An Information-Theoretic investigation of cheating in traditional examinations

Submitted in partial fulfilment

of the requirements of the degree of

Bachelor of Science (Honours)

of Rhodes University

Greg Pennefather

Supervisor: Yusuf Motara

Grahamstown, South Africa

November 1, 2013

Abstract

An investigation of the cheating problem was performed, using an Information-theoretic approach. This approach aimed to discover the applicability of Information Theory, Steganography, and definitions of knowledge in understanding the problem. From this understanding the three categories of techniques for solving the problem were identified, and basic techniques for two categories were created. The process of creation identified the lack of applicability of Modern Steganography and the potential for further research into the use of Information Theory. Limitations of techniques that may aid in prevention of cheating were identified.

ACM Computing Classification System Classification

Thesis classification under the ACM Computing Classification System (1998 version, valid through 2013:

H.1.1 [Systems and Information Theory]: Information Theory

General-Terms: Theory, Human Factors

Acknowledgements

I would like to thank my supervisor, Mr Yusuf Motara for his guidance and support for this project.

I would like to acknowledge the financial and technical support of Telkom, Tellabs, Stortech, Genband, Easttel, Bright Ideas 39 and THRIP through the Telkom Centre of Excellence in the Department of Computer Science at Rhodes University.

I would also like to acknowledge the support and encouragement offered throughout the year by my parents, Reginald and Allyson Pennefather.

Contents

List of Figures

Chapter 1

Introduction

1.1 Motivation

Academic integrity is the moral code that controls the conduct of those involved in academia, committing them to honesty and fairness. An institute that lacks integrity will develop a reputation for dishonesty and works produced by that university will likely be distrusted by academics from other institutions. This principle extends to the production of educated minds by an institution. In order to ensure a positive reputation an institution must ensure that all graduates deserve their degrees. To achieve this the institution must enforce academic integrity amongst its candidates.

The act of skewing assessment in one's favour, hereforth called cheating, is in direct conflict with academic integrity. Cheating in a traditional examination (where written questions require written answers) is defined as access to information that is external to a candidates knowledge. Accessing this external information despite the restrictions of the examination environment is referred to as the cheating problem.

A cheater is attempting to receive that which they do not deserve through dishonesty. This makes cheating a concern of all academic institutes and considerable research has been performed with the aim of understanding cheating. Research attempts to answer

questions similar to 'Why do students cheat?, 'How do students justify cheating morally?' and 'Is cheating socially acceptable?' by performing surveys and interviews with students. While these questions should be asked, they do not question or study the techniques used by students to solve the cheating problem. An investigation of cheating that attempts to classify, understand, and identify their limitations of cheating is non-existent.

1.2 Research Goal

This paper will explore cheating techniques using knowledge from the fields of Steganography, the science of hiding information, and Information Theory, which is based on the work of mathematicians Shannon & Weaver (1949). It aims to discover the extent to which the knowledge of those two fields can be applied to the cheating problem, and what understanding we can gain about cheating techniques and their limitations because of this.

The primary aim of this paper is to understand cheating, and develop basic techniques for solving the cheating problem. Understanding should provide insight into the limitations of cheating techniques, and the relevance of information-theoretic concepts to cheating and other similar problems.

The following questions are of particular interest to this work:

- what defines cheating?
- how can we categorise cheating techniques?
- to what extent can concepts from the field of Information Theory be applied to these category?
- to what extent can concepts from the field of Steganography be applied to these category?
- what are the limitations of techniques in each category?

1.3 Methodology

Chapter 2 will begin with a review of existing research, focusing on the epistemological side of assessment, the psychological side of cheating and the examination environment. From there it will review research on the topics of Information theory and Steganography.

Research into assessment will attempt to establish a definition of data, information and knowledge that will be used throughout this paper. These definitions are of great importance to both assessment and the information-theoretic fields of Information Theory and Steganography. In order to establish a current set of definitions literature from both the fields of Information Systems and Modern Philosophy will be reviewed.

An understanding of the environment surrounding cheating will provided through reviews of literature and examination procedures. Literature and surveys that focus on the moral and psychological factors of cheating will be used to establish the motivations of a candidate. These motivations will help guide examples and theories in Chapter 3. An exploration of examination environment guidelines will outline restrictions that will need to be considered throughout the paper.

A review of the relevant literature will provide insight into the concepts and definitions used in Information Theory. This insight will be used to analyse the cheating problem, and develop basic techniques for solving the problem. Basic information theory is also present in the final field being reviewed, Steganography. The concepts and definitions of Steganography will be outlined for later use in the same way as those of Information Theory.

In Chapter 3 this paper will begin by defining cheating in terms of epistemological definition of information and knowledge. From this definition it will attempt to group cheating techniques into different categories, which will be explored individually at the end of the chapter. The relevant categories, Information Smuggling and Communicated Information, will be explored through creation of basic cheating systems. It will also explore the concepts of motivation, selection of cheating technique, representation of information,

knowledge internalisation and provide an example of a typical cheating candidate that will be used throughout the chapter.

In Chapter 4 we will conclude by determining the feasibility of applying information theoretic practices to cheating, and touch on the impact of the limitations that may have been revealed by the investigation.

Chapter 2

Related Work

2.1 Data, information and knowledge

Data, information and knowledge form the building blocks for all cognitive content. They are concepts that form a fundamental part of every human mind, even if they are not consciously considered. In academia, where the workings of the human mind and the representation of facts and opinion is important, different definitions of these concepts can be found depending on who is being asked, making it a very subjective topic for discussion. It has been discussed since the time of early Greek philosophers such as Plato and Aristotle, and forms a key part of Modern Philosophy. Another field with an interest in these concepts is the field of Information Systems. While the majority of information theorists agree on the general definition of each concept, they often disagree on the finer details of each. The views of both philosophers and information theorists will be taken into consideration when defining data, information and knowledge as they will be used throughout this paper.

2.1.1 Definitions from Information Theorists

The general consensus regarding the content that can be stored in the human mind, breaks it into three categories: data, information, knowledge; with some sources defining

understanding and wisdom in addition to this (Ackoff, 1989, Zins, 2007). However sources differ on the exact definition of each category. This section will explore the different definitions offered by information theorists, and will state the definition that will be used in this paper.

The work of Ackoff

Ackoff (1989), an information theorist, categorises cognitive content using following hierarchy: data, information, knowledge, understanding, wisdom. He approached the definitions from the position of an Information Systems Manager concerned with the creation of Expert and Understanding Systems. He believed that each category was reliant on the category that came before it.

He perceived data as the basic form of mental content (Ackoff, 1989). It is the product of observation and has no value by itself. Data only becomes usable when processed, often through reduction and simplification, to create information (Ackoff, 1989). Information is therefore the functional form of a collection of data, containing descriptions such as what, who, where, when and how many (Ackoff, 1989). Knowledge is know-how required to act upon information. Ackoff argues that knowledge allows someone to control a system, and in doing so increase its efficiency. Knowledge is generated through learning, either in the form of instruction or personal experience. The ability of an individual to learn is that individual's intelligence. Understanding is the ability to synthesize new knowledge or information from existing knowledge (Ackoff, 1989, Bellinger *et al.*, 2004). Wisdom is less relevant to this paper, but explores a higher level of understanding that deals with the future and allows for an increase in effectiveness. It is a uniquely human process for extrapolating non-deterministic, non-probabilistic answers to questions (Ackoff, 1989, Bellinger *et al.*, 2004).

Results of a survey of Information Systems Scholars

Zins (2007) performed a study seeking to map the conceptual approaches for defining

data, information and knowledge in the field of Information Systems (IS). His motivation for this study was the need for the regular reviewing of definitions in the IS field owing to its constantly changing nature. Zins employed the Critical Delphi methodology which aims to facilitate critical, moderated discussion among experts on a panel of 57 leading scholars in the field. Of the 57 panel members, 44 offered definitions of data, information and wisdom. What follows is a collection of classifications for each term, reinforced by quotations from the respondents. Zins defined classifications of his own, based on a number of different characteristics. These classifications were not relevant as they did not define each concept individually, but rather listed the characteristics that applied to each concept.

Three repeated descriptions of data could be identified in the responses. The first characterises data as static content, describing it as "coded invariances" or something "which is stated" (Albrechtsen). The second describes the means or source it is acquired from, stating that it "refer(s) to statistical observations" (Davis), is the product of "sensory stimuli that we perceive through our senses" (Baruchson-Arbib), or is "the raw' material obtained from observation" (Ekbia). The third description is of data as a member of a symbol set, forming part of "a symbol set that is quantified and/or qualified" (Barreto), "a set of symbols representing a perception of raw facts" (Dragulanescu). (Zins, 2007)

Information is also described as being part of a symbol set, however these symbols represent knowledge rather than raw observations, "[information] is a set of symbols that represent knowledge" (Wormell) or "represented knowledge is information" (Childers) or "the act of communicating knowledge" (Oxford English Dictionary). Information comes from the Latin *informatio* meaning "to give a form". From this we can describe information as something that allows for the creation of knowledge inside the mind of an individual; "a set of significant signs that has the ability to create knowledge" (Wersig and Neveling, 1975). The ability to create knowledge implies that information has meaning associated with it. Other respondents note that information can be created by adding meaning to data through the act of processing it. "Collocations of data that thereby become meaningful to human beings" (Ess) or "data that has been processed into a form that is meaningful to the recipient". (Zins, 2007)

Knowledge "exists in the mind of the knower" (Childers) and reflects the knower's ability to find "definition in meaning and understanding" (Debons). Furthermore, as Herold said "It is heavily internally orientated, understood completely only by the person possessing it". Knowledge is created in the mind of the knower as a product of appropriated information (Barreto). It "emerges from analysis, reflection upon, and synthesis of information" (Hawkins) and "is by definition subjective" (Childers). Knowledge makes the processing and utilisation of information possible, and Fidel states that knowledge is a "personal framework that makes it possible for humans to use information", while Tenopir stated it "can be used to make decisions". (Zins, 2007)

2.1.2 Definitions from Modern Philosophy

Philosophers approach knowledge in a more abstract way than Information Theorists. Rather than focusing on definitions of different concepts, philosophers are concerned with the nature and origin of knowledge. This branch of philosophy is referred to as epistemology. While there are many branches of epistemology that deal with different theories of knowledge, Rationalism and Empiricism are two branches that specifically deal with the origin and nature of knowledge.

Rationalism

Rationalism is rooted in the work of the early Greek philosopher Plato. He argued that when something is experienced in the physical world, that experience simply reminded us of something that we had always known. Everything that is experienced (love, pain, beauty) was already present in our minds before we experienced it; we simply needed an external influence to recall it. This is referred to as Plato's Theory of Forms and is commonly explained using the cave metaphor. Plato believed that all things experienced were the shadows of the most perfect form of that thing, which exists outside what we can experience, much like the shadow cast onto the walls of a cave by an object outside of a cave. (Schunk, 2003)

This led to a belief in a mind-matter dualism. Plato believed that the world simply consisted of raw matter, with information existing solely in the brain of an individual. The act of absorbing sense data from raw matter allows for the recall of information that exists in the brain.(Schunk, 2003)

The theory of mind-matter dualism was expanded on by the work of the German philosopher Kant. He argued that the world consisted of a large amount of disorganised data, and that the brain organised that data into information. This implies that we cannot be sure of anything about the world based on the sense data we receive, we can only know that this is how we perceive the world. Thus the only true knowledge comes from reasoning, as anything based on sensed information may be influenced by misperception. This idea originated from French philosopher René Descartes, who stated that the only thing that can be proved is one's own existence, leading to his well-known phrase I think therefore I am, in his work *Meditation on First Philosophy* (Descartes, 1967).

The work of Kant and Descartes have resulted in Rationalists in Modern Philosophy moving away from Plato's Theory of Forms. They agree that knowledge is created by the mind and limited to the mind as sense information can potentially be untrustworthy.

Empiricism

Aristotle, a student of Plato, disagreed with his theory of forms. He believed that since the perfect form of an experience could not exist in the physical world, that it was not perfect. Furthermore he believed that knowledge was part of a physical thing and that the two concepts could not be separated. Rather, sensory information is the source of knowledge, and it is not subject to change. However, his views were not as popular as those of Plato. Plato's theory went largely unopposed until the work of English political philosopher John Locke.

Locke disagreed with Plato, arguing that there was no innate knowledge and that the mind was a blank slate, or *tabula rasa*, upon birth (Petryszak, 1981). While Plato insisted that reasoning alone was enough to create knowledge, Locke stated that it was a product

of experience, which came in two forms: sensory information and personal awareness. Knowledge may be refined and developed through personal awareness alone, but for that knowledge to exist originally it must have been acquired from the physical world through sensory information. Thus any idea regardless of complexity originated from a simple piece of knowledge acquired from the real world. These simple pieces of knowledge are associated with one another and developed through personal awareness in order to form a complex piece of knowledge. Thus all knowledge can be broken down atomically. (Schunk, 2003)

Both branches believe that knowledge is a part of the mind, though they disagree on the origin of this knowledge. This paper will deal with Information that has been constructed from knowledge, and as such is less concerned with the origin on knowledge.

2.1.3 Definitions used in this paper

For the purposes of this paper data, information and knowledge will be defined in the following ways.

Data is raw content obtained from real world observations that form part of a symbol set. This means that data

- is unprocessed and has no context, resulting in an absence of meaning
- is capable of being communicated and recorded in symbolic form

Information is the meaningful representation of knowledge, created either by the possessor of the knowledge or the act of processing raw data. This means that information

- is capable of being communicated and recorded
- is the bridge between raw data and internal knowledge
- has meaning to human beings

Knowledge is the collection of the internal, subjective understanding and meaning of the knower, that can be used functionally in decision making and the creation of new knowledge. This means knowledge

- is subjective and exists in the mind of the knower
- is a collection of the understanding and meaning possessed by the knower

2.2 Classification of Assessment

Becker & Trowler (1989) created a framework for grouping disciplines by their style of assessment and type of knowledge. Groupings categorise the type of knowledge as hard or soft, and the use of the knowledge as pure or applied. The style of assessment for a discipline is relevant when considering cheating during an examination, as different styles will require different preparations. This section will explain the aims, processes of assessment and type of knowledge for each grouping.

2.2.1 Hard disciplines

The knowledge of a hard discipline is composed of cumulative, atomic units of knowledge (Becker & Trowler, 1989). Hard disciplines are quantitative in nature, with examples including Physics, Mathematics and Engineering (Becher, 1994). Education of candidates begins with the fundamental basics of the field that are expanded upon with more advanced concepts. For example, in Physics, equations of motion are taught in early in the education process and then expanded on by Einstein's theory of relativity. The two subjects are taught independently of each other, but the ability to understand relativity is dependent on the prerequisite understanding of equations of motion.

A hard pure discipline seeks to create new knowledge through research. The research environment is competitive, but research is often performed gregariously by a number of

authors. New research builds upon previous research, and usually attempts to simplify the representation of, and create universal rules for, a problem. Hard applied disciplines are built upon the knowledge generated by hard pure disciplines. Knowledge is applied to the physical world, and used in the development of techniques and products that solve practical problems. (Becker & Trowler, 1989)

The assessment of a hard pure discipline aims to ascertain the extent to which a candidate has acquired a piece of knowledge. Hard applied disciplines aim to test the extent to which a candidate is able to apply a piece of knowledge. To achieve this, examination questions are specific, unambiguous and focused on a particular section of knowledge. Testing is objective and requires fewer safeguards for unbiased assessment, that would make assessment laborious. This allows for regular assessment, used to test the acquisition of each cumulative block of knowledge before progressing to the next. Candidates tend to require the ability to retain more knowledge and have a greater aptitude for problem solving than those in a soft pure discipline.

2.2.2 Soft disciplines

In contrast to the hard disciplines, soft disciplines are qualitative in nature. Knowledge is holistic, biased and not superseded by subsequent work. Examples of soft disciplines include Philosophy, English Literature and Management; with the latter being an example of a soft applied discipline (Becher, 1994). Soft pure disciplines often focus on the study of a single work, such as *Nicomachean Ethics* by Aristotle. Individual works and concepts, while not completely independent of others in the discipline, can often be studied without prerequisite knowledge. (Becker & Trowler, 1989)

A pure soft discipline, much like a pure hard discipline, also focuses on enquiry (Becker & Trowler, 1989). However it does so primarily through reiterative research that refines knowledge over time in addition to the creation of new research. Research also tends to take place independent of other research, with very limited co-authorship on new material (Becker & Trowler, 1989).

The assessment of pure soft subjects makes use of broad essay-style questions that test the holistic knowledge of a candidate and their ability to argue a particular theory, substantiating their argument with evidence from the studied text. This element of personal opinion means that there is no singular correct answer to a question. In this case marking is largely impressionistic, with examiners following assessment guidelines to ensure that candidates are marked similarly. Multiple examiners are used to grade candidates in order to ensure consistent assessment. The result of this is an increased workload for everyone involved in assessment. Coupled with hard pure discipline's focus on the entire scope of the subject, this means that assessment only takes place once or twice over the course of a subject. Soft applied disciplines build on the principle of developing a candidate and their understandings. They focus on enhancing a candidate and developing their skills. This is done through skills demonstrations that often have vague assessment guidelines (Toohey, 1999).

A candidate in a soft discipline must have the ability to formulate their own opinions and express them in an appropriate manner. They require skills such as prose and writing coherence, and many times focus on the degree of refinement or elegance in the opinion and expression of opinion of a candidate (Becker & Trowler, 1989).

This paper will only be concerned with the assessment of pure disciplines, as the assessment of applied disciplines does not typically take place in a traditional examination environment.

2.3 Survey of cheating in Academia

Western Universities have been aware of cheating and academic dishonesty for sometime (Bjorklund & Wenestam, 2000). It is a problem that is hard to tackle owing to its deceptive nature. In recent decades there have been a large number of different new evaluation methods put in place with the aim of preventing cheating (Joyce, 2002). These can be broken up into three distinct groupings: preventing access to information during evaluation, evaluating those who share information equally and an increased accountability in

the integrity of academic work. Prevention of access to information attempts to cut down on information sharing though individual examinations. These examinations can come in the form of closed book, open book and oral examinations in which students are prevented from communicating with each other for the duration of the examination (Joyce, 2002). Group evaluations with larger workloads allow for those who share information to be given an equal grade (Joyce, 2002). Finally the use of plagiarism warnings and 'own work' declarations discourage cheating by making students aware of the consequences and repercussions of being caught (Joyce, 2002).

However this has not managed to discourage students from cheating. Many students believe that cheating is socially acceptable (Bjorklund & Wenestam, 2000) and do not discourage or report their peers whom they find cheating (Lim & See, 2001). More concerning is the number of academic supervisors who have ignored evidence of cheating (Bjorklund & Wenestam, 2000). Many supervisors choose not to do so because of the discomfort caused by reporting them to a university authority. Rather supervisors attempt to handle the students personally without involving the university (Bjorklund & Wenestam, 2000).

Current studies in the field tend to focus on the moral and social issues surrounding cheating. They do not explore the way that students cheat or when students cheat. Instead they try to explain why paticular students are cheating.

2.3.1 Previous Studies

A Swedish-Finnish university performed a study that attempted to discover the frequency of confessed cheating, the most common kinds of cheating, what the relation between cheating and gender is, and how their results related to British results (Bjorklund & Wenestam, 2000).

The study distinguished between four types of cheating behaviors:

- individual opportunistic cheating

- individual planned cheating
- active social cheating
- passive social cheating

2.3.2 Frequency

Students participating in the study were given a questionnaire containing a list of different cheating methods and asked to mark off which they had engaged in. The list contained 23 different methods that could be used under different circumstances including coursework, research and examination environments. The study also classified the methods on the list as either social or individual cheating and had a special classification for altruistic cheating (Bjorklund & Wenestam, 2000). The results of the questionnaire revealed that three quarters of students had engaged in at least one of the methods on the list (Bjorklund & Wenestam, 2000). However in a final question that asked whether they felt they had ever cheated overall only 63.5% of students felt they had (Bjorklund & Wenestam, 2000). This again indicates that students who engage in academic dishonesty sometimes do not believe that they have done anything wrong (Bjorklund & Wenestam, 2000).

2.3.3 Methods

The results of the questionnaire identify the following methods:

- copying during an exam
- illicitly gaining advance information about the contents of an examination paper
- taking unauthorised material into an examination (e.g. ‘cribs’)
- premeditated collusion between 2 or more students to communicate answers to each other during an examination

- lying about medical or other circumstances to get special consideration by examiner
- taking an examination for someone else or having someone else take an examination for you

2.4 Examinations at Rhodes University

This section reviews the examination enviroment utilised at Rhodes University. This will be used as an example of what can be reasonably expected when considering an examination enviroment.

2.4.1 Examination Initiation

An examination is run by an examination commissioner, who must arrive 15 minutes before the initiation of an examination session (Rhodes Academic Administration, 2013). The commissioner is assisted by examiners in the distribution of papers and the seating of candidates (Rhodes Academic Administration, 2013). Candidate seating is prearranged, with question papers distributed to the tables at which candidates will write the examination.

The examination must be initiated by the examination commissioner, but thereafter it is the responsibility of the examiners to collect attendance slips and invigilate the examination (Rhodes Academic Administration, 2013). They maintain this responsibility until the examination commissioner returns and ends the examination, their responsibility is then to supervise the collection of scripts by the relevant Departmental representatives.

The prearranged seating does not explicitly follow a set of guidelines, but the norm at Rhodes University is to use to be an alphabetical order.

2.4.2 Superintendence

Official guidelines of the duties of an examiner in an examination venue dictate that an examiner 'should give their undivided attention to superintendence and patrol from time to time so that opportunities do not exist for infringement of regulations' (Rhodes Academic Administration, 2013). There is no further expansion on this point before listing a set of actions to follow should the examiner be 'satisfied that a candidate may be liable for disqualification' (Rhodes Academic Administration, 2013). Liability for disqualification is found through violations of the Student Disciplinary Code (Rhodes Academic Administration, 2013). The guideline states that the purpose of this is simply to remove opportunities for the candidate to cheat, presuming that supervision will remove opportunity.

These guidelines presume that an examiner is informed enough on cheating techniques to become aware of a candidate who may be cheating. In practice, noticing conspicuous behaviour is a challenging owing to a number of factors:

- detection from a distance is obfuscated by fidgeting and the human tendency to make small movements.
- patrolling is noticeable and gives candidates forewarning of supervision.
- conspicuous behaviour is only conspicuous if one is aware of what to look for.
- some forms of communication may be near undetectable without express knowledge of the channel of communication and/or the language.

The guidelines provide no information for overcoming these factors.

2.4.3 Malpractices

Malpractices are the actions that make a candidate liable for disqualification from an examination (Rhodes Academic Administration, 2013). As previously mentioned, liability

is in accordance with the Student Disciplinary Code. In the event of malpractice examiners are instructed to do the following:

- confiscate incriminating material
- confiscate the answer book of the candidate and record the time
- provide the candidate with a new answer book, record the time and allocate the candidate no additional time.
- inform the candidate that the incident will be reported
- expel the candidate from the venue on repeated offence.

The guidelines provided by Rhodes University are primarily concerned with the administration surrounding the commencement and completion of an examination, and the role of the examiner. Guidelines do not specify the exact operations of an examiner, past being required to be required to attempt to prevent opportunities, and rely on the examiners decision making to detect infringements. A typical traditional examination environment is therefore considered one where access to external information is restricted by the discretion of examiners.

2.5 Information Theory

Information theory is a subcategory of communication theory. Communication theory derives from the work of Shannon & Weaver (1949) in their paper *A Mathematical Theory of Communication*. It encapsulates a collection of theories from the fields of physics, mathematics, electrical engineering and computer science (Cover *et al.*, 1994), and was created to provide a deeper understanding of communication and the problems associated with it (Shannon & Weaver, 1949).

Information theory itself is concerned with the quantification of information. It aims to introduce data compression and error correction to information representation (Murphy,

1998). This chapter will first discuss these concepts on a conceptual level, before covering the statistical proofs and methods upon which they are based. Prior to this the basic components and considerations of a communication system will be discussed.

2.5.1 Components of an Communication System

According to Shannon & Weaver (1949) a communication system is made up of the following five essential parts, arranged as shown in Figure 2.1:

- An information source that produces a message or a sequence of messages that are to be communicated to the destination. An example of a message is someone speaking into the microphone on a telephone.
- A transmitter that converts the message from the information source into a form that is suitable for transmission over the desired channel.
- A channel which is a medium for transmitting the signal form of the message being sent. This medium is subject to noise, which can occur in a number of forms depending on the medium. For example, copper wires experience noise in the form of electromagnetic fields.
- A receiver that reconstructs the message from the signal transmitted over the medium.
- The destination where the message was intended to be sent.

2.5.2 Defining Messages

A message is the unit of communication. In the context of Information Theory, the term 'information' applies to a sequential collection of symbols that can be identified as a message. A message is thus a container for the information being communicated. This definition can be consolidated with information theorists' definitions, as sequential

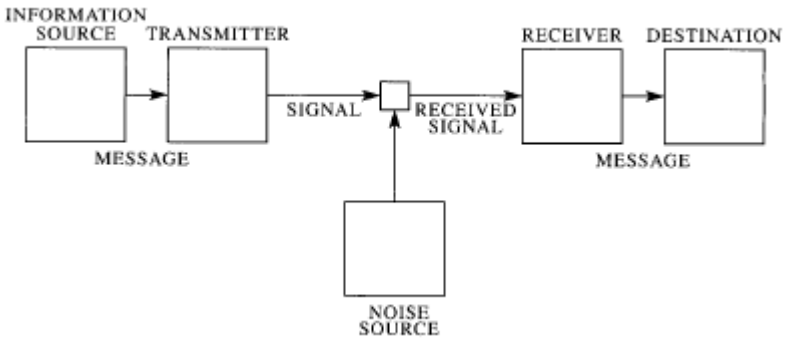

Figure 2.1: Schematic diagram of a general communication system (taken from Shannon & Weaver, 1949, p. 4)

symbols imply a relationship between those symbols, which in turn implies meaning. The symbols contained in a message are data units. The collection of these symbols in sequential order transforms them into information. The term "message" is largely synonymous with the term "information" in this context.

While it can be said that messages have meaning, Shannon & Weaver (1949) believed that the meaning of the message is not as relevant as the message itself. They explained that a message is simply one selected message from a set of equally possible messages. Should that set be finite in size, the information possible to transfer by any message in that set is measurable. Generally we cannot predict the information that has yet to be received. We cannot be certain of which message will be received next. This uncertainty of an event can be measured as the inverse of the probability of it occurring.

2.5.3 Communication Channel

When communicating, the channel for communication should always be considered. A channel is almost always affected by noise. Noise comes in many forms, such as telephone line cross-talk; background radiation for radio waves; and even faults in the transmitter and receiver hardware (MacKay, 2003). The result of noise is that the information recieved at the source is not the same as was transmitted (Shannon & Weaver, 1949). Information theory provides two ways of combating this problem: error correction and compression, which are encapsulated in the coding theory (Murphy, 1998).

Noise

Noise can be said to be a function of the communication channel, such when the noise function is applied to the transmitted message, it produces the received message (Shannon & Weaver, 1949). A discrete noise function will always produce the same received message for each transmitted message. This is referred to as distortion. In principle, distortion can be corrected by simply applying the inverse of the noise function to the received message (Shannon & Weaver, 1949).

It is also possible that a noise function will result in a transmitted message producing a number of different recieved messages. This case has been found to be far more common than distortion (Shannon & Weaver, 1949). It can be assumed in this case that the recieved message R is the result of the function f which is based on the transmitted message T and the noise of the channel N. Here noise is considered to be the product of a stochastic process (Shannon & Weaver, 1949).

$$R = f(T, N) \tag{2.1}$$

Channel Capacity

Assume that a given channel is capable of transmitting a 1000 bits per second, with the $p_0 = p_1 = \frac{1}{2}$, where p_0 and p_1 are the probabilities of a message being a 0 or a 1 respectively. The noise function over the channel results in 1 in 100 bits being received incorrectly. This implies that less than 1000 bits are being transmitted each second, as 1% of bits are received incorrectly (Shannon & Weaver, 1949). This is because incorrect bits have not communicated any information to the destination. When the same channel has a different noise function applied to it, one that results in all 1000 bits being affected, it is find that on average 500 bits received are the same as the bits that were transmitted. However this channel has not actually transmitted any information, as the bits received were the product of noise and not the transmitter. From this it can be said that the capacity of a channel to communicate information is not necessarily equal to its rate of

communication. Rather, the capacity of the channel is the result of the uncertainty created by noise being subtracted from the rate of communication. *A Mathematical Theory of Communication* can be seen for further reading on this topic.

2.5.4 Coding theory

Error correction and compression are two different approaches to improving the representation and communication of information. Compression allows for information to be reduced to a form that makes more efficient use of the medium used to communicate or store it (Lelewer & Hirschberg, 1987). Error correction introduces redundancy into information to combat the effects of noise on a transmitted message (MacKay, 2003).

Error Correction

Noise comes in many forms, such as telephone line cross-talk; background radiation for radio waves; and even faults in the transmitter and receiver hardware (MacKay, 2003). For a binary string, the probability of a bit not being affected by noise is $(1 - f)$ and the probability of it being distorted is f (MacKay, 2003). Given this, consider the bitmap example in Figure 2.5.4 that is being communicated over a channel where the noise function has a probability of distortion of $f = 0.1$.

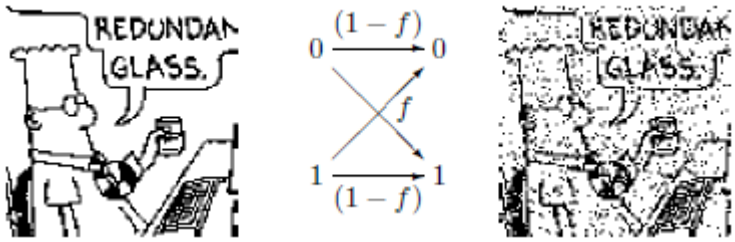

Figure 2.2: Example of channel noise (taken from MacKay, 2003, p. 4)

With as little as 10% of the bits flipped the image is significantly distorted. Ideally there would be no probability of distortion, and no bits would be flipped. There are two approaches to reducing the probability to the point where no bits are distorted. The

first involves improving the hardware and making it more reliable. Research into the physical design of components is a requirement of this approach, which along with the cost of producing new components increases the cost of the channel used to communicate (MacKay, 2003).

The other solution is to accept the probability of distortion and create a communication system that will allow for the detection and correction of errors. Error correction functions by introducing redundancy to transmitted messages. This redundancy makes it possible to recreate the original message after a certain acceptable level of distortion. While a hardware approach will increase the cost of the channel medium, this approach results in a more affordable computational cost (MacKay, 2003). In Figure 2.1 we see a transmitter and receiver connect the information source and destination to the channel. The transmitter acts as an encoder, adding redundant information to the message (MacKay, 2003). The receiver acts as the decoder, using the known redundancy to recreate the message to the best of its ability (MacKay, 2003).

Compression

In addition to correcting errors through redundancy, coding theory is also used to create systems for reducing the length of a message (Lelewer & Hirschberg, 1987). Known as data compression, these systems often transform the symbol set that a message is represented with into one that is more efficient (Murphy, 1998). This increase in effeciency leads to an increase in the capacity of the communication channel (Lelewer & Hirschberg, 1987).

Typical compression involves the mapping of messages from a given alphabet, α to an alphabet of codewords, β (Lelewer & Hirschberg, 1987). For example consider $\alpha = a, b, c, d, e$ and $\beta = 0, 1$ (Murphy, 1998). From this we can map

$$a \rightarrow 000, b \rightarrow 001, c \rightarrow 010, d \rightarrow 011, e \rightarrow 100 \tag{2.2}$$

In order to represent this language, we require 3N bits of information to be communicated (Murphy, 1998). If the string “acce” would be encoded it would require $3 * 4 = 12$ bits

and would map to "000010010100". This is referred to as fixed-block encoding (Lelewer & Hirschberg, 1987). Suppose that α has a probability distribution

$$p(a) = 0.25, p(b) = 0.25, p(c) = 0.2, p(d) = 0.15, p(e) = 0.15 \tag{2.3}$$

It stands to reason that we might represent messages with a greater probability of occurence in a shorter string of bits. This is called variable-block encoding (Lelewer & Hirschberg, 1987). If we map α to β as follows

$$a \rightarrow 00, b \rightarrow 10, c \rightarrow 11, d \rightarrow 010, e \rightarrow 011 \tag{2.4}$$

we will be able to represent the string "acce" in $2 + (2 * 2) + 3 = 9$ bits as "001111011". The same amount of information is being encoded as with the previous example, but it is being encoded into fewer bits. This means that the channel capacity of the communication system using this encoding is higher than one not using it

Going futher we can calculate the average number of bits required to represent α as

$$0.25 * 2 + 0.25 * 2 + 0.2 * 2 + 0.15 * 3 + 0.15 * 3 = 2.3bits \tag{2.5}$$

.

Thus this encoding will on average produce a shorter message than one that uses fixed-block encoding. Shannon & Weaver (1949) proves that there is a minimum number of bits needed to encode a message using a concept he named entropy (Murphy, 1998).

2.5.5 Entropy

A finite set of messages has a finite number of possible values. The total number of possible values from a set represents the uncertainty of a message. In the majority of cases a set

will have not have an equal probability distribution, resulting in some messages being less likely than others. Shannon & Weaver (1949) showed that messages that were less likely convey more information. This means that each message has its own quantifiable information content based on its probability.

In order to quantify the uncertainty of a message from a given set, an average of the information content (uncertainty) for all the messages is calculated (Shannon & Weaver, 1949). This is called the entropy of the message. For example consider the message set $A_x = \{a, b, c, d, e\}$ with the probabilities $P_x = \{0.25, 0.25, 0.2, 0.15, 0.15\}$. To calculate the entropy we use the following formula (Murphy, 1998):

$$H(x) = -\sum p(X = k) log_2 p(X = k) \tag{2.6}$$

Where $H(x)$ is the entropy of a message and X is a random message from the set. Here the $H(x) = 2.2855$. This means an average of 2.2855 bits are required to store a message. However with the probability $P_x = \{0.2, 0.2, 0.2, 0.2\}$ entropy $H(x) = 3$, results in an increased average number of bits required of 3. This is because entropy is maximised over a uniform probability distribution (Murphy, 1998). In contrast for the probability $P_x = \{1, 0, 0, 0, 0\}$ entropy is minimised ($H(x) = 0$) as the outcome is deterministic.

This principle is represented in the binary entropy function, seen in Figure 2.5.5. This function contains only two possible messages 0 or 1 (and is therefore binary). The probability of messages are supplementary and can be defined by $p(X = 0) = \theta$ and $p(X = 1) = 1 - \theta$.

Here we see that entropy maximisation ($H(\theta) = 1$) occurs when the probability is evenly distributed ($p(X = 1) = 0.5$) and minimisation occurs when probability is deterministic ($p(X = 0) = \theta$) (Murphy, 1998).

2.5.6 Use of Logarithmic Base

Shannon & Weaver (1949) claimed that the logarithmic function is the most appropriate

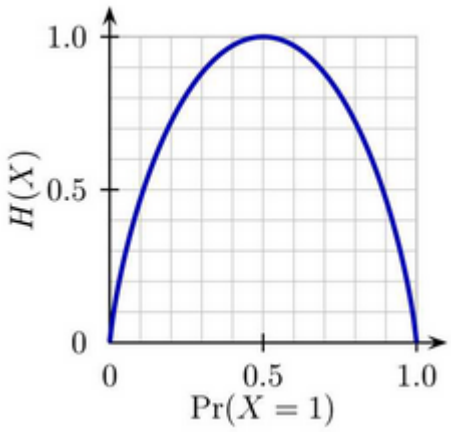

Figure 2.3: Graph of Unit Entropy (taken from Murphy, 1998, p. 4)

function for calculating the quantity of information in a message. Shannon motivated this claim with the following reasons:

- It was practically applicable to his field of research, telecommunications. Many of the important considerations in this field such as time, bandwidth and number of relays tend to have a linear relationship with the logarithm of their possibilities. Shannon explained this in terms of additional relays, where adding a relay will double the number of possible states a group of relays can represent.
- The logarithm closely relates to our intuitive feeling of a measure of information. We tend to feel for example that two devices should hold double the information of one, or two channels should have twice the capacity of one.
- It is mathematically suitable. It is easier to handle the logarithm than the number of possibilities.

For this reason Shannon decided to use logarithms in his measures of entropy and channel capacity.

2.5.7 Numeral Systems

A numeral system is another term used for a compression encoding mapping an alphabet α to a smaller simpler alphabet β. Parallels can be drawn to encoding theories in compression, which typically use a binary alphabet of $\beta = 0, 1$. However the size of the alphabet should be considered when determining how to encode α.

Information Theory makes use of a number of different logarithmic bases. Each has a different unit of measurement associated with it and is better suited to different applications. The numeral system used for representing these units has a digit selection size equal to the logarithmic base.

The logarithmic base 2 allows for use of the most simple unit of data, the binary digit. A binary digit is commonly referred to as a bit (a name suggested by J. W Tukey). It represents two possible states: on or off, true or false, powered or unpowered. This makes it particularly applicable to electronic circuitry, where many devices such as a transistor are either activated or not. A transistor is capable of representing one bit. N transistors will represent N bits, combined together they would be able to represent 2^N different possible states. This allows for a short series of simple data units to represent a large number of possible states, for example $2^4 = 16$ therefore 4 bits can represent 16 possibilities.

Decimal numeral systems make use of 10 symbols. Units of base 10 are referred to as a ban, but are also occasionally called a hartley or a dit (decimal digit). The use of a larger base allows for a ban to represent more possibilities than a bit, which has a base of 2. Representing a single ban in bits would require $3\frac{1}{3}$ bits.

The natural logarithmic base is also a commonly used base. Generally referred to as a nat or nit (following the convention used for bit and dit) it uses the base e.

The relationship between two bases a and b can be calculated using

$$log_a M = log_b M / log_a b \tag{2.7}$$

or a transformation from base a to base b can be made by multiplying by $log_b a$ (Shannon & Weaver, 1949).

2.5.8 Joint Entropy

Joint entropy is a calculation of the combination of two random variables using the dependency between their probabilities. Given variables X and Y the following formula is used (Murphy, 1998):

$$H(X,Y) = -\sum p(x,y) log_2 p(x,y) \tag{2.8}$$

The following example, where $X(n)$ represents n being odd and $Y(n)$ represents n being divisible by 3.

n	1	2	3	4	5	6
X(n)	1	0	1	0	1	0
Y(n)	0	0	1	0	0	1

In this example the probability of $X(n)$ being odd is the same as the probability of it being even. Or $p(X = 1) = p(X = 0) = 0.5$. Owing to the binary unity function we know that the $H(X)$ is 1. However $p(Y = 1) = 2/3$ and $p(Y = 0) = 1/3$ resulting in $H(Y) = 0.78$.

If X and Y are independent of each other then $H(X,Y) = H(X) + H(Y)$ (Murphy, 1998). If not, owing to the principles of mutual information, $H(X,Y) \leq H(X) + H(Y)$ (Murphy, 1998). In the above example X and Y are not independent and must have their joint probability distribution calculated.

P(X,Y)	0	1
0	2/6	2/6
1	1/6	1/6

$H(X,Y)$ =1.7233 in the above example.

Joint Entropy falls out of the scope of this investigation, but has applications in a deeper understanding of a communication system that may prove useful to solving the cheating problem.

2.5.9 Conditional Entropy

Conditional entropy is the entropy of an event A after the event B has occurred. It is the uncertainty of A knowing B. It is represented by the formula $H(A|B) = H(A,B) - H(B)$ (Murphy, 1998).

If A is completely independent of B, then $H(A|B) = H(A)$ and likewise if B determines A: $H(A|B) = 0$. Mutual information tells us that $H(A,B)H(A) + H(B)$ thus $H(A|B)H(B)$, achieving equality only when A and B are independent. This shows that receiving additional data will never increase the average uncertainty (Murphy, 1998).

2.5.10 Mutual Information

A string of messages describes an object, where X {hot, cold,...}.The first message describes the object as hot. We know that there is a connection between the concepts of hot and cold, this means that the message hot will alter the probability of the outcome cold, reducing the uncertainty of the next message (Shannon & Weaver, 1949).

This is described as mutual information and is quantified with the following formulae (Murphy, 1998):

$$I(X;Y) = p(x,y)log(p(x,y)p(x)p(y) \tag{2.9}$$

An equivalent simplification of this formula is as follows:

$$I(A;B) = H(A) - H(A|B) = H(B) - H(A|B) \tag{2.10}$$

Substituting in $H(A|B) = H(A, B) = H(A)$ results in

$$I(A, B) = H(A) + H(B) - H(A, B) \tag{2.11}$$

Resulting in

$$H(A, B) = H(A|B) + H(B|A) + I(X, Y) \tag{2.12}$$

Mutual information gives a measure of the amount of information given random variable holds about another (Cover *et al.*, 1994). Use of mutual information falls outside scope of this investigation, though it was originally intended to provide insight into the connections between information pieces being used in Section 3.3 and Section 3.4.

2.5.11 Relative Entropy

Given X that represents all possible representations of an image of a fixed size or all possible outcomes of a lottery draw, there is a very complex probability distribution $p(X)$. Modelling this probability distribution is challenging (Murphy, 1998). Given a more accurate model of this distribution $q(X)$, which will allow for a more efficient representation of it, we can measure the 'distance' between these two distribution models (Murphy, 1998). This distance is called relative entropy, and can be measured using the Kullback-Leibler divergence. See Murphy (1998) for further reading. Relative entropy is used in Steganography to determine the strength of a stegosystem, see Section 2.6.6.

2.6 Steganography

2.6.1 Introduction

Cryptology is the act of concealing the information stored inside a message. This however is often not sufficient, as information can be ascertained from the communication between

two parties, even if the message itself cannot be read (Anderson & Petitcolas, 1998). Instead of simply making a message unreadable, steganography conceals the existence of the message (Bennett, 2004). It is both the art and science of embedding a message in a cover medium that will not draw attention to the existence of the message (Provos & Honeyman, 2003).

The ideas behind Steganography have existed for many years. The word itself is taken from *Steganographia* written by Trithemus sometime between 1462 and 1516, origination from the Greek for "covered writing" (Bennett, 2004). The practices of Traditional Steganography involved the use of a physical medium, for example ancient Greeks writing a message underneath the wax on a writing tablet (Anderson & Petitcolas, 1998) or tattooing a message onto the shaven head of a slave and allowing the hair to grow back, thus concealing the message(Bennett, 2004). A more advanced medium used by Traditional Steganography was invisible inks. The medium in which the hidden message is placed is known as the stego medium (Provos & Honeyman, 2003).

Other Traditional Steganographic techniques involved systems for embedding messages inside of other messages. The communicating parties agreed upon the system, for example that every misspelt or corrected word is part of the message. In this case, two messages are being sent. The first is the message that a third party will be able to read which the intended message is encoded into. This is called the covertext or cover medium. The second is the message we wish to encode, sometimes referred to as the stegotext. The system for encoding the message into the covertext is known as the stegosystem (Cachin, 1998).

2.6.2 Modern Steganography

The invention of modern computing and communication systems caused Steganography to evolve from physical mediums to digital mediums. Since all information on a digital system is represented as a binary value of either 1 or 0, it is possible to embed a message in any digital medium. This makes modern steganography more powerful and usable than

physical steganography, as not only can messages be embedded in any data structure but embedding and extraction can be automated (Bennett, 2004). The more common covertexts are files like images, audio recordings or XML documents (Bennett, 2004, Memon *et al.*, 2008).

Information is embedded into a digital cover medium using the redundant bits contained by the medium (Provos & Honeyman, 2003). The most popular form of steganographic system revolves around the Least Significant Bit (LSB) of a particular word or section in the medium (Munuera, 2007). For example, changing the LSB in each pixel in an image has an effect that is hard to notice with the human eye (Munuera, 2007). Other techniques may have a greater effect on the integrity of the medium, and can lead to noticeable changes such as fuzzy images or audio. For this reason stegosystems are often concerned with the extent to which an medium is altered, with the objective being to minimise the change (Provos & Honeyman, 2003).

Embedding a message in a cover medium makes the embedded medium statistically distinguishable from the original (Provos & Honeyman, 2003). This distortion can be detected through a process called statistical steganalysis (Provos & Honeyman, 2003). Should it be possible to statistically detect a change in the medium then the stegosystem has failed (Anderson & Petitcolas, 1998). For this reason it is important to consider the cover medium and how the embedding will alter that cover text (Bennett, 2004). When comparing two systems, the one that alters the cover medium the least is the stronger of the two (Bennett, 2004).

Kerckoff's principle is widely accepted in Cryptology (Provos & Honeyman, 2003) and states that the strength of a cryptographic system should be based solely on the secrecy of its key, and should treat all other factors as public knowledge (Cachin, 1998). This is to minimise the effect of one of the other factors, for example the cover text, being exposed would have on the system (Provos & Honeyman, 2003). Steganography applies this principle in a similar way, with the strength of the system being placed on as few key bits of secret information as possible (Provos & Honeyman, 2003).

2.6.3 Factors in a Steganographic System

Capacity, security and robustness are three factors that are relevant to an information-hiding system (Provos & Honeyman, 2003). Capacity is the amount of information that can be hidden in the stego medium. Security measures how undetectable a message is in the medium. Robustness measures the amount of modification that can be made to a stego medium before the hidden information is destroyed.

An example of an information-hiding system other than Steganography is watermarking. Watermarking systems attempt to hide a piece of identification information in a medium, so that it can be traced (Wang *et al.*, 2009). Examples of this information are serial numbers and insignias. Watermarking therefore requires high robustness. The aim is to make the removal of the watermark without the destruction of the medium impossible (Provos & Honeyman, 2003). Capacity is not a concern passed a system having the capacity to contain the defined piece of information, and the security of the watermark changes depending on the situation. Currency for example requires low security, as everyone should be able to authenticate it.

Steganography on the other hand aims to maximise the capacity and security of a steganographic system, the process for doing so often makes the data fragile with minimal robustness (Provos & Honeyman, 2003).

2.6.4 The third party

In a steganographic communication between two parties, a third party is the party that is not aware of the communication but wishes to detect it. Without this third party, steganography would not have a purpose (Anderson & Petitcolas, 1998). The third party can operate either actively or passively.

When passive, the third party will simply observe the flow of messages (Bierbrauer & Fridrich, 2008). They will not make any attempt to alter the message or medium and

will attempt to discover a hidden message only through analysis (Anderson & Petitcolas, 1998).

A active third party will modify messages in an attempt to gain further insight into the stegosystem (Simmons, 1985). Messages may be created to elicit information from one of the communicating parties, or edited to obscure the intended message (Anderson & Petitcolas, 1998). A famous example of active third parties in Traditional Steganography was the prevention of the assassination of Queen Elizabeth of England by Mary Queen of Scots. Queen Mary conspired to claim the English Throne and used a cipher to communicate to her conspirators. However the cipher was broken, and a forged message asking for "the names and qualities of the six gentlemen which are to accomplish the designment" allowed the would-be assassins to be indentified and Mary to be arrested and executed (Anderson & Petitcolas, 1998). Another example is postal censoring and rephrasing of telegrams in the 20th Century (Anderson & Petitcolas, 1998). Common examples removed the X's from messages between lovers and changed the phrasing of "father is dead" to "father is deceased" as the two phrasings had different meanings in the stegosystem used (Anderson & Petitcolas, 1998).

One of the primary examples for discussion of Modern Steganography is the Prisoners' Dilemma created by Simmons in 1983 (Anderson & Petitcolas, 1998). In this problem the warden, who is the third party, can act either actively or passively (Simmons, 1985).

2.6.5 Prisoners' Dilemma Problem

The Prisoners' Dilemma was a problem proposed by Simmons (1985) and is a widely used example in Steganography (Bierbrauer & Fridrich, 2008, Cachin, 1998). The problem is concerned with the communication between two separated prisoners, who are allowed by the warden to communicate through messages transported by the warden's agents. The prisoners are allowed to communicate because they have agreed to make their messages readable by the warden, either by sharing their method for decryption or communicating in unencrypted messages (Simmons, 1984, Bierbrauer & Fridrich, 2008).

The aim of the prisoners, Alice and Bob, is to plan their escape from the prison. The aim of the warden, Eve, differs between scenarios. In a passive warden scenario, Eve will simply read the messages and end communication should she discover the prisoners are planning an escape (Bierbrauer & Fridrich, 2008). Here the communication channel is treated as being noise free, and messages are not distorted. In the active warden scenario Eve aims to deceive one prisoner into believing a message received, one she either tampered with or created, is actually a genuine message from the prisoner's accomplice (Simmons, 1985). In this scenario prisoners must authenticate that messages are indeed from their accomplices, and cannot trust the communication channel to not cause distortion.

Alice and Bob communicate through a covertext, that appears to be an innocent message , with a stegotext encoded into it containing the message they do not wish to be detected (Cachin, 1998). Simmons (1985) established that in order for Alice and Bob to achieve their aim, they must establish a channel for communicating without Eve knowing. He referred to this as a subliminal channel.

When communicating with an active warden Alice and Bob must authenticate the received message as coming from their accomplice, and not from Eve. To authenticate messages the prisoners must include redundant information, which when present in a message will imply that it is genuine (Simmons, 1985). The probability of Eve choosing the correct redundant information for a genuine message, P_a, is dependent on the amount of redundant authentication information, Hr (Simmons, 1985). The probability P_a can be calculated as $P_a = 2^{-Hr}$.

The opponent, Eve, must be able to verify that the message contains only accepted information. She must be allowed to decrypt the information into a form that is readable, thus authentication information alone will not allow for secrecy. Rather this establishes an authenticated channel of communication between the prisoners, which contains no secrecy. Alice and Bob must construct a steganographic system for communicating using a subliminal communication channel.

2.6.6 A steganographic system

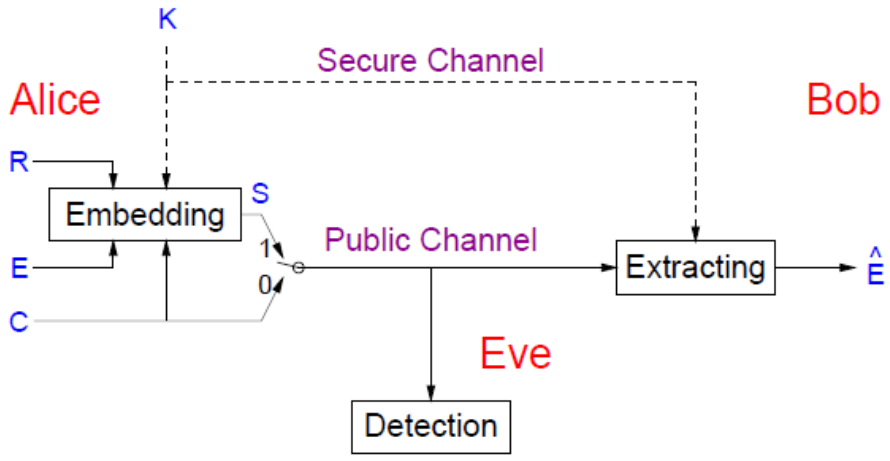

Figure 2.4: A secret-key stegosystem (taken from Cachