to culture effects. In smaller settings, news travels fast and therefore there could have been a very rapid dissemination that a phishing e-mail was abound. This could explain the very low click rate and the high ones. Someone may have been tipped off that there was a SPT happening and not a real phishing attack and they may have all decided to give bogus response. These are things that we do not know and we cannot control. Thus, although value to the client is provided, the results cannot be used to scientific conclusions.

In summary, by examining the entire dataset we can deduce that the data is very heterogeneous. The e-mails were designed in an ad-hoc way according to the clients' specifications. During the preparation phase there was no standardized procedure to generate the phishing e-mails. In addition, the e-mails were sent to very different organizations which probably have demographically very different employees. Thus, the click rate is not dependent on the e-mail itself but also on the organization and the environmental conditions within which the test occurred. For example, in smaller organizations, due to the small number of employees, information travels faster. If one person knows that a phishing e-mail is coming in a small organization then if they share the information the impact on the percentage of those "in the know" increases more significantly than in a smaller organization. Small group effects significantly impact the click rate of the e-mail. These factors preclude the use of any regressive approach to generate a predictive model with the current dataset. However, we can still qualitatively

evaluate the e-mails and extract the relevant characteristics from them without any expectation on how these characteristics impact the click rate. We will now move onto exploring representative samples from the dataset study them more in depth.

4.2. Exploration of representative examples

In this section we will explore some characteristic e-mails from the dataset. Due to confidentiality reasons, the actual bank names, livery and identifying information has been removed or changed.

We will begin with the e-mail sent to "BANKB1ROM" which is the Romanian branch of a major Dutch bank. The original text was written in Romanian and was translated to English from a native Romanian speaker with a good command of the English language. The e-mail is shown below (the bank's name was changed to AAB Bank):

Dear colleagues,

Professional-benefits.com is an organization that gives the possibility to take advantage of discounts and attractive prices for all your purchases. AAB Bank Romania initiated a collaboration process with Professional-benefits.com and it is currently conducting a pilot with a restricted number of employees. You have been selected for this test.

The first discount offered by Professional-benefits.com consist on traveling vouchers. As an employee of AAB Bank Romania you can benefit with discounts between 25% and 50% for the tourism offers of several affiliated travel agencies.

If you want to benefit from this offer, we ask you to act instantly. You can subscribe here. You will receive an answer with more information and specific offers. (If the link does not work, copy the following URL into your browser: {LINK})

We hope that many of you will benefit from this offer. Cooperation with profesional-benefits.com allows us to improve the benefit program offered to the AAB Bank Romania´s employees.
Your participation in this sample is extremely important to us.

[BANK'S LINE, LOGO AND NAME]

Kind regards

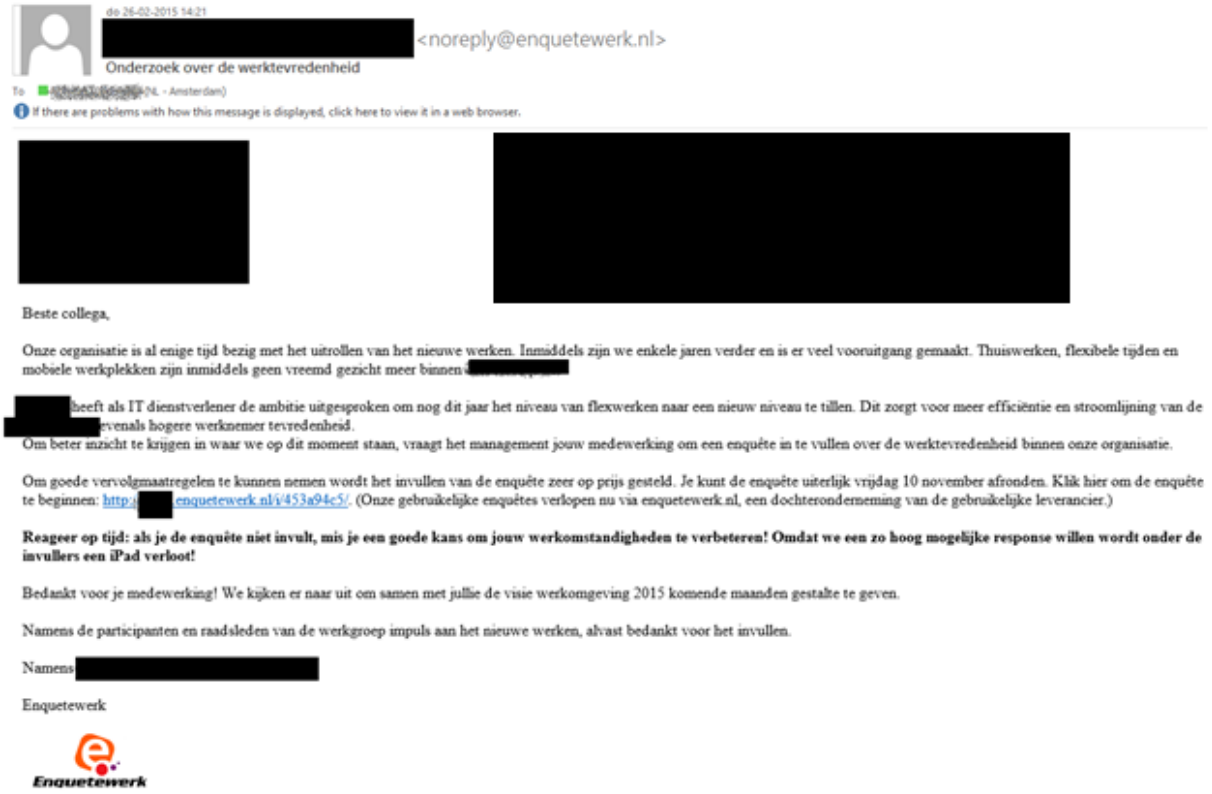
Figure 11: A SPT conducted to a banking institution

The above e-mail is a very typical scenario used in our SPT dataset, namely the "professional benefits" scenario. As can be seen, it involves the cooperation of the target company with an organization which gives discounts to employees. It asks employees to click on the link and provide their credentials in order to benefit from the offer. It is important to note that in this type of scenario the employees are not told *they have to* click the link as part of their duties. This offer is available to those who would like to thus, this can party explain the reason why for this case there was such a low click-rate. However, the same scenario was used in other engagements with much higher click-rates. The other e-mail sent to the same

bank in Romania also had a low click-rate as were the e-mails from this bank in general had low click rates.

This bank's resilience is further supported from other tests performed on them, namely the "FRAUDBANKB", "EMAILBANKB" and "SASBANKB" which were part of the same campaign to the bank's Dutch offices. The Fraud e-mail was asking employees to login with their company credentials to gain access to a new article which described a purported new way that fraudsters use to defraud bank customers. The Email scenario was asking employees if they wanted to expand their e-mail's storage if they could prove why they needed it. The SAS e-mail was promoting a new beta version of a software kit which is in use by the bank. In all three cases the request from the e-mail was in an "if you like" fashion. There were no repercussions in case the recipient failed to respond. Therefore, apart from the bank's resilience, the scenario was not imploring enough to warrant many responses. In all three cases the e-mails were purportedly sent internally from the bank to its own employees. The sender was a non-existent person but the bank's standard e-mail structure and disclaimers were used. Due to the small number of samples, however, it is difficult to pinpoint which of these factors were responsible for the low click-rate. The "ITCOMP1" scenario was a questionnaire regarding the adoption of flexible working hours by the company. It was purportedly sent by an external council which oversees the adoption of these hours and not from the company itself. A tablet was also offered as an incentive due to the need to "ensure we get the most responses possible". This scenario is a recurring theme in the database yet in this case the click rate was only 11.11%. This can be attributed to the nature of the company which is an IT provider. It is highly probable that the employees at this company are well exposed to the latest trends of IT and IT security. Thus, they were probably much more suspicious to e-mails which include links, offer iPads and ask for credentials. The last two e-mails of the "low" group were to a different major bank (Bank A) and a pilot test to a Belgian mortgage company. The "CINEMABANKA" scenario asked employees to answer a questionnaire about their movie going habits and they would be rewarded with free tickets which could be used in any Dutch cinema. The offer was provided by a fictitious company called MaxCinema. It is very likely that mindful employees searched online for this particular company and discovered it does not exist, thus explaining the low click-rate. It is interesting to see that this e-mail was part of a campaign to this bank with the other two e-mails having almost double the click rate and the fourth having more than three times the click rate. Given that these e-mails were sent to the same organization this variability between the e-mails could indeed be explained by the differences found between them. The "STARTING1" addressed to the mortgage company was intentionally riddled with spelling and grammar errors. It was a work from home scenario where interested employees could participate in a questionnaire and afterwards some would be selected to work from home and provided with a laptop and smartphone. Although the scenario was a typical one, it seems that the large number of spelling mistakes were a giveaway on the nature of the e-mail as phishing.

Looking at a different example from the low group further supports the notion that the nature of the organization is highly influential on the click rate of the e-mail. An e-mail was sent to the IT department of a Dutch part of government, had a click-rate of 11.11% and looked like this:



do 26-02-2015 14:21
[redacted] <noreply@enquetewerk.nl>
Onderzoek over de werktevredenheid
To [redacted] (NL - Amsterdam)
If there are problems with how this message is displayed, click here to view it in a web browser.

[redacted]

Beste collega,

Onze organisatie is al enige tijd bezig met het uitrollen van het nieuwe werken. Inmiddels zijn we enkele jaren verder en is er veel vooruitgang gemaakt. Thuiswerken, flexibele tijden en mobiele werkplekken zijn inmiddels geen vreemd gezicht meer binnen [redacted]

[redacted] heeft als IT dienstverlener de ambitie uitgesproken om nog dit jaar het niveau van flexwerken naar een nieuw niveau te tillen. Dit zorgt voor meer efficiëntie en stroomlijning van de [redacted] evenals hogere werknemer tevredenhed.
Om beter inzicht te krijgen in waar we op dit moment staan, vraagt het management jouw medewerking om een enquête in te vullen over de werktevredenheid binnen onze organisatie.

Om goede vervolgmaatregelen te kunnen nemen wordt het invullen van de enquête zeer op prijs gesteld. Je kunt de enquête uiterlijk vrijdag 10 november afronden. Klik hier om de enquête te beginnen: http://[redacted]enquetewerk.nl/i/453a94c5/. (Onze gebruikelijke enquêtes verlopen nu via enquetewerk.nl, een dochteronderneming van de gebruikelijke leverancier.)

Reageer op tijd: als je de enquête niet invult, mis je een goede kans om jouw werkomstandigheden te verbeteren! Omdat we een zo hoog mogelijke response willen wordt onder de invullers een iPad verloot!

Bedankt voor je medewerking! We kijken er naar uit om samen met jullie de visie werkomgeving 2015 komende maanden gestalte te geven.

Namens de participanten en raadsleden van de werkgroep impuls aan het nieuwe werken, alvast bedankt voor het invullen.

Namens [redacted]

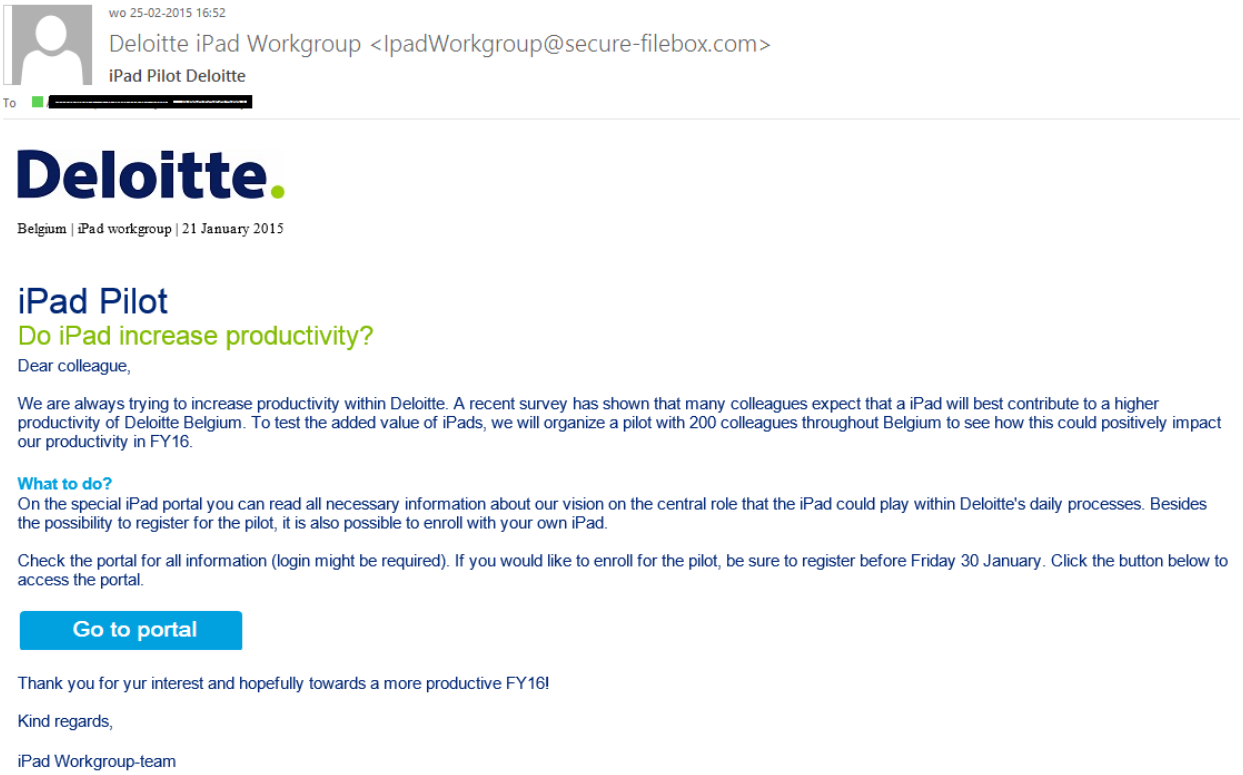
Enquetewerk

Enquetewerk

Figure 12: An SPT sent to a public body.

As can be seen above, this e-mail was very well produced. There are no spelling or grammar errors and standard e-mail writing etiquette is followed. The domain used is the enquetewerk.nl which is usually used in scenarios where employees are asked to answer some questions and then provide their credentials so they can be considered for the reward, if any.

The “middle” group consists of eight cases which range from 20.72% to 30.06%. Bank A has three cases here. The sample with 20.72% was sent to Deloitte Belgium as part of an internal phishing exercise. Since it is an internal e-mail, it can be shown is its entirety. The actual e-mail is shown below:



wo 25-02-2015 16:52

Deloitte iPad Workgroup <IpadWorkgroup@secure-filebox.com>

iPad Pilot Deloitte

To [redacted]

Deloitte.

Belgium | iPad workgroup | 21 January 2015

iPad Pilot

Do iPad increase productivity?

Dear colleague,

We are always trying to increase productivity within Deloitte. A recent survey has shown that many colleagues expect that a iPad will best contribute to a higher productivity of Deloitte Belgium. To test the added value of iPads, we will organize a pilot with 200 colleagues throughout Belgium to see how this could positively impact our productivity in FY16.

What to do?

On the special iPad portal you can read all necessary information about our vision on the central role that the iPad could play within Deloitte's daily processes. Besides the possibility to register for the pilot, it is also possible to enroll with your own iPad.

Check the portal for all information (login might be required). If you would like to enroll for the pilot, be sure to register before Friday 30 January. Click the button below to access the portal.

Go to portal

Thank you for yur interest and hopefully towards a more productive FY16!

Kind regards,

iPad Workgroup-team

Figure 13: An SPT sent to Deloitte Belgium

The above e-mail is representative of various elements which are recurring within the dataset. Starting from the top of the e-mail, we see the sender address. The e-mail is sent from the "Deloitte iPad Workgroup" with the sender address being 'IpadWorkgroup@secure-filebox.com'. The e-mail has all the Deloitte logos and color scheme yet the sender address is from secure-filebox.com domain. This is a first tell-tale sign that is intentionally included to see whether the e-mail is phishing or not. Further reading of the e-mail reveals a number of spelling and grammar errors such as in the sub-title where it asks "Do iPad increase productivity?" with using iPad instead of iPads. The overall English of the text seems to be of a lower level than that expected by an organization such as Deloitte. Also, a clear spelling mistake can be seen right under the "Go to portal" button where it reads "Thank you for yur interest […]". It is obvious that the e-mail was intentionally made to include various tell-tale signs of phishing. Yet, it had a click rate of 20.72% and 21.64% as it was sent to two groups of Deloitte employees. Looking at the e-mail without any precondition from the literature it seems that the offering of an iPad is a very strong incentive to follow the e-mails direction and apparently it made people overlook the many problematic elements of the e-mail. Summing up the elements of this e-mail we can say that it has spelling and grammar errors and purports to be sent internally[6]. It also uses logos and the internal color scheme.

[6] "Internally" here means that the e-mail is purportedly sent from the target organization to the target organization i.e. an internal department. "Externally" means that the e-mail is supposedly sent from an external organization.

Since e-mails sent to internally do not fall under confidentiality clauses, we can show another example that was sent to internal employees in the Netherlands during Christmas with the name, “Sinterklaasactie”. This exercise had a click rate of 30.06%. The e-mail is shown below:



do 26-02-2015 13:11
Van Zwieten, Marcus (NL - Rotterdam) <MarvanZwieten@deloitte.nl>
Sinterklaasactie
To [redacted] (NL - Amsterdam)

- Is dit bericht niet goed te lezen, bekijk dan hier de online versie -

Deloitte.

Nederland | Employee Services & Workplace Services | 21 november 2014

Sinterklaasactie
Wat staat er op jouw verlanglijstje?

Beste collega,

Binnenkort is het alweer 5 December, de meest gezellige tijd van het jaar. Met deze gezelligheid is het echter ook de duurste maand van het jaar. Daar willen wij jou, in samenwerking met onze partner, in tegemoed komen in de vorm van een kortingactie.

Sinterklaasactie
De actie is vrij simpel en zit in de kracht van het ‘as one’ gedachte. Door je interesse in deze actie kenbaar te maken via de actiepagina leggen we jou interesse vast. Vervolgens gaan we na de sluitingsdatum op 26 november naar onze partner met het aantal inschrijvingen. Op basis van het aantal inschrijvingen zal Boll ons een passende korting aanbieden. Deze korting begint vanaf 15% en kan oplopen tot wel 40% korting op al jouw decembercadeaus!

Registreer snel via onderstaande link om je interesse kenbaar te maken. Doe dit voor **26 november 12:00** en onthoudt; hoe meer inschrijvingen hoe hoger de uiteindelijke korting!

Sinterklaasactie portaal: https://sinterklaasactie.deloitte.nl/i/866e9bea/.

We rekenen op veel interesse en zien er naar uit jullie een geweldig aanbod te doen!

Namens het Sinterklaasactie-team,

Marcus van Zwieten
Employee Services & Workplace Services

Figure 14: An SPT sent to Deloitte Netherlands

In this example we see some differences and similarities to the previous e-mails. In this scenario, the e-mail purports to be sent internally from a realistic looking yet non-existent person. It offers a discount given the looming Christmas season. The difference from the previous example is that in this case the sender address is spoofed to look like an internal Deloitte e-mail. This e-mail has only one isolated and discreet spelling mistake where “tegemoet” is changed to “tegemoed”. It has, however, several grammar mistakes scattered throughout the e-mail. The overall theme of providing a discount is a recurring theme of the dataset. This particular example is one of the better designed ones in the sense that it does not have that many tell-tale signs of phishing. Even the signature at the end is a common element of internal Deloitte communications. There is however, a tricky but undisputable way to tell that this e-mail is phishing.

If one was to hover their mouse pointer over the sender address the following would be seen:

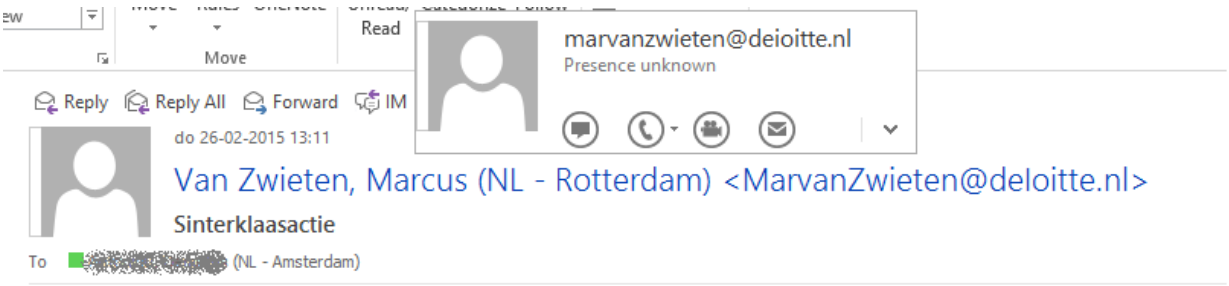


Figure 15: Hoovering over the address reveals the true sender domain.

As can be seen in the small box above, the actual sender e-mail address uses a "deioitte.nl" domain. However, it required that the receiver would hover over the sender address and look carefully for this highly obscured sign.

The "high" group is the most populous with eleven cases. The presence of this group is important since it demonstrates that the receiving organization is a major factor in the click rate of the e-mail. As stated earlier, the gardening company had an extremely high click-rate which may or may not be the result of the e-mail. So let's take a closer look at the two most successful e-mails in the dataset both sent to the gardening company starting with "GARD2". The e-mail is shown below:



Figure 16: An SPT conducted to a Dutch client.

The above e-mail has perfect spelling and grammar and purports to be sent internally. In addition, it uses the company's actual logo and e-mail theme. The sender address is from "ethischerichtlijnen.nl" whereas the e-mail is supposedly internal. This discrepancy between the sender address and the sender signature is a recurring theme in the dataset. The scenario revolves around the acceptance of new ethical guidelines for the company which could be considered a powerful drive for action. In contrast to the other examples presented here, this scenario is unique in that it *implores* that action is taken rather

than being provided in an “if-you-wish” fashion. This can be explained as part of the persuasiveness of the e-mail yet it seems that the major drive for this click rate was the nature of the company. This is because the second most successful e-mail was sent to the same company and looked like this:



Figure 17: An SPT conducted to a Dutch client.

The grammar is horrendous in this example, it does not have any legitimizing elements and generally looks like a spam e-mail. However it still had a click rate of 52.94%. It is unreasonable to try and explain this click rate from the properties of the e-mail since to most this e-mail seems as characteristically spam. For the explanation it would make sense to look at the organization itself. For one, the employees could have had trust in the spam filters of their corporate accounts and felt that any e-mail that makes it through the e-mail filters should be safe. Also, the nature of the employees’ work probably does not expose them to the dangers of phishing in general. In any case, the reason for this click rate does not seem to be a generalizable property but rather the specific conditions of this engagement.

The above examples have illustrated the heterogeneity that characterizes the dataset. Due to the large variability of organizations and the small sample size it makes it difficult, at present, to make e-mail specific prediction models based on this dataset. There can be, however, a way which could enable the creation of such a model in the future. To do so, as much formalization and generalization as possible needs to be done so as the essence of each e-mail can be fully captured.

By inspecting the dataset a number of literature-independent variables were identified which will be formalized in the next section.

4.3. Identification of common themes

The preceding section has demonstrated some characteristics of the current dataset. These characteristics are different from the literature derived ones which were explored earlier in the sense that they come from a qualitative inspection of the dataset. Trying to extract variables from the dataset presents a tricky proposition. It has to be ensured that the characteristics that are extracted from the dataset have not been covered by the literature. If they have been covered, then they serve as indicators of literature-based variables. The process by which the following variables were identified was by going over the entire dataset with the literature-backed variables in hand and try to identify if there are elements in the e-mail which could have an impact on the eventual click-rate. This specific dataset is the result of engagements done between two organizations thus, it is natural that there are some things which are bound to be unique to the dataset. After coming up with the unique characteristics, it became evident that they were indeed discussed in literature.

The table below shows the identified characteristics:

Table 15: The unique variables from the dataset.

Name	Description	Rationale
Spelling Errors(SP), Prominent or Isolated(P or I)	The e-mail contains spelling errors either throughout or in isolated cases	Literature discusses spelling mistakes but no distinction is made between prominent and isolated examples.
Grammar Errors(GR), Prominent or Isolated (P or I)	The e-mail contains grammar errors either throughout or in isolated cases	Literature discusses grammar mistakes but no distinction is made between prominent and isolated examples.
Internal/External(IM/EM)	The e-mail either purports to be sent from an internal department or an external organization	Due to the fact that the exercises are done as part of an agreement, there is the choice of making the e-mail look external or internal.
Address & Signature Discrepancy(ASD)	The e-mail's sender address and signature in the end do not match	Has been discussed by Lee & Song but not in the form found in our dataset.
Uses company logo(LOG)	If the e-mail contains the target or sender company logo	Discussed by Lee & Song.
Uses company e-mail structure(STRU)	If the e-mail uses the target company colors and e-mail structure(i.e. disclaimers at the end)	This characteristic results only by knowing in advance on how the e-mail should look like.

The above table contains the unique characteristics of the dataset. These are to be included in the final coding scheme in Chapter 5. The unique variables are straightforward in the sense that they can be measured with minimal subjectivity on the behalf of coders. Spelling Errors are there or not regardless how an individual understands it. They could be automated in the future but for now the process of enumerating the unique variables seems easy and not subject to much interpretation.

4.4. Extraction of phishing characteristics from the dataset

The previous section outlined some characteristics and recurring themes found in the current dataset. Using these characteristics each e-mail in the dataset was rated according to them.

The following table outlines the current dataset along with the specific characteristics of each e-mail:

Table 16: The unique variables of the existing dataset.

#	**Name**	**Characteristics**
1	GARD2	IM, ASD, LOG, STRU
2	GARD1	SP-P, GR-P, EM, ASD
3	MAR	GR-I, IM, ASD, LOG, STRU
4	PERO	IM, ASD, LOG
5	NORDIC1	IM, ASD, STRU, LOG
6	FOREMAN1	IM, ASD, LOG, STRU
7	BANKA3	EM, LOG
8	BANKA1	IM, ASD, LOG
9	NORDIC2	IM, ASD, STRU, LOG
10	BANKA2	IM, ASD, LOG, STRU
11	THUISBANKA	IM, ASD, LOG
12	CONSNL1	SP-I, GR-I, IM, LOG, STRU
13	TUBE1	IM, LOG, STRU, ASD
14	SSL1	GR-I, IM, ASD, LOG
15	BANKA4	EM, LOG, STRU
16	CONSBEL1	SP-I, GR-I, IM, ASD, LOG
17	COFFEEBANKA	EM, ASD, LOG
18	CLOUDBANKA	IM, LOG, STRU
19	CONSBEL2	SP-I, GR-I, IM, ASD, LOG
20	STARTING1	SP-P, GR-P, IM, LOG
21	CINEMABANKA	IM, ASD, LOG, STRU
22	BANKB2ROM	IM, ASD, STRU
23	ITCOMP1	EM, LOG
24	FRAUDBANKB	IM, ASD, LOG, STRU
25	EMAILBANKB	IM, ASD, LOG, STRU

26	SASBANKB	IM, ASD, LOG, STRU
27	BANKB1ROM	IM, ASD, STRU

As can be seen above, except in five cases, there was a discrepancy between the sender address and the signature found at the end of the e-mail. For example, an e-mail had a sender address of bedrijsaanbiendingen.nl but the signature was from an internal person of the company. This is quite an eminent element of the phishing e-mails yet it does not seem that the lack of this discrepancy leads to a higher click-rate. This particular phenomenon is also widely discussed in follow-up training sessions that industry holds after the exercise and is the most prominent and reliable way to discern a phishing e-mail as such. In cases where the sender address is obfuscated, it requires the extra step of "hovering over" in order to see the actual sender address (given that the recipient uses MS Outlook).

Currently, analyzing the dataset has been impossible. The various approaches outlined above were attempted with the existing data and the results were conflicting and inconsistent. So why was it impossible to make a predictive model using the existing dataset? We answer this question in section 4.5.

4.5. Limitations of the dataset

The inspection of the dataset has already shown some intrinsic inconsistencies of the dataset. We have a very badly written e-mail that had a 52.2% click rate and a very well-produced e-mail which only had a 20% click rate. What caused this? Was it the nature of the organization? Did someone spill the beans that a phishing e-mail was coming and made everyone click/not click the email? Did one organization have much higher awareness than the other on the dangers of phishing? This list of questions could go on and on and this is precisely the reason why we cannot conduct reliable and valid statistical analysis on our current dataset.

The presence of those questions is what makes the dataset unreliable at the moment for statistical analysis. The conditions under which the various engagements were conducted were widely different from client to client. What we essentially have is a field study of e-mails which were created in an unstandardized way and sent to an unstandardized group of people. One organization may have been comprised mostly of men. Would that be the reason that that particular organization had that particular click rate? There are many sensible ways to explain the variability within the dataset due to reasons beyond the e-mail as was shown in section 4.2. The heterogeneity of the data is extremely large. A way to offset this variability is to use a standardized approach when creating the e-mails and ensure that comparisons between organizations are done very carefully. In order to offset the variability between organizations there should be a very large number of samples with all receiving the same measurement based on the coding scheme. Ideally, each variable should be measured in an experimental condition where all other influencing variables are kept constant whilst the variable of concern is changed. The setup of such experiments is outlined in section 5.2.

4.6. Possibilities of the dataset

The dataset, although not suitable for predictive analytics has provided us with useful information which was incorporated in our coding scheme in sections 4.3 and 4.4. We were able to qualitatively inspect the dataset and extracted information which was not available from literature. The unique variables are

specific to the e-mails generated by SPTs and will be very likely be present in future engagements. The existing dataset has two possibilities with one taken advantage of. First, it can be used for qualitative review to formulate unique variables. This has been done in section 4.3 where we established our SPT specific variables. It can also be used for guiding future experimentation. For example, the cases where there were extreme values could be investigated with examining prominent elements of those e-mails. In evaluating our coding scheme we will use the work of Rourke et al. who investigated coding of computer mediated communication through text.[52]

Expansion of the dataset is perhaps the biggest opportunity. With the adoption of the proposed coding scheme the new entries to the dataset will be accompanied by several measured variables which will increase the data quality considerably. Any new entry to the dataset will not only be measured by the coding scheme but also be designed with the coding scheme as a guide.

5. The coding scheme and its application.

This section will coalesce the findings of the previous sections and generate the coding scheme. The resultant coding scheme will be both based off the literature and the unique properties identified from inspecting the dataset. In the previous sections we generated variables from three distinct sources: deception, persuasion and dataset inspection. All are included in the final coding scheme.

5.1. The coding scheme

The table below shows the final coding scheme which results from the combination of the three sources. All the identified variables from deception, persuasion and those specific to STPs were included in the final coding scheme:

Table 18: The coding scheme

Variable Name(abbreviation)	Group	Operationalization
Number of Words(WRD)	Deception	Absolute number of words
Number of Verbs(VRB)	Deception	Absolute number of verbs
Modifier(MOD)	Deception	Absolute number of modifiers
Noun Phrase(NPHR)	Deception	Absolute number of noun phrases
Pausality(PAUS)	Deception	Ratio of punctuation marks over number of sentences
Lexical Diversity(LDIV)	Deception	Ratio of unique words over total words
Content Word Diversity(CWD)	Deception	Ratio of unique content words over total content words
Uses Honorifics(HOMS)	Persuasion	“Does the e-mail contain titles which denote the sender’s position or expertise?”
Appeals to Authority(AUTH)	Persuasion	“Does the e-mail make any appeals to authority?”
Legitimate e-mail sender prefix(PREF)	Persuasion	“Does the e-mail use e-mail prefixes characteristic of e-mails?”
Legitimate domain name(DOM)	Persuasion	“Does the e-mail come from a legitimate domain name?”

Clearly Specified Sender(CSEND)	Persuasion	"Does the e-mail come from a clearly specified sender such as a person or a group?"
Contact telephone number(PHON)	Persuasion	"Does the e-mail include a telephone number?"
Contact e-mail address(EMAL)	Persuasion	"Does the e-mail contain a contact e-mail address? i.e. a reply address?"
Evidence Plausibility(EPLA)	Persuasion	"Is the evidence provided in the text plausible?"
Reasoning by Cause(RCAU)	Persuasion	"Does the text use reasoning by cause?"
Reasoning by Sign(RSIG)	Persuasion	"Does the text use reasoning by sign?"
Reasoning by Analogy(RANA)	Persuasion	"Does the text use reasoning by analogy?"
Reciprocity(RECIP)	Persuasion	"Does the text use the Reciprocity principle?"
Commitment and Consistency(COMCON)	Persuasion	"Does the text appeal to the commitment of the reader?"
Social Proof(SOCPROO)	Persuasion	"Does the text make use of the Social Proof principle?"
Liking(LIKI)	Persuasion	"Does the text make use of the Liking principle?"
Scarcity(SCARCI)	Persuasion	"Does the text make use of the Scarcity principle?"
Fear Appeal(FEAR)	Persuasion	"Does the text induce emotions of fear?"
Affection Appeal(AFFECT)	Persuasion	"Does the text make use of affective events?"
Happiness Appeal(HAPPY)	Persuasion	"Does the text induce feelings of happiness?"
Safety Needs(SAFE)	Persuasion	"Does the text appeal to the safety needs of the reader?"
Self-Esteem Needs(SELF)	Persuasion	"Does the text appeal to the self-esteem needs of the reader?"
Belongingness Needs(BELONG)	Persuasion	"Does the text appeal to the belongingness needs of the reader?"
Spelling Errors(SP), Prominent or Isolated(P or I)	SPT Specific	The e-mail contains spelling errors either throughout or in isolated cases
Grammar Errors(GR), Prominent or Isolated (P or I)	SPT Specific	The e-mail contains grammar errors either throughout or in isolated cases
Internal/External(IM/EM)	SPT Specific	The e-mail either purports to be sent from an internal department or an external organization
Address & Signature Discrepancy(ASD)	SPT Specific	The e-mail's sender address and signature in the end do not match
Uses company logo(LOG)	SPT Specific	If the e-mail contains the target or sender company logo

Uses company e-mail structure(STRU)	SPT Specific	If the e-mail uses the target company colors and e-mail structure(i.e. disclaimers at the end)

The above table is the final coding scheme which has a total of 35 distinct variables which describe a phishing e-mail. We have decided to include all the variables which were identified in the preceding sections since it is our goal to capture as much information as possible about the phishing e-mails. The deception based variables are derived from work which attempted to automatically measure deception in A-CMC and thus they can be measured using software implementation. The persuasion and SPT specific variables are dependent on human coders to be evaluated with some being more objective than others. The issue of subjectivity of coders will be further explored in Section 5.2.

5.1.1. On the method used in analyzing the e-mails

During the writing of the project, it became clear that no formalized method was used which describes how to analyze the e-mails. This identification came along later during the project so what was done was to go over the thesis and try to identify if, unwittingly, such a method was used. As stated in Section 1.2, we are looking for a "context-independent" method of analyzing the e-mails. In addition, the foundation for the analysis was from Aristotle's Rhetoric. Thus, the method of "Rhetorical Analysis" was used.[53] Rhetorical analysis is the analysis of text where the form, or structure, of the text is analyzed rather than the content. When conducting Rhetorical Analysis one attempts to see how well they make their argument and, in the classical sense, to what extend they make use of *Pathos, Logos* and *Virtue.* Although this has not been explicitly stated during the early stages of the project, it has become clear that the coding scheme is essentially a structured methodology to make a Rhetorical Analysis of the text. In the Aristotelian view, persuasion is the application of Rhetoric in attempting to influence the beliefs and decisions of an audience. The content of the message is not what is of interest here. The degree to which the argumentation is persuasive enough is what is. For example, an e-mail may offer an iPad in exchange for personal information. It can be merely a statement of facts on the benefits of an iPad but it could also appeal to all the possibilities that an iPad can offer. Moreover, the e-mail could implore on how much value is placed on the receiver and how that value results in the offer being made. The essence of rhetoric lies in making seemingly mundane phenomena appear as appealing or revolting and, consequently, effecting the desired behavior. The phishing examples that we outlined employ rhetoric to various degrees to bring about the desired behavior.

5.2. Informed Assumptions on the impact of variables.

The table below shows the informed assumptions for each variable, accompanied by the relevant rationale. For each variable a critical view was taken on how it is expected to impact the eventual click-rate of the e-mail. The reasoning from most of them comes from the understanding using common sense on what would make a phishing e-mail more successful or not. Naturally, experimental testing can support either way of each variable and these assumptions server only to provide an intuition based direction for using the coding scheme:

Table 20: The informed assumptions table

Variable Name	Assumption	Rationale
Number of Words(WRD)	Will decrease the click rate	In the original paper[42], more of these variables is an *indicator* of deception. We reverse the conclusion and say that the more of these variables will decrease the chance of them being seen as deceptive, thus decrease the click rate.
Number of Verbs(VRB)	Will decrease the click rate	
Modifier(MOD)	Will decrease the click rate	
Noun Phrase(NPHR)	Will decrease the click rate	
Pausality(PAUS)	Will decrease the click rate	
Lexical Diversity(LDIV)	Will decrease the click rate	
Content Word Diversity(CWD)	Will decrease the click rate	
Uses Honorifics(HOMS)	Will increase the click rate	The use of honorifics is usually done to convey authority. Authority in general assists in persuasion, according to Cialdini. (Virtue)
Appeals to Authority(AUTH)	Will increase the click rate	
Legitimate e-mail sender prefix(PREF)	Will increase the click rate	These variables are e-mail specific and denote the overall legitimacy of the e-mail. Their presence is expected to increase the click rate.
Legitimate domain name(DOM)	Will increase the click rate	
Clearly Specified Sender(CSEND)	Will increase the click rate	
Contact telephone number(PHON)	Will increase the click rate	
Contact e-mail address(EMAL)	Will increase the click rate	
Evidence Plausibility(EPLA)	Will increase the click rate	This variable can be expanded in terms of absurdity of the evidence. Less absurd evidence will lead to higher click rate.
Reasoning by Cause(RCAU)	Will increase the click rate	A cause reasoning scheme is the most prominent and widely used. The use of other than cause is expected to reduce the click rate. (Logos)
Reasoning by Sign(RSIG)	Will decrease the click rate	
Reasoning by Analogy(RANA)	Will decrease the click rate	
Reciprocity(RECIP)	Will increase the click rate	The Cialdini variables have shown to increase the success of marketing materials. Thus, we assume that their presence will increase the click rate.
Commitment and Consistency(COMCON)	Will increase the click rate	
Social Proof(SOCPROO)	Will increase the click rate	
Liking(LIKI)	Will increase the click rate	
Scarcity(SCARCI)	Will increase the click rate	
Fear Appeal(FEAR)	Will increase the click rate	The use of emotional appeals is expected to stir the emotions of the recipients and increase the click rate. (Pathos)
Affection Appeal(AFFECT)	Will increase the click rate	
Happiness Appeal(HAPPY)	Will increase the click rate	

Safety Needs(SAFE)	Will increase the click rate	Motivational appeals are expected to increase the click rate.
Self-Esteem Needs(SELF)	Will increase the click rate	
Belongingness Needs(BELONG)	Will increase the click rate	
Spelling Errors(SP), Prominent or Isolated(P or I)	Will decrease the click rate. If prominent, even more.	Spelling errors will decrease the quality of the email. More errors even more.
Grammar Errors(GR), Prominent or Isolated (P or I)	Will decrease the click rate. If prominent, even more.	Grammar errors will decrease the quality of the email. More errors even more.
Internal/External(IM/EM)	Internal e-mails will increase the click rate.	People will more likely trust an internal email than an external
Address & Signature Discrepancy(ASD)	Will decrease the click rate	The presence of such discrepancy is one of the original tell-tale signs of phishing emails
Uses company logo(LOG)	Will increase the click rate	The use of logos gives substance to emails
Uses company e-mail structure(STRU)	Will increase the click rate	An email which follows a defined and familiar structure will be more likely to be believed and clicked

The table above shows the informed assumptions on what the impact of each variable would be on the click rate. As stated earlier, we do not make any claims on the actual impact of each variable. Moreover, some variables may accentuate or cancel-out each other when used in tandem. The above assumptions can be the first step toward formulating hypotheses which can then be experimentally tested as explored in section 5.4. Although we cannot make any conclusion on the extent of the impact of each variables, using the informed assumptions table we can discuss the *loading* of each e-mail. What is meant by loading is the presence of more variables in the e-mail. For example, an e-mail which contains authority, happiness appeal, time pressure and uses company logo will be considered more "loaded" than an e-mail that has only time pressure and the company logo. The direction on whether a particular variable will increase or decrease the click-rate can be derived from the assumptions table but not with full confidence. In section 5.4. we explore more in depth how we can reach the point to talk about actual e-mail difficulty. For future engagements without the results of experimentation the informed assumptions table can be used to describe e-mails in terms of difficulty, derived from the e-mail's loading.

5.3. The importance of coders and practical considerations.

Coders is the term given, in the realm of social science, to the people whose job is to assign values to a construct based on a set of guidelines. What makes coders important is the training they have received so as to be able to search for the elements of interest to the construct. In our case, we would like to train our coders in evaluating the existing e-mail dataset as well as future e-mails. In addition, the coding scheme will also serve as a guide for the generation of future e-mails. As we have explored in section

5.1., the coding process is quite subjective and achieving ideal coding conditions is next to impossible. Continuing the application of Rourke's work, we will look more closely on how a coder perceives a phishing e-mail.

Rourke has made the distinction between two variable types which are present in a text to be coded: *manifest content* variables and *latent content* variables. Manifest content is, quoting, "content that resides on the surface of communication and is therefore easily observable". It is easy to make out the meaning from this statement. Manifest content is the content which is beyond doubt present or not present in the text and represents elements which are observable to all and do not require much thinking to decide their entity. On the other hand, latent content is much more subjective and requires "input of meaning" and inference on the part of the coder. Latent content is further split into *latent pattern* variables and *latent projective* variables. Latent pattern variables are the less subjective than their projective sibling where the coder will rely on "an inventory of clues" to decide whether a variable is present or not. It does require *some* cognitions and could give rise to disagreements but it is expected that latent pattern variables will be identified similarly by the majority of coders. Truly subjective variables fall under the latent projective grouping. This requires cognition and opinion by the coder to be identified. In variables like these is where cultural differences, age, sex and reasoning schema will have impact on the rating. It is here that training is required to ensure reliability of the coders.

The above paragraph has outlined the three types of variables which, according to Rourke, comprise coding scheme which aim to cade text: manifest content variables, latent pattern variables and latent projective variables. The table below assigns each variable of our coding scheme to the groupings above:

Table 19: The types of variables in the coding scheme.

Variable Name	Variable Type	Rationale
Number of Words(WRD)	Manifest	Absolute, computer measurable number
Number of Verbs(VRB)	Manifest	Absolute, computer measurable number
Modifier(MOD)	Manifest	Absolute, computer measurable number
Noun Phrase(NPHR)	Manifest	Absolute, computer measurable number
Pausality(PAUS)	Manifest	Ratio, computer measurable number
Lexical Diversity(LDIV)	Manifest	Ratio, computer measurable number
Content Word Diversity(CWD)	Manifest	Ratio, computer measurable number
Uses Honorifics(HOMS)	Manifest	Very easy to recognize
Appeals to Authority(AUTH)	Latent Projective	Requires coder cognition
Legitimate e-mail sender prefix(PREF)	Manifest	Very easy to recognize
Legitimate domain name(DOM)	Manifest	Very easy to recognize
Clearly Specified Sender(CSEND)	Latent Pattern	Presence is evident, yet coders may disagree

Contact telephone number(PHON)	Manifest	Very easy to recognize
Contact e-mail address(EMAL)	Manifest	Very easy to recognize
Evidence Plausibility(EPLA)	Latent Projective	Open to interpretation in some cases
Reasoning by Cause(RCAU)	Latent Projective	Coders need training to distinguish
Reasoning by Sign(RSIG)	Latent Projective	Coders need training to distinguish
Reasoning by Analogy(RANA)	Latent Projective	Coders need training to distinguish
Reciprocity(RECIP)	Latent Projective	Requires coder cognition
Commitment and Consistency(COMCON)	Latent Projective	Requires coder cognition
Social Proof(SOCPROO)	Latent Projective	Requires coder cognition
Liking(LIKI)	Latent Projective	Requires coder cognition
Scarcity(SCARCI)	Latent Projective	Requires coder cognition
Fear Appeal(FEAR)	Latent Projective	Requires coder cognition
Affection Appeal(AFFECT)	Latent Projective	Requires coder cognition
Happiness Appeal(HAPPY)	Latent Projective	Requires coder cognition
Safety Needs(SAFE)	Latent Projective	Requires coder cognition
Self-Esteem Needs(SELF)	Latent Projective	Requires coder cognition
Belongingness Needs(BELONG)	Latent Projective	Requires coder cognition
Spelling Errors(SP), Prominent or Isolated(P or I)	Manifest	Very easy to recognize
Grammar Errors(GR), Prominent or Isolated (P or I)	Manifest	Very easy to recognize
Internal/External(IM/EM)	Manifest	Very easy to recognize
Address & Signature Discrepancy(ASD)	Manifest	Very easy to recognize
Uses company logo(LOG)	Manifest	Very easy to recognize
Uses company e-mail structure(STRU)	Manifest	Very easy to recognize

The table above shows the categorization of the coding scheme's variables. The deception variables are sourced from literature which attempted to make deception computer detectable. Thus they represent variables which can be quantitatively measured by the computer and do not have any openness to interpretation. The use of honorifics is also a manifest variable; all that the coder has to do is to look for the use of honorifics and then mark whether or not they are used. The appeals to authority is a latent projective variable since it requires the coder to think whether the variable is present or not which may not be evident at a first glance. Training is necessary to enable the coders to identify *all* variables but for the latent projective ones even more so. We will discuss training more in depth in section 5.2.1. The recognition of a legitimate e-mail sender prefix and domain name are straightforward as well with their presence or not being not open to interpretation. The variable of the clearly specified sender may get a little tricky sometimes since to some it may mean that there is clearly specified sender as a person whereas others may view a department or committee as a sender. Contact telephone number and e-mail address are also manifest variables and easily identifiable. The evidence plausibility and the type of

reasoning are latent projective since we ask our coders to opinionate on whether they view the evidence as plausible and identify a reasoning scheme. The plausibility of evidence should be straightforward but in some cases there could be disagreements which arise wholly from the differences between coders. For example, to some, the evidence of a particular event may seem plausible because the event actually happened and was reported in the news even if it was strange. If the coder knows that something weird is true they would find the evidence as plausible. There are cases, for example, where a dog helps out a cat; this would be not plausible in most cases but weird events like these do happen. So the plausibility of evidence is a subjective variable which depends on what the coders know about the world and to what extent they are willing to consider absurd evidence as plausible. Realistically, only in very few cases the coders would disagree on the plausibility of evidence. The reasoning scheme gets trickier. In the original paper the authors discovered that mostly there is reasoning by cause with very few cases using reasoning by sign and analogy. Nonetheless, the training regimen should enable coders to distinguish between the different types of reasoning.

Cialdini's principles represent perhaps the most subjective element of the coding scheme. Their presence or not is dependent on whether or not they gave that impression to the coder. In addition, Cialdini's influence principles has seen use in marketing over many years which may have made our coders-to-be desensitized in recognizing them. Ideally, if we would like our coders to recognize reliably the presence of Cialdini's principles then training will have to include exercises where they are asked to recognize their presence or not. The same goes for the motivation-inspired variables; the differences between coders will come into play in recognizing their presence or not.

The last, unique variables are manifest variables in the sense that there is not much interpretation when it comes to recognizing them. The only source of disagreement that could arise would be differences in the distinction between prominent or isolated spelling and grammar errors.

With the explanation of the nature of each variable we will now proceed in outlining the need for training and provide recommendations on how it should be done with special emphasis on the latent projective variables.

5.3.1. Training of coders.

As we have explored in the preceding section, some variables are more open to interpretation than others. Therefore, a training regimen is necessary to ensure that the coders who will use the scheme to measure and design phishing e-mails will be on the same page. This is especially important because consistency of rating is the only way to produce high quality data. As we have explored in section 5.2. some variables' values will be dependent on the understanding of the coder. Thus in this section we propose a training approach to tackle the issue of subjectivity.

A good starting point for training future coders would be the current dataset. We can use the current dataset to provide examples in which the various variables are present or not present. This does have some implications, however, with the first being that the existing dataset will serve as the standard for future encoding. Ideally, we should look for examples where there is undoubtedly the presence or not of a variable. However, this means that the coding of the existing dataset will end up defining all further coding attempts. This is bad in terms of the replicability of our coding scheme since we will develop an understanding of the coding scheme which is based off our initial dataset. However, in terms of internal

reliability, using a common set of examples with decided properties will allow to minimize the differences in coding between all the people who received the said training.

According to Hak, coder training can be split into two sections: theoretical training and practical socialization.[54] Theoretical training is the guidelines that the researcher gives to the coders when they are to analyze a piece of text. In our case, in the most Spartan sense, the guidelines are the operationalization questions shown in table 18. However, our training needs to go deeper than that. As shown in section 5.2., some of our variables are manifest, some are latent projective and some latent pattern. For the manifest and latent pattern variables there is not much need for training since those variables' presence or not is not subject to much interpretation. Coder training becomes important with latent projective variables which are highly dependent on the understanding of the coder. In Hak's literature review (where he also identified the lack of empiricism on coder training) he mentions that not only the training provided to the coders is important but also whether or not they have similar cultural backgrounds. In our case this can be illustrated with an example. Consider asking a coder to identify whether appeals to authority are used in an e-mail. A person with a western cultural background would try to see whether honorifics are used or if the e-mail purports to be from someone of perceived authority. Someone with an oriental background and different language would be looking at the language itself to recognize authority. In the Japanese language, for example, the rank difference between two people is explicitly expressed in the overall language of a text. Even when the coders share similar backgrounds, there can be differences in upbringing which will lead to different understanding. To overcome this cultural differences, we will follow Hak's recommendation of *practical socialization*.

In practical socialization the coder team along with the researcher have extensive discussions on samples of the text to be analyzed on how they would have coded the e-mail. The coders are encouraged to externalize their thinking process, also known as the "think-aloud protocol".[55] This protocol has seen application in several fields and its essence is straightforward. Subjects are given a task and are asked to verbally explain their thinking process while they do the task. In our case, we will egg the coders on to use the protocol to explain how they coded a particular variable of the e-mail. This would happen in a communal setting where the thinking processes of each coder can be shared with the rest. The point of this would be to initiate a researcher-led discussion about the different viewpoints and reach a consensus. This will ensure that the coders have developed a uniform understanding on the presence or not of variables in the e-mail. This approach is more involved than simply providing the coders with descriptions of each variable and then leaving them to their own devices to deduce the presence or not of the variables.

A training session for using the scheme is not expected to be longer than a few hours, perhaps one half of a working day. As shown in section 5.3, some variables are merely manifest and less open to interpretation whereas others require coder cognition. The reliability and the replicability of a coding scheme, as we will see in section 6.1, are what make a good coding scheme. To ensure reliability and replicability the training sessions will require that the latent variables would be coded the same value across all potential coders. The training regime would follow the following steps:

1. A brief introduction to social engineering and phishing
2. Description of phishing e-mails from previous STPs and identification of stark variables taken from examples. It would be wise to begin with the more projective variables and ask

participants to identify their presence. In this part, it is not expected to see much disagreement between the participants as the presence or not of a variable is quite clear.

3. We would then move on to the latent variables which are harder to find consensus on. As other researchers suggest[52], the coders should express their opinion on whether they view the presence of a variable or not and then move on to have a discussion which will reach a consensus. The moderator will have examples where it is definitely decided that a variable is present or not. It is then the moderator's duty to guide the discussion toward the conclusion he wants building on the ideas of those who agree and trying to establish a common way of thinking amongst the participants.
4. With the consensus achieved, a small test can be done to see those participants who can effectively follow the scheme's motif.

Cyber-security companies conduct training sessions often for their employees who work in cyber-security. Some of these trainings are also focused on social engineering since consultancies employ social engineering methods in their penetration tests.[56] For employees there could be a phishing training segment where the coding scheme is explained and discussions are done between the participants. The current dataset can be used in examples presented to the trainees where the researcher identifies and defines the presence or not of a particular variable.

5.4. Quantitative attempts at the dataset and recommendations on analysis.

Originally this thesis was meant to generate a linear model which could predict the click rate of phishing e-mails using the existing dataset. As has been shown in section 4.2 the dataset's small size and heterogeneity between the various organizations made this almost impossible. The conditions of each e-mail were widely different from engagement to engagement therefore the number of factors which are exogenous to the e-mail are too many to enable any meaningful inference. This heterogeneity can be offset by increasing the size of the dataset to such a level where the impact of the e-mail can be identified. So, what would the best way to look at this impact be? In this section we will explore some statistical approaches which the author attempted at the existing dataset and recommendations will be made with the advantages and disadvantages of each. Finally, the most appropriate statistical approach will be recommended.

We will begin the Linear Regression. Linear Regression is one of the oldest and most studied predictive tools available, first published in 1908 by Francis Galton.[57] Linear regression attempts to generate a simple linear equation which tries to predict a response or dependent variable based on a number of explanatory variables. There are many variations of linear regression; in our case we have one response variable and many explanatory variables. Thus, for this dataset multiple linear regression would be the type of regression suitable. Linear regression generates a simple and easily understandable model which can, sometimes, achieve good prediction yet at the same time being explanatory as well. Linear regression, however, has a number of assumptions that have to be met in order to produce reasonable results. The most important assumption is that there needs to be knowledge of the prior distribution. For example, linear regression can be performed if we know that the response variable is linearly or normally distributed. If one wanted to measure, for example, whether drinking milk influences height linear regression is suitable because *we already know* that height in the entire population is normally distributed. As shown in section 4.1 our data is not normally distributed. There are more generalizations of linear models which do not require such assumptions but then they trade it off with less robustness

(i.e. the ability to predict out of sample cases). In addition, the issue of collinearity is especially impactful on the results of linear regression. As will be shown in section 5.2, very often some variables are almost always present together. This collinearity would throw off the results and produce a meaningless model. Finally, standard regression does not include any form of variable selection which would enable it to extract the more relevant features from the set of available predictive variables. This often results to the problem of overfitting where the model manages to predict almost 100% of the variability in the dataset using complex linear terms and interaction effects. Although this would be useful if one's goal to solely explain what happens in the current dataset without need for future predictions, these over fitted models do not usually produce meaningful results. In our situation linear regression seems as the most straightforward approach which can be easily understood and is widely used in numerous applications. However, the fact that the relationships can *only be linear* severely limits the explanatory power of the resulting model. There could have been non-linear relationships which have, at best, been approximated in a linear fashion. At worst, the resulting model could claim that there is no effect of a particular variable while in reality it may have an impact. The deficiency of over-fitting, however, has been overcome with regularization (or penalization) techniques like stepwise linear regression and LASSO regression.

Stepwise regression works by iterating over different numbers and combinations of predictors with the decision to add or remove a predictor based on whether the inclusion or the exclusion of a predictor improves the model based on a defined criterion.[58] This approach is more computationally efficient than a "best subset selection" approach with 2^p where *p* is the number of predictors. It is an extension of such an approach which allows for the development of sparse models. Stepwise regression can be forward (where it starts from zero variables and adds), backward (where it starts with all the variables and removes) and hybrid where multiple add and remove iterations are performed. Freedman explains very eloquently how regression's pitfalls may not be apparent at first and violations of the assumptions can lead to untrue models, an issue not solved by stepwise regression.[59] It is considered by some researchers as not a very reliable method since the approach of adding or removing variables according to the increase or decrease of a criterion using the same data will almost always give a model which seems "reasonable" but in essence it may be something totally unrelated.[60] The researchers propose more rigorous variable selection methods such as LASSO regression.

LASSO regression is a form of penalized regression developed by Tiribashi.[58] It works by introducing a penalization factor which, as it increases, it reduces some variables to very small coefficients and some are even set to zero. This results in a sparse model which uses the smaller number of variables to explain the response variable. This approach is useful in high-dimensional problems where there is a large number of predictors and we want to not only know which predictors are the strongest but also to what extent. In the field of genetics methods like this are widely used where, for example, with a sample of thousands of subjects they want to find the genes responsible for a particular illness with *millions* of genes to choose from. In our case, the dataset was too small to make any LASSO-style analysis; although we could have a variable to cases ratio which is what LASSO is made for, the absolute low numbers make such approach unfeasible.

Going beyond linear models, we will look at Fuzzy Logic, Neural Networks and Neuro-Fuzzy logic. Fuzzy logic was first postulated by Zadeh in his work "Fuzzy Sets".[61] In essence, in fuzzy logic, the truth of any statement is a matter of degree. For example, the temperature of 20 degrees Celsius can be somewhat warm and somewhat cold. This matter of degree is achieved by creating membership

functions; in our example we have the membership functions of warm and cold. These functions can take any shape that the researcher defines based on their expertise or previous research. The image below shows an example for triangular membership functions:

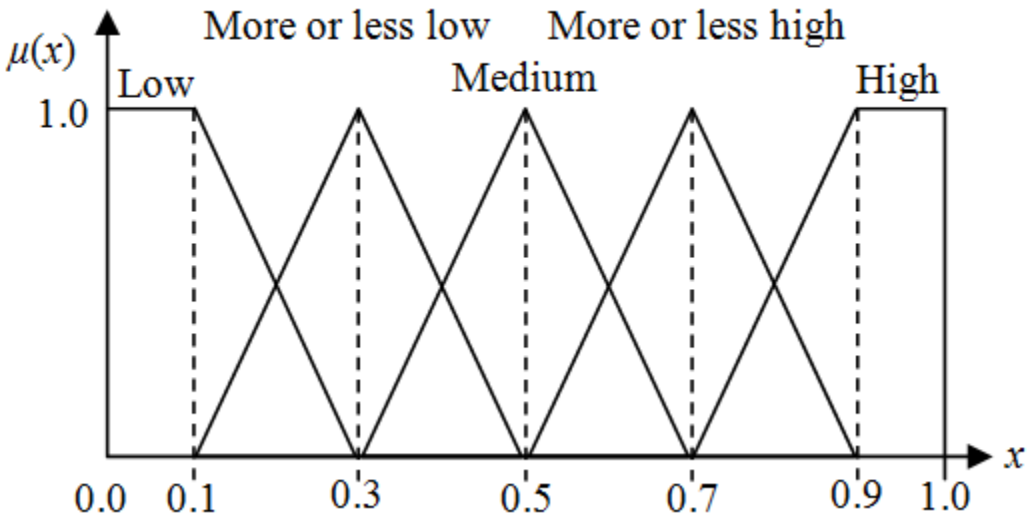


Figure 18: An example of membership functions in Fuzzy Logic. From Hameed and Sorensen. [56]

As the image above shows, the numerical value of x will define to what degree the value is a member of low, medium and high with "more or less" in between. Fuzzy logic is especially useful as it can incorporate linguistic vagueness which is often encountered in social sciences. For example, an answer such as "I find this e-mail persuasive" could be from 1 to 5. If 1 then it is not persuasive at all or at 5 is fully persuasive (these are called crisp sets in Fuzzy Logic). The in-between answers however can be expressed in terms of limited membership in the fully persuasive and not persuasive groups. This linguistic vagueness can be incorporated in the fuzzy logic set. After defining the predictors in a fuzzy set, fuzzy inference systems then need a set of rules which give structure to the outcome of the system. These rules follow a simple IF-THEN rationale. Using our own example, we could create a rule stipulating:

IF time pressure = true AND internal sender = true THEN click-rate = high

IF time pressure = false AND internal sender = true THEN click-rate = medium

The click-rate could also be expressed in absolute numbers as a scalar response number (which is what our response actually is).

Fuzzy logic systems have been applied in a number of applications with much success. Their design, however, can be tricky since there is a large part of the model which is dependent on the designer's insight and understanding. One could say that experience and knowledge of the context are extremely important if one is to design a good fuzzy model. There needs to be a prior understanding of the context-specific rules which are pertinent to the shape of the membership functions. Neuro-fuzzy approaches have offset the need for expertise and will be discussed later in this section. Looking at our example with the total number of words as our predictor. We could assume that the membership

function follows a normal function which translates to very long and very short e-mails having a negative impact on click-rate with the most successful e-mails being situated around the mean with one standard deviation. Although this “makes sense” it is still the researcher’s opinion unless earlier work which studied only that variable in a controlled setting has demonstrated that the distribution is indeed normal. In our case, since we ask whether a variable is present or not (save for the deception variables) a fuzzy logic system would be working with crisp sets. If fuzzy logic is to be used then the operationalization of the variables will have to change. This could introduce more information than there is now; each variable will be present to a certain degree. This is also the main advantage of Fuzzy Systems; they can handle human linguistic input well. However, the results will be highly dependent on the coder/designer which may lead to ungeneralizable results. The model that results from a Fuzzy Inference System is also understandable as it gives a clear model with the membership functions and the IF-THEN rules. But what method would be suitable if we were to have no previous knowledge or assumptions on what the effect of each predictor?

Artificial Neural Networks (ANNs) are a method for control systems and predictive models which have seen a renewed interest lately. In contrast to the previous methods explained, ANNs can be constructed without any prior structure or assumptions and can detect patterns by comparing the predictor and response variables. An explanation of the more technical workings of ANNs would be beyond the scope of the thesis yet it is useful to look at some characteristics of ANNs and how well ANNs could work with our problem. ANNs are capable of working with noisy and small datasets. They can do this by recognizing small non-linear interactions between the predictors and the response variables including interactions between the predictors. The big drawback, however, is that the resulting model is a ‘black-box’. This means that it is next to impossible to understand which the underlying factors of the predictive model are. Although neural networks have a very good predictive power, analysis of the model cannot be done thus there is a lack of explanation on what caused the result. This makes neural networks not very useful in our case since we want to know which variables have an impact on the click-rate. However, the ability of neural networks to build a ground-up model without any prior assumptions is an extremely powerful feature which has been combined with Fuzzy Set theory to create Adaptive Neuro-Fuzzy Inference Systems. (ANFISS)

ANFISS systems are an interesting combination of Fuzzy Logic and Neural Networks. They maintain the adaptability and predictive power of Neural Networks yet they still enable the analysis of the resultant model within the confines of Fuzzy Logic. When using ANFISS a researcher can either choose to have the membership functions or the IF-THEN rules be discovered by the system itself. In most applications the designer defines the membership functions and the system will generate the IF-THEN rules. This approach was used by Zhou et al. where they wanted to see how Deception Detection occurs in A-CMC under uncertainty.[62] They achieved, what they called, knowledge creation since the result of their work was the set of IF-THEN rules which defined whether a text was deceptive or not. A major difference from our case, however, is that their response variable was a true/false statement if the text was deceptive.

ANFISS systems follow the ANN approach of learning but maintain the linguistic input of Fuzzy systems. This makes ANFISS an extremely useful approach when we want to know how humans understand something. ANNs are an attempt to emulate the functioning of neurons and Fuzzy systems take into account the vagueness (or fuzziness) that encircles the decision making process of a human. Therefore, the author recommends that ANFISS is used for future analysis of future phishing e-mails whilst keeping

the current coding scheme; a 100% true/false statement is called a 'crisp' set in Fuzzy Set theory which are special cases of the general membership functions. By using ANFISS from the onset, we can still gain insight with the crisp responses that we have now and still enable the analysis of more vague statements in the future.

The author has attempted all the above methods in his quest to generate a predictive model which could predict the click-rate of future e-mails. Standard regression produced overfitting models which did not have any predictive power whatsoever. LASSO approaches and Stepwise regression did manage to produce sparse models with good errors but then the resulting models failed the ANOVA lack-of-fit test. ANNs and ANFISS systems did show some promising results but they were erratic. This is because the samples were split into model generation and validation sets which led to extremely small sample sizes in both cases. With every iteration of the model a very different model would arise instead of optimizing towards a solution. Fuzzy Logic was given consideration but it also produced erratic results since the shape of the membership functions and the IF-THEN rules had to be built from scratch with a massive amount of researcher bias.

This begs the question, which statistical method would be best for SPT analysis? Firstly, there cannot be any reliable statistical inference before the accumulation of at least 100 cases. Although the number 100 may see arbitrary it nevertheless marks the entry into the "hundreds of cases" space where statistical inference becomes much more reliable. Regression is the most conservative analytical tool and general rules of thumb state that at least 10 cases per variable under study be used. Since we have many variables per case w at least 100 cases are needed to start analyzing the entire e-mail.[63][64] Given enough variables then we have to choose which method to use. It is suggested that all are at least attempted and then their results be qualitatively reviewed. The table below shows the advantages and disadvantages of each approach:

Table 17: The various statistical tools, their advantages and disadvantages.

Name	Description	Advantages	Disadvantages
Linear Regression	Linear relationship between predictors and response.	Easy to use and apply, understandable model, widely used.	Many assumptions on data, only linear relationships, complex models with interaction effects
Stepwise Regression	Same as linear regression with add/remove of variables according to their impact on criterion	Sparse modelling, removes variables which do not contribute, can be optimized for out-of-sample prediction	Controversial method with questionable reliability, removing a variable does not mean it is not important, only linear relationships, complex models with interaction effects
LASSO regression	Similar to Stepwise Regression with a more rigorous removal of variables	Sparse modelling, removes variables which do not contribute, can be optimized for out-of-sample prediction. The	Removing a variable does not mean it is not important, only linear relationships, complex models with

	based on a penalization factor.	penalization factor can be decreased as sample size increases	interaction effects. Many assumptions.
Fuzzy Logic	Truth of any statement is a matter of degree. Something can be somewhat true or false.	Handles linguistic input well. Good approximation of human reasoning. Allows for non-linear relationships to be studied.	Requires extensive design from the researcher; introduces researcher bias. Needs to be supported by adjunct research on the shape of membership functions and IF-THEN rules.
Neural Networks	Emulation of biological brain with multiple learning neurons. Learns from the data.	Best predictive performance of all methods in the table. Learns and adapts as data grows. Can handle noisy and low-quality data well.	Next to impossible to review the resulting model; black-box modelling. Requires data to train itself before making reliable predictions; large datasets needed.
Neuro-Fuzzy Inference	Combination of Fuzzy Logic and Neural Networks.	Learns from data and handles linguistic input. Can either learn the rules or the membership functions. Very good predictive power and model is understandable by humans.	Requires researcher bias or adjunct research to create the model. Needs large dataset to train for future prediction. New, unproven approach.

Looking at the above table we can say that all the above methods would be suitable for analyzing our dataset. However, it is suggested that the analysis methods follow the increase of available cases. Before 100 cases it is suggested that the cases are only qualitatively reviewed and measured according to the coding scheme. At 100 cases linear regression could be used to demonstrate which variables *seem* to have the most impact. If we would like to pinpoint the variables which have the most impact then stepwise regression and LASSO can be used as a way of Principal Component Analysis. There are inherent risks when using linear methods; we cannot know if the impact that a variable has is linear or not and we use a model which assumes this. They can however give some reliable information initially. More importantly, we can use linear methods to investigate each variable and then we can incorporate the results in a non-linear model such as Fuzzy Logic or Neuro-Fuzzy inference.

Given enough cases it is recommended that more complex analysis methods are used. Neural networks would be the simplest to implement. Since they need training, it is recommended that 30% of the dataset is set aside for training(split for training and validation).[65] Usually the training set is randomly selected from the overall dataset but this has some risks to it. As we have shown, our dataset can have some very deviant results so it is necessary to be careful when splitting the dataset. MATLAB's Neural Network Toolbox allows for other splitting methods such as interleaving where the training and

validation datasets have similar numbers in their variables. This is to ensure, for example, that the training dataset does not end up with extreme values and then the validation set has more “regular” values. If the dataset is large enough then true randomization would ensure equal representation of all classes of cases in both the training and validation sets. Fuzzy Logic requires that the researcher constructs several membership functions and IF-THEN rules. If the number of predictor variables is small and their nature can be reliably assumed then a Fuzzy System could be constructed. However, in our case we have over 30 variables which translates to over 30 sets of membership functions. We also do not know the nature of these membership functions. We could conduct linear tests on individual variables and then use those results in the membership functions. We also need to set the IF-THEN rules as well which would require that we investigate the interaction effects for the variables as well. This is quite the research necessary to make a good fuzzy logic system.

Adaptive Neuro-Fuzzy Inference Systems are the best approach for our case. It overcomes the problem of researcher bias by having a learning element in where it can adapt according to the data. We still have to create either the membership functions or the IF-THEN rules. It is recommended that A/B testing is conducted to understand the nature of impact of each predictor on the response variable and then construct an ANFISS with the knowledge gained applied in the shape of the membership function. Then the system itself would generate the relevant IF-THEN rules by itself. Because it has a learning element and can handle linguistic input well ANFIS is recommended as the tool to be used when the dataset expands to sufficient levels.

5.5. Towards prediction: Experimentation using the coding scheme

The engagements conducted present a massive opportunity to conduct proper scientific experimentation so as to make conclusions on the nature of phishing exercises. As the above sections have shown, the data that is generated with the current approach does not allow for reliable results due to the many factors which are unaccounted for. In this section, a set of guidelines will be provided to enable experiments and sound conclusions.

Before we make the guidelines we first need to look back and see what parts of the current approach prohibit the generation of good scientific data. The SPT engagements are conducted over different organizations which precludes most comparisons to be done among them. If the number of engagements was in the hundreds then the effect of the e-mail could be understood since the impact of the organizations would be diluted over the large sample. The way with which we can generate sound data is by experimenting with the same group of subjects once and treating the organization as the population under study. As was shown in section 4.2, e-mails which would at first glance have very low click rates had much higher click rates than others which were very well produced. We therefore need to develop a methodology which will enable the testing of e-mail differences *within* the *same* population.

To do so the use of A/B testing is suggested with the improvements proposed by Deng et al. on how to reducing metric variance using pre-existing metrics.[66] A/B testing is rooted in the original t-test published by Student.[67] A/B testing is essentially comparing one variable whilst keeping everything else equal. It has seen widespread use by online companies where it is easy to manipulate one condition and measure the impact that the manipulation has. For example, Microsoft conducts thousands of A/B tests every year on Bing where they manipulate only a small element of an advertisement and measure

the impact of the manipulation on the advertisement's click-rate[68]. In order to reduce the variance of the results however, Deng proposes the use of pre-experimental variables in the experimental analysis. We could also call these pre-experimental variables demographic variables; these variables define the subjects of the experiment and it is necessary to ensure proportional representation of these demographic factors. For example, we send a phishing e-mail to a company with a total of 1000 recipients spread across three departments. The client would want to know how each of these departments fared. Thus, each recipient will have their department included along with the e-mail list provided. In some cases, the position of each recipient is also provided (e.g. manager). If we were to conduct an experiment with this group of people we need to split the recipients into two groups *in which all demographic factors are equally represented.* In our example, we would make two e-mails to be sent to 500 people each. In these two groups of 500 there needs to be the same number of managers and the same number of people from each department. The graph below shows the sampling approach proposed:

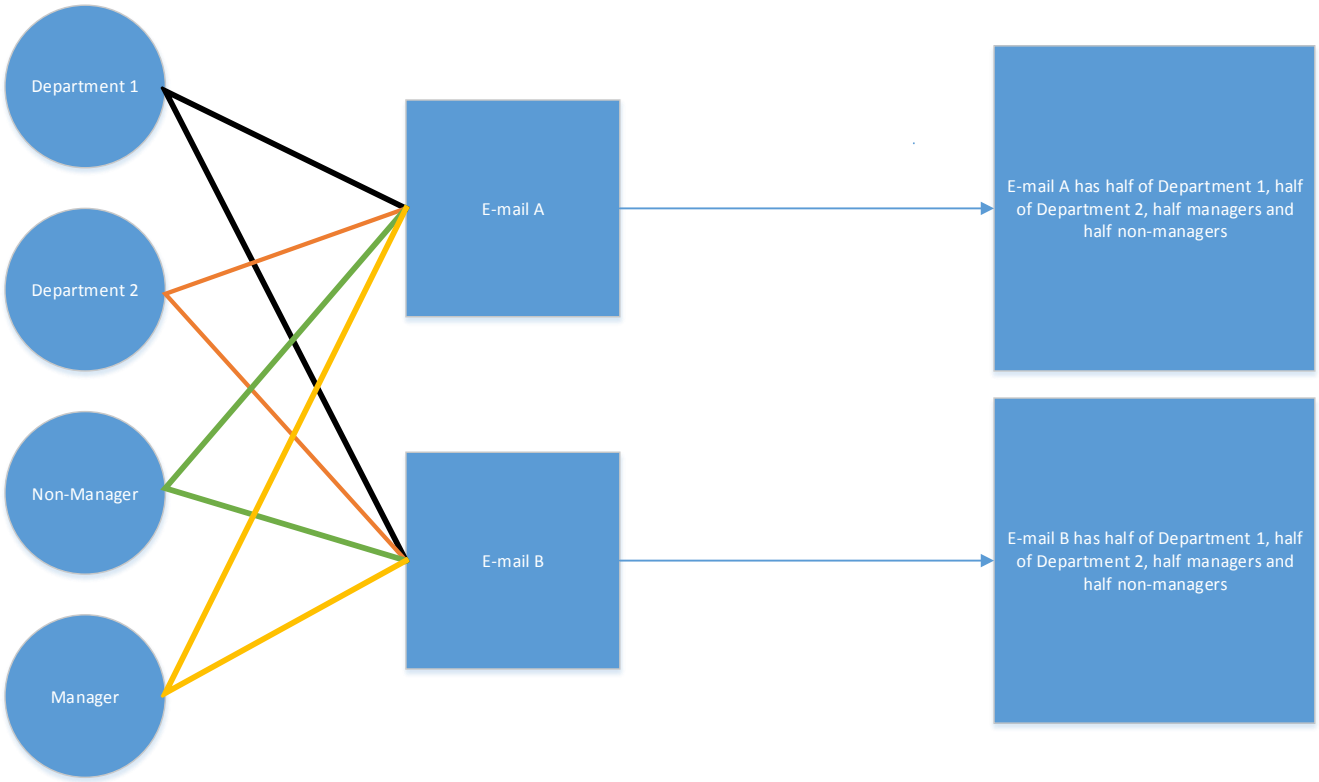


Figure 19: The A/B testing paradigm.

The above diagram represents the sampling methodology proposed. We need to take into account also the fact that there is overlap between the different demographic factors. Moreover, there can be more demographic factors. In reality, though, the amount of demographics available during engagements is very little; the departments and the position at best. This doesn't preclude, however, the ability to include them if they are available is necessary. Thus we propose that a modular approach is used where there can be an infinite number of demographics available. For example, the infrastructure can be used to conduct an academic study into phishing with the use of student volunteers. Since it is a non-work

environment more demographic information can be collected about the subjects such as age, sex, country of origin, income and so forth. The two groups of A/B testing will then be split with the same approach; ensure that all the demographics are equally represented in both conditions. This would enable the conduct of true experiments instead of quasi-experiments where the control and test conditions receive truly random assignments from the population. Each group will have a defined number of members, and out of each group the members to either A or B will be assigned randomly.

To illustrate we use the following example: let us say we have 1500 recipients which are from three departments, Department 1, 2 and 3. Some of our recipients are managers and some of our recipients are women. Stating exact numbers is unnecessary since the procedure can be used for any number of recipients and any number of demographic factors. The following steps need to be taken:

1. All permutations of all demographic combinations need to be explicitly stated and treated as unique. In our above example we would have Dept1, Manager, Male; Dept1, Manager, Female; Dept1, Non-Manager, Male; Dept1, Non-Manager, Female etc.
2. We assign each recipient to their respective grouping. We should randomize our recipient list before assigning them to groups to reduce bias.
3. We then split the members of each permutation and assign half of them to group A and the other half to Group B.
4. In case we have an odd number of members in a permutation we arbitrarily add the odd one to either group but ensuring that the next odd member will be added to the other. This problem of odd numbers is severely reduced as the sample size increases.

By following the above method we end up with two groups which have, as much as possible, equal representation of all the different subjects. Thus this would reduce the impact of these factors when comparing between e-mails. With the groups split, we can now proceed in sending them the e-mails. With two groups we can test the effect of one variable. If we would like to test more than one variable (including interaction effects) then the formula that dictates the number of groups is $2^{Number\ of\ variables}$. So if we wanted to test two variables we would need four groups, for three variables we would need 8 groups. Two to the power of number of variables under study. The schematic below shows how A/B testing can be applied with a working example:

- We have 1000 subjects
- Some from Dept. 1 some from Dept. 2
- Some are managers
- We can have as many *demographics* as we want
- We want to see if time pressure leads to higher click-rate

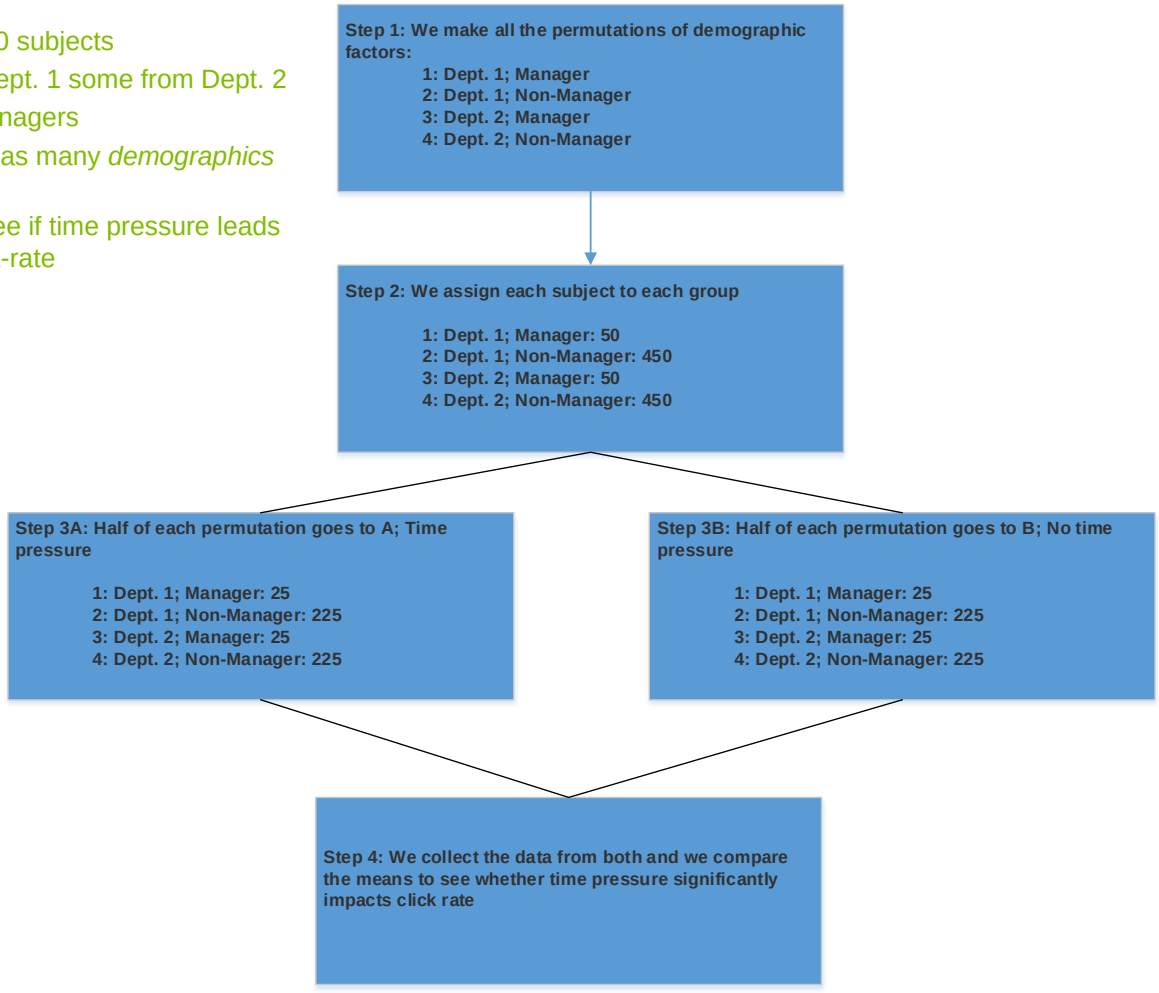


Figure 20: An example for A/B testing.

Below we show how these two e-mails would differ:

Dear Colleague,

We are pleased to announce a new raffle where you can win a trip for two to Cyprus.

Please click the link below to enter:

www.phishlink2cyprus.com.cy

Sincerely,

Anton de Metre

Dear Colleague,

We are pleased to announce a new raffle where you can win a trip for two to Cyprus.

This offer expires in 3 days so please sign up while it lasts!

Please click the link below to enter:

www.phishlink2cyprus.com.cy

Sincerely,

Anton de Metre

As shown in the example above, the e-mails are essentially the same except the sentence which implores that the offer is available only for 3 days. If we were to send these two e-mails to two groups of people with minimal demographic differences then we could discuss about the results more soundly. The hypothesis on whether time-pressure can have an impact on the click-rate of an e-mail can now be investigated.

6. Evaluation of the coding scheme.

In this section we will evaluate our coding scheme based on literature-derived evaluative methods for coding schemes. We will see to what extend the research questions were answered and how can the work done here be applied in practice.

6.1. Evaluation of the coding scheme: a first test

Currently there is no software implementation which enables the creation of data based on the above variables. As a proof of concept however, a Google Forms instance was generated and the questions were presented to two female university master students. One had a background in Education and the other in Business Administration. Their answers produced some interesting results. By using Cohen's Kappa as a measure of inter-rater agreement some questions had a full agreement whilst some full disagreement. This can be attributed to the nature of the questions. For example, the question of time pressure had 100% agreement because it is very obvious when an e-mail asks for action to be taken immediately or not. However, there were disagreements on the question on authority. One coder viewed the fact that the e-mails are corporate as an indicator of authority. The other viewed authority as the presence of a person's position within the company. In social science the training of coders is of paramount importance and will be further explored in Section 5.2. This section will continue with a discussion on how coding schemes are rated and what can be done to improve our coding scheme.

A coding scheme, also called a Quantitative Content Analysis, is a protocol used to give structure, meaning and understanding of communication.[52] The work of Rourke et al. on coding schemes will serve as the basis for our evaluation of our coding scheme. Rourke has identified that there are four criteria which are necessary to evaluate a coding scheme: its objectivity, its reliability, replicability and systemic coherence with one criterion leading to the next in the order they are presented here. So objectivity issues lead to reliability issues which lead to replicability issues which finish in the systemic coherence of the coding scheme. We will begin with objectivity.

6.1.1. The objectivity of our coding scheme

Objectivity is, in our case, the degree to which the coding of variables is open to interpretation by the coders or, in other words, how subjective is the answering of a question. To illustrate we will use a variable from our dataset, Authority. Cialdini discusses authority in terms of the presence of a person who is perceived to be authoritative in the eyes of the individual. In essence, if the individual thinks that the e-mail comes from a person who has some sort of authority then the recipient will be more likely to do what they ask, ex-officio of their authoritative position. Now if we want to ask someone whether a particular e-mail makes use of the Authority principle then the answer is highly dependent on what the coder perceives as authority. In the attempt to code the dataset one coder marked *all* e-mails as having the authority principle whereas the other only some. When they were asked, the coder who marked all with authority mentioned that e-mails sent from corporations always make use of the authority that

companies in general project. The other coder assumed authority was projected whenever a name was used as the sender instead of the e-mail being anonymous or being sent from a group or committee. This disagreement demonstrates two things: first, some variables to be coded seem to be less objective than others and that training is necessary to ensure a common mindset between coders. We will discuss training more in depth and make the distinction between objective and subjective variables in section 5.2. The presence of non-objective questions gives rise to the problem of reliability between coders or *interrater reliability*.

6.1.2. The reliability of our coding scheme

As we discussed in section 5.1.1. some questions are far more subjective than others. For coding schemes we talk about interrater reliability. This is the extent to which two or more coders coded the same value to the same piece of content, in our case, the e-mail. If we see consistent disagreements in the entirety of the coding scheme then this means that our coding scheme needs to be reworked. If we see that one particular question is consistently disagreed upon then the question either needs to be changed or we need to train our coders better. Reliability has many ways to measure; it is recommended that for this case we use Cohen's Kappa as a measure for interrater agreement. The reasons for this choice are two. First, Cohen's Kappa is one of the most conservative approaches in measuring interrater reliability and second it takes into account the possibility that the agreement was by chance. Values over 0.75(out of 1) constitute an excellent agreement. We would, ideally, want to get over 0.75 of agreement to consider our results to have any gravity. This agreement level is inextricably connected to the subjectivity of the individual questions which was discussed in section 5.1.1. In section 5.2., we will look closer where do disagreements arise in our set.

6.1.3. The replicability of our coding scheme

Replicability of the dataset is the situation where when the coding scheme has been applied elsewhere, the results are similar to the original results which were by the original creator of the coding scheme. According to Rourke's investigation, he has not been able to find a paper where the coding scheme is used from other researchers. He claims that application by someone else and production of similar results would be the highest verification of reliability of the coding scheme, a statement with which we agree.

6.1.4. The systematicity of our coding scheme

Quoting Reber, systematic refers to "a more or less well-structured set of ideas, assumptions, concepts and interpretative tendencies, which serves to structure the data of an area"[69]. From this definition is where our project takes its name from. The purpose of the coding scheme is to systematize the generation and evaluation of phishing e-mails. Rourke operationalizes the systematic nature of a coding scheme by naming a coding scheme as systematic if it is based on compatible theoretical foundations. In our case, we make the claim that phishing is, by definition, a deceptive text which in turn is persuasive. We have used the theories developed for deception and persuasion which we think are one-way related. Deception always has persuasion but persuasion is not always deceptive. We thus believe that our coding scheme is systematic in the sense that it uses related theoretical foundations with unique variables added from the dataset inspection.

6.2. Using the coding scheme for future SPT engagements

For future engagements, there are two main applications for the coding scheme. First, it can be used to guide the generation of new e-mails. This is the main purpose of the coding scheme as it will enable the systematic and measurable generation of phishing e-mails. Whenever there is to be a new phishing exercise the phisher can design the e-mail based on the variables. Special care was taken not to make any claims on to what extend each variable will impact the final click-rate. Although we can make informed assumptions on each individual variable whether or not it could increase or decrease the click rate we still refrain from making any conclusive statement on the direction and magnitude of the variables' effects. If one were to make conclusive statements on the impact of specific variables then proper experimentation should be conducted. Guidelines on how to conduct experiments which could conclude on the impact of variables were described in section 5.4.

6.1. Limitations of the coding scheme

The coding scheme's focus is mainly the text of the e-mail. This was party motivated from the fact that there are plenty existing solutions whose aim is to detect spam e-mail based on technical control. In our case, we wanted to look deeper into the understanding of phishing e-mails by humans and thus we focused on the deception and persuasion of the text. We have included more functional elements of phishing e-mails such as logos and structure but we have not explored these in depth. We only test for presence of them or not. We could have discussed about the quality of the logos that are in the e-mail yet we only study its presence. It has to be noted that the focus of the coding scheme is on the text; it enables it on one hand to find applicability elsewhere but also limits its scope.

The greatest limitation of the coding scheme is the fact that the human coded variables can only have a true or false value. This was done to make analysis in the future easier and to reduce opinion bias in the responses of the coders. They could be reformulated to include the level of each variables. For example, we could have a 1 to 5 scale and ask the coders' questions like "To what extent does the e-mail use the Authority principle?" This would have resulted in much more informative coding but it would also increase the bias of coders. This extra information could be used with advanced statistics like Neuro-Fuzzy logic which can handle linguistic input well.

On rational persuasion, the coding scheme discusses the plausibility of evidence and the rationality of the argument. The plausibility of evidence could be expanded to include the notion of mentioning facts that happen around the world and investigating whether they have an impact. The coding scheme, in retrospect, seems to be focused mostly on the emotional side of persuasion rather the rational. This does not necessarily mean that our coding scheme fails to capture the essence of a phishing e-mail but there could be an extension which focuses mostly on the rationality of the argument.

Care should be taken when using the coding scheme to measure e-mails from the wild. Although we believe that this coding scheme encompasses all the deceptive and persuasive elements of an e-mail they have been based on the results of SPTs and not of real-world phishing attacks. A useful application would be to see whether threat actors who do spear-phishing attacks could be categorized according to their phishing e-mail characteristics. On a cautionary note however, there may be the need to add unique variables which are specific to real world phishing attacks.

7. Conclusion

In this section we summarize the findings of the project and we discuss the implementation of the thesis in future SPTs. We also take a look back and see to what extend we have achieved our research objective and answered the research questions.

7.1. The deliverable

The main deliverable of this project has been the coding scheme which will allow the generation of new SPTs and also the measurement of existing e-mails. In addition, it can be used to measure real world phishing attacks.

This project has been conducted to enable the systematic generation of SPT e-mails and the enabling of their future statistical analysis. This has led to the creation of a coding scheme of phishing e-mails which, to our opinion, encompasses all the elements which could have an impact on the click-rate. The title of the thesis uses the word "systematic" and in this conclusion we will evaluate whether or not we have really systematized the generation of SPTs. We touched on the word systematic when we discussed whether or not our coding scheme is systematic or not in section 5.1.4. We restate Reber's definition of systematicity, "a more or less well-structured set of ideas, assumptions, concepts and interpretative tendencies, which serves to structure the data of an area". The definition uses the word structure twice which leads us to the understanding that a systematic process is structured. This is contrasted to an ad-hoc, tacit knowledge approach which characterizes nonsystematic approaches. An interesting treatise on this distinction comes from Bohn who studied the evolution of technological knowledge from a craft-based system to a fully systematized process[70]. The table below shows the differences between the two, taken from Bohm:

	Embryonic technology: "Art"	**Ideal technology: "Science"**
How *activities* are executed	Zero procedure; idiosyncratic	Fully specified procedure
What *results* are achieved	Each one different, mostly poor	Consistent and excellent
Characteristics of *knowledge*:		
How knowledge specified	Tacit	Codified
What knowledge about	Purely know-how	Also know-why
Extent of knowledge	Minimal; can distinguish good from bad results, but little more	Complete

Figure 21: Bohn's table; the difference between Art and Science. [63]

Using the above table we will see how our coding scheme has achieved moving from Art to Science. Starting with the first row, the table makes the distinction in the process. The overall process of the entire engagement is outlined in section 1.1. This coding scheme does not aim to change the entire process but only the process dealing with the making of the e-mail. As discussed in section 1.1., currently the e-mails are done in consultation with the client on deciding the context of the e-mail in a pretty much ad-hoc way, or in other words, idiosyncratic. With the application of the coding scheme there will be a fully specified procedure on how to write the e-mail. The coding scheme can be used to guide the generation of phishing e-mails more procedurally. For example, the designer can decide with the client whether they want to include any of Cialdini's principles of influence or motivational appeals. There is a clear set of design decisions to be taken when writing an e-mail or in other words, there is a "fully specified procedure".

The results row is where the table may not be entirely applicable. It claims that embryonic technology achieved heterogeneous results but also mostly poor results, a statement with which we disagree. The desired result of every SPT conducted was to raise awareness of employees and measure their susceptibility to phishing attacks, a result which has been consistently achieved. However, this does not mean that the results *are* different in the sense that we do not exactly what was responsible for their difference. This leads to the second part of the table which deals with the knowledge.

As explained earlier, the phishing e-mails are generated by specialists who have learned how to make good phishing e-mails through experience and repetition. However, it has remained up to this point an empirical situation where we have very highly-skilled specialists but it has been difficult to pinpoint the exact elements of this "high-skill". The coding scheme codifies the elements which distinguish between good or bad phishing e-mails or more precisely, between high and low click rates. The second row of the second part is where our coding scheme really shines. So far it has been the know-how of the specialists who developed it empirically. With the coding scheme, we are now in a position on why a phishing e-mail had the click rate it did. Although at the moment we can only make informed assumptions on the direction that each variable has on the click rate we are in a position to measure and estimate of the impact of each variable on the click rate. Finally, specialists and laymen are able to tell between a good and bad phishing email yet explaining why becomes a little trickier to extrapolate beyond the gut feeling. With the coding scheme, we hope that we have achieved a complete understanding of the inner workings of phishing e-mails. Finally, with this coding scheme the knowledge has evolved from personal to collective as the same understanding of phishing can be shared among different people.

This coding scheme will be incorporated in a new Phishing platform currently in development by Deloitte. The deception based variables will be measured by computer code and the human-coded variables can be ticked by the e-mail author while the e-mails are being written. The results of the measurement and the ticks will be included in the data file of the e-mail. The platform will also be able to add/remove variables which will enable the platform to adapt to changes that may occur in the future, such as a more elaborate coding scheme. The same applies for the A/B testing. The new platform will enable the generation of almost-identical e-mails with only one difference, as explained in section 5.5. Again, there will be a variable number of demographic factors that can be added/removed as per the needs of the particular instance. For example, if we were to use the platform to do an A/B test in an academic setting we could ask the subjects to provide more personal information(more than we could ask a client) and use that information to answer hypotheses such as "Does the city of origin affect the likelihood to click on a phishing e-mail?".

7.2. Conclusion

In this section a final evaluation of the extent to which the research objective was fulfilled is explored. We begin by revisiting the research objective:

To create a coding scheme which will enable the generation of SPT e-mails in a structured way and which will enable the analysis of the results of SPT engagements.

The first part of the objective requires that we make a coding scheme, something that has been achieved with the deliverable. We have managed to bring together variables from deception, persuasion and phishing-specific and coalesced them in a coding scheme which, to our opinion, encompasses the elements of the text which have an impact on the click-rate. The next part of the objective asked for the generation of e-mails in a structured way. We have achieved this as well since with the application of the coding scheme in the creation of a new e-mail, the maker has a clear set of defining variables out which they can choose. We have transformed the notion of making an e-mail more or less "phishy" into concrete variables. The phishing e-mail maker does not need to resort to his intuition to make a phishing e-mail but they can follow the set of scientifically backed variables. Thus we can say that the creation of new e-mails can be done in "a structured way". The second part of the research objective stated that we should enable the analysis of the results of the SPTs. To achieve this we have provided two things. First, the coding scheme which enumerates each e-mail with its defining variables. These enumerated variables can be correlated to the click-rate and we can identify the intrinsic differences between the e-mails. Second, is the experimentation guidelines that we have provided in the form of A/B testing. A/B experimentation will enable us to make conclusions on the actual impact of each variable on the click-rate.

7.3. Recommendations

To our knowledge, this work represents the first attempt to quantify the variables which define a phishing e-mail which has been sent as part of a SPT. More importantly, it is the first stepping stone in understanding what could make a phishing e-mail successful and rigorously distinguish the good from the bad phishing e-mails.

The coding scheme has two major uses for which recommendations are made. As requested in the research objective the need was to be able to create phishing e-mails scientifically and to measure them. In this section we will provide recommendations on how the coding scheme can be implemented to achieve the research objective.

The coding scheme provides a clear, decision-driven approach in the generation of SPTs. The phishing e-mail maker can follow and choose the value of each characteristics according to the context of the exercise. The context can be defined when the engagement is planned with the receiving party. In addition, using the informed assumptions table the e-mail can be assigned a set difficulty. If A/B testing is conducted then the actual contribution of each variable can be used and the resulting e-mail can have a prediction of click-rate in terms of percentages. As for the actual implementation of the coding scheme in a software package there are two approaches. For the deception variables, which are drawn from literature which tries to computerize deception detection, simple text analysis scripts can be made which can count the numbers of the deceptive variables. These variables do not constitute strictly a design decision but they can be used as such. In the case of A/B testing, for example, two e-mails of the

same length can be made with only different ratios of lexical diversity. It is recommended that the deception variables be measured only and not be used to guide the design of the e-mail as much. The values of these variables could be considered as research targets themselves and correlations can be made with the other human coded variables. These correlations can give new insights on how the deceptive variables can change as the persuasive elements of an e-mail change. For the persuasion variables coding is much more straightforward. A simple yes-no input can be created which the maker of the phishing e-mail has to answer. The answer will depend on the designer's decision to include or exclude a particular variable. This decision can be motivate by the consultations with the client or the desired difficulty of the e-mail.

This coding scheme also fills in a cog in the wider scheme of a successful phishing attack explained in Section 2.5.1. We have focused on the properties of the e-mail; with this coding scheme in hand we can now measure the susceptibility of individuals whilst having a formalized method of measuring the e-mail. In the future, one could make answer questions such as "What phishing e-mail would be more effective when sent to individuals who have high conscientiousness?" We now have knowledge of the elements of a phishing e-mail and we can, in the future, ask how these elements interact with the intrinsic characteristics of those receiving them. We could also incorporate other facts such as the time the e-mail was sent.

We could also move towards fully automated analysis and rating of phishing e-mails. If the coding scheme is used extensively, we will have a large dataset which has the values of the different persuasive principles in use. We could then perform a keyword analysis of all the e-mails which, for example, use authority and see what keywords represent the use of authority in an e-mail. We can make use of the human knowledge that is embedded in the e-mail to enable computers to detect the more latent parts of an e-mail.

7.4. Reflection

This thesis has been first and foremost a very involved learning experience. A new problem was presented and in order to tackle it many different fields had to be brought together and generate this final document. It began with many iterations on what the subject should be and it was essentially defined by Deloitte's kind provision of data. At first the focus on the individuals; the question was what makes *someone* susceptible to a phishing attack. There is considerable work done on the matter by others, however, in addition to the difficulties in finding subjects who would be willing and able to be subjected to testing. Focusing on the e-mail was definitely a new approach since, to our knowledge, there has not been a rigorous attempt to compare quantitatively between phishing e-mails. This comparison was the initial goal of the thesis where it was desired that a predictive model was generated which would be based on the data generated by previous SPTs. The inspection of the dataset has demonstrated that with the current amount of data that would be impossible. First, there was the heterogeneity of the data and second the small number of samples accentuated any extreme and bizarre data-points. Second, the conditions of each engagement were not accounted for which is necessary for the creation of high-quality scientific data. The proposals outlined in this project try to overcome these issues that arise and pave the way for meaningful inference on the results and causes of phishing click-rates.

Looking back at the thesis it is evident that themes from many diverse fields are taken. Apart from phishing specific literature we investigated deception, persuasion, motivation, emotion, we went all the way back to Aristotle and discussed leading-edge statistics. Various methodological tools were used in various sections of the thesis, from the definition of the research objective to the rules of the coding scheme. It has definitely been an arduous but extremely rewarding experience not so much from the context of the thesis but by the process. It required many iterations, a lot of feedback and changes but nonetheless a scientific document was produced.

During the making of this thesis the author was immersed in the concepts of deception and persuasion with persuasion being more prominent. As he delved more into the theory of persuasion, he began to understand the application of these techniques in everyday life and especially in non-phishing e-mails. The principles of persuasion are constantly applied in interactions between people that occur daily and mastery of such principles can prove very beneficial to the user. A word of caution however that is the content of someone's speech is the most important aspect. Using the principles to further one's own end without consideration of the consequences results in immoral outcomes. Examples of this are Hitler's masterful use of rhetoric or the persuasiveness of the Enron board in sustaining their deceptive operations.

Contribution to the Management of Technology Program.

This project is the last requirement to fulfill the obligations necessary to complete the Management of Technology (MOT) program at the Technical University of Delft, the Netherlands. To understand what the essence of MOT is, we will go through the description found on the program's website; the description which I read 2-and-a-half years ago and I knew that this was the right program for me.

We will quote a part of the website which best connects with this project, "Students that have completed the programme are well-versed to analyse technologies and their commercial impact and implement these in the organizational context of the firm."[7] Phishing is a phenomenon born out of technology; e-mail needs computers which in turn need electricity which in turn needs (or needed) steam engines. With this project we combined a wholly technology based construct (phishing) in the "organizational context of the firm". We investigated the nexus between technology and human communication, something that can undoubtedly be considered as a part of management.

A large number of courses were taught ranging from the management of people to the evolution of technology. This thesis explored the interaction between humans and technology as expressed in the phenomenon of phishing. Deception and persuasion are enduring elements of the human condition and phishing is resulted from the evolution of technology.

With this thesis, we further improve the capacity of a corporation to provide value to its customers. Simulated phishing tests can be done more informatively for the client whilst at the same time paving the way for scientific analysis in the future.

[7] http://www.tudelft.nl/studeren/masteropl/masteropleidingen/management-of-technology/

List of Figures.

Figure 1: The spear-phishing e-mail which compromised RSA. Source: wired.com 9
Figure 2: An ontological model for social engineering attacks by Mouton et al. [12] 16
Figure 3: The adapted ontological model for industry's SPTs. 17
Figure 4: Source-based model by Algarni et al. 20
Figure 5: Phase-based model developed by Algarni et al. 20
Figure 6: The various metrics which could predict the success of a social engineering attack by Workman (2008) 21
Figure 7: Berlo's Model of Communication. From http://communicationtheory.org/berlos-smcr-model-of-communication 23
Figure 8: The adapted SMCR model for Simulated Phishing Tests 24
Figure 9: The Q-Q plot of the click percentages of the current dataset. 38
Figure 10: The click rate groupings over the total number of recipients. 40
Figure 11: A SPT conducted to a banking institution 41
Figure 12: An SPT sent to a public body. 43
Figure 13: An SPT sent to Deloitte Belgium 44
Figure 14: An SPT sent to Deloitte Netherlands 45
Figure 15: Hoovering over the address reveals the true sender domain. 46
Figure 16: An SPT conducted to a Dutch client. 46
Figure 17: An SPT conducted to a Dutch client. 47
Figure 18: An example of membership functions in Fuzzy Logic. From Hameed and Sorensen. [56] 62
Figure 19: The A/B testing paradigm. 67
Figure 20: An example for A/B testing. 69
Figure 21: Bohn's table; the difference between Art and Science. [63] 73

Bibliography.

[1] K. J. Higgins, “Cybercrime, Cosa Nostra-Style,” *DarkReading*, 2008. [Online]. Available: http://www.darkreading.com/risk/cybercrime-cosa-nostra-style/d/d-id/1129602? [Accessed: 09-Feb-2015].

[2] I. S. Winkler and B. Dealy, “Information Security Technology?...Don’t Rely on It A Case Study in Social Engineering,” *Sci. Appl. Int. Corp.*, no. June, pp. 1–6, 1995.

[3] M. Bezuidenhout, F. Mouton, and H. S. Venter, “Social engineering attack detection model: SEADM,” *2010 Inf. Secur. South Africa*, pp. 1–8, 2010.

[4] F. Björck and L. Yngström, “IFIP World Computer Congress / SEC 2000 Revisited,” *WISE 2, Proc. IFIP TC11 WG 11.8 Second World Conf. Inf. Secur. Educ. held Perth July 12-14*, pp. 209–223, 2001.

[5] Z. Ramzan, “A Brief History of Phishing: Part I,” *Symantec Official Blog*, 2007. [Online]. Available: http://www.symantec.com/connect/blogs/brief-history-phishing-part-i. [Accessed: 02-Sep-2015].

[6] A. Algarni, Y. Xu, T. Chan, and Y.-C. Tian, “Social Engineering in Social Networking Sites : How Good Become Evil,” in *Proceedings of The 18th Pacific Asia Conference on Information Systems*, 2014.

[7] T. N. Jagatic, N. A. Johnson, M. Jakobsson, and F. Menczer, “Social phishing,” *Communications of the ACM*, vol. 50, no. 10. pp. 94–100, 2007.

[8] I. Fette, N. Sadeh, and A. Tomasic, “Learning to detect phishing emails,” *Proc. 16th Int. Conf. World Wide Web - WWW ’07*, p. 649, 2007.

[9] A. Walls and L. Akshay, “Innovation Insight for Anti-Phishing Behavior Management What You Need to Know,” 2015.

[10] K. Peffers, T. Tuunanen, C. E. Gengler, M. Rossi, W. Hui, V. Virtanen, and J. Bragge, “The Design Science Research Process: A Model for Producing and Presenting Information Systems Research,” in *the Proceedings of Design Research in Information Systems and Technology DESRIST’06*, 2006, vol. 24, pp. 83–106.

[11] J. Saldana, *The coding manual for qualitative researchers. No. 14*. Sage, 2012.

[12] U. S. Odaro and B. G. Sanders, “Social Engineering : Phishing for a Solution,” *Adv. Commun.*

Comput. Networks Secur., vol. 8, no. 88, 2009.

[13] K. D. Mitnick and W. L. Simon, "The Art of Deception: Controlling the Human Element in Security," *BMJ Br. Med. J.*, p. 368, 2003.

[14] T. R. Gruber, "Technical Report KSL 92-71 Revised April 1993 A Translation Approach to Portable Ontology Specifications by A Translation Approach to Portable Ontology Specifications," *Knowl. Creat. Diffus. Util.*, vol. 5, no. April, pp. 199–220, 1993.

[15] F. Mouton, L. Leenen, M. M. Malan, and H. S. Venter, "Towards an Ontological Model Defining," pp. 266–279.

[16] E. E. H. Lastdrager, "Achieving a consensual definition of phishing based on a systematic review of the literature," *Crime Sci.*, vol. 3, no. 1, pp. 1–10, 2014.

[17] FireEye, "Spear Phishing Attacks — Why They are Successful and How to Stop Them Combating the Attack of Choice for Cybercriminals," 2012.

[18] Cisco, "Email Attacks : This Time It ' s Personal," 2011.

[19] S. Abraham and I. Chengalur-Smith, "An overview of social engineering malware: Trends, tactics, and implications," *Technol. Soc.*, vol. 32, no. 3, pp. 183–196, 2010.

[20] B. Parmar, "Protecting against spear-phishing," *Comput. Fraud Secur.*, vol. 2012, no. 1, pp. 8–11, 2012.

[21] A. Algarni and Y. Xu, "Social engineering in social networking sites: phase-based and source-based models.," *Int. J. e-Education, e-Business, e-Management e-Learning*, vol. 3, no. 6, pp. 456–462, 2013.

[22] A. Algarni, Y. Xu, and T. Chan, "Social Engineering in Social Networking Sites : The Art of Impersonation," 2014.

[23] "The devil is in the details : Social Engineering by means of Social Media Research Project," 2012.

[24] J. Wang, T. Herath, R. Chen, A. Vishwanath, and H. R. Rao, "Research article phishing susceptibility: An investigation into the processing of a targeted spear phishing email," *IEEE Trans. Prof. Commun.*, vol. 55, no. 4, pp. 345–362, 2012.

[25] M. Jakobsson, A. Tsow, A. Shah, E. Blevis, and Y. K. Lim, “What instills trust? A qualitative study of phishing,” *Proc. 11th Int. Conf. Financ. Cryptogr. 1st Int. Conf. Usable Secur.*, pp. 356–361, 2007.

[26] M. Workman, “Wisecrackers: A theory-grounded investigation of phishing and pretext social engineering threats to information security,” *J. Am. Soc. Inf. Sci. Technol.*, vol. 59, no. 4, pp. 662–674, 2008.

[27] Deloitte, “Stop Phishing Attacks,” London, UK, 2014.

[28] A. Vishwanath, T. Herath, R. Chen, J. Wang, and H. R. Rao, “Why do people get phished? Testing individual differences in phishing vulnerability within an integrated, information processing model,” *Decis. Support Syst.*, vol. 51, no. 3, pp. 576–586, 2011.

[29] S. Sheng, M. Holbrook, P. Kumaraguru, L. F. Cranor, and J. Downs, “Who falls for phish? A Demographic Analysis of Phishing Susceptibility and Effectiveness of Interventions,” *Proc. 28th Int. Conf. Hum. factors Comput. Syst. - CHI ’10*, pp. 373 – 382, 2010.

[30] D. K. Berlo, *The process of communication: an introduction to theory and practice*. Fundo de Cultura, 1960.

[31] J. E. Mahon, “Kant on lies, candour and reticence.,” *Kantian Rev.*, vol. 7, pp. 102–133, 2003.

[32] B. M. DePaulo, J. J. Lindsay, B. E. Malone, L. Muhlenbruck, K. Charlton, and H. Cooper, “Cues to deception,” *Psychol. Bull.*, vol. 129, no. 1, pp. 74–118, 2003.

[33] J. K. Burgoon, J. P. Blair, and R. E. Strom, “Cognitive biases and nonverbal cue availability in detecting deception,” *Hum. Commun. Res.*, vol. 34, no. 4, pp. 572–599, 2008.

[34] R. . Petty and J. . Cacioppo, “The elaboration likelihood model of persuasion,” vol. 19, pp. 1–24, 1986.

[35] C. E. Shannon and W. Weaver, *The mathematical theory of communication*. University of Illinois press, 2015.

[36] R. L. Ackoff, “Towards a Behavioral Theory of Communication,” *Manage. Sci.*, vol. 4, no. 3, pp. 218–234, 1958.

[37] T. Liqing, “What Do We Know About Email - An Existing and Emerging Literacy Vehicle?,” in *National Reading Conference*, 1995.

[38] L. Zhou, J. K. Burgoon, J. F. Nunamaker, and D. Twitchell, "Automating Linguistics-Based Cues for detecting deception in text-based asynchronous computer-mediated communication," *Gr. Decis. Negot.*, vol. 13, no. Cmc, pp. 81–106, 2004.

[39] P. S. Keila and D. B. Skillicorn, "Detecting Unusual and Deceptive Communication in Email," *Proc. 2005 Conf. Cent. Adv. Stud. Collab. Res.*, no. June, pp. 17–20, 2005.

[40] S. Grazioli, "Where did they go wrong? An analysis of the failure of knowledgeable Internet consumers to detect deception over the internet," *Gr. Decis. Negot.*, vol. 13, pp. 149–172, 2004.

[41] T. Q. T. Qi, "An Investigation of Heuristics of Human Judgment in Detecting Deception and Potential Implications in Countering Social Engineering," *2007 IEEE Intell. Secur. Informatics*, no. September, 2007.

[42] L. Zhou, D. P. Twitchell, T. Q. T. Qin, J. K. Burgoon, and J. Nunamaker, J.F., "An exploratory study into deception detection in text-based computer-mediated communication," *36th Annu. Hawaii Int. Conf. Syst. Sci. 2003. Proc.*, vol. 00, no. C, pp. 1–10, 2003.

[43] M. Friestad and P. Wright, "The Knowledge Persuasion Model: How People Cope with Persuasion Attempts," *J. Consum. Res.*, vol. 21, no. 1, pp. 1–31, 1994.

[44] C. Rapp, "Aristotle's Rhetoric," *The Stanford Encyclopedia of Philosophy*, 2010. [Online]. Available: http://plato.stanford.edu/archives/spr2010/entries/aristotle-rhetoric/. [Accessed: 04-Sep-2015].

[45] R. B. Cialdini, *Influence*. A. Michel, 1987.

[46] N. Akbar, "Analyzing Persuasion Principles in Phishing Emails," 2014.

[47] K. J. Lee and I. Song, "Investigating Information Structure of Phishing Emails Based on Persuasive Communication Perspective," *J. Digit. Forensics, Secur. Law*, vol. 2, no. 3, pp. 29–44, 2007.

[48] D. Kim and J. H. Kim, "Understanding persuasive elements in phishing e-mails: A categorical content and semantic network analysis," *Online Inf. Rev.*, vol. 37, pp. 835–850, 2013.

[49] B. Weiner, "An attributional theory of achievement motivation and emotion.," *Psychol. Rev.*, vol. 92, no. 4, pp. 548–573, 1985.

[50] R. Buck, “Prime theory: An integrated view of motivation and emotion.,” *Psychol. Rev.*, vol. 92, no. 3, pp. 389–413, 1985.

[51] A. H. Maslow, “A theory of human motivation.,” *Psychol. Rev.*, vol. 50, no. 4, pp. 370–396, 1943.

[52] L. Rourke, T. Anderson, D. R. Garrison, and W. Archer, “Methodological Issues in the Content Analysis of Computer Conference Transcripts,” *Int. J. Artif. Intell. Educ.*, vol. 12, no. July 2000, pp. 8–22, 2001.

[53] D. McCloskey, “How to do a rhetorical analysis, and why.,” *New Dir. Econ. Methodol.*, pp. 319–342, 1994.

[54] T. Hak and T. Bernts, “Coder training: Theoretical training or practical socialization?,” *Qualitative Sociology*, vol. 19, no. 2. pp. 235–257, 1996.

[55] K. A. Ericsson and H. a Simon, “Verbal reports on thinking,” *Introsp. Second langauge Res.*, pp. 24–53, 1987.

[56] G. Watson, A. Mason, and R. Ackroyd, *Social Engineering Penetration Testing: Executing Social Engineering Pen Tests, Assessments and Defense.* Syngress, 2014.

[57] D. A. Freedman, *Statistical Models: Theory and Practice*. Cambridge University Press, 2009.

[58] G. Casella, S. Fienberg, I. Olkin, and Tiribashi, *An Introduction to Statistical Learning*, vol. 102. 2006.

[59] D. Freedman, “Statistical models and shoe leather,” *Mathematical Social Sciences*, vol. 18, no. 2. p. 192, 1989.

[60] P. L. Flom and D. L. Cassell, “Stopping Stepwise,” in *NESUG 2007*, 2007, pp. 1–7.

[61] L. A. Zadeh, “Fuzzy sets.,” *Inf. Control*, vol. 8, no. 3, pp. 338–353, 1965.

[62] L. Zhou and A. Zenebe, “Representation and reasoning under uncertainty in deception detection: A neuro-fuzzy approach,” *IEEE Trans. Fuzzy Syst.*, vol. 16, no. 2, pp. 442–454, 2008.

[63] P. Peduzzi, J. Concato, E. Kemper, T. R. Holford, and A. R. Feinstem, “A simulation study of the number of events per variable in logistic regression analysis,” *J. Clin. Epidemiol.*, vol. 49, no. 12,

pp. 1373–1379, 1996.

[64] J. Concato, A. R. Feinstein, and T. R. Holford, “The risk of determining risk with multivariable models,” *Ann. Intern. Med.*, vol. 118, no. 3, pp. 201–210, 1993.

[65] J. Khan, J. S. Wei, M. Ringnér, L. H. Saal, M. Ladanyi, F. Westermann, F. Berthold, M. Schwab, C. R. Antonescu, C. Peterson, and P. S. Meltzer, “Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks.,” *Nat. Med.*, vol. 7, no. 6, pp. 673–679, 2001.

[66] A. Deng, Y. Xu, R. Kohavi, and T. Walker, “Improving the sensitivity of online controlled experiments by utilizing pre-experiment data,” *Proc. sixth ACM Int. Conf. Web search data Min. - WSDM ’13*, no. FEBRUARY, p. 123, 2013.

[67] T. Student, “The probable error of a mean,” *Biometrika*, vol. 6, no. 1, pp. 1–25, 1908.

[68] T. Lowdermilk, *User-Centered Design: A Developer’s Guide to Building User-Friendly Applications*. O’Reilly Media, Inc., 2013.

[69] A. S. Reber, *The penguin dictionary of psychology*, 2nd ed. Toronto: Penguin Books, 1995.

[70] R. E. Bohn, “From Art to Science in Manufacturing : The Evolution of Technological Knowledge,” *Found. Trends Technol. Information, Oper. Manag.*, vol. 1, no. 2, pp. 1–82, 2005.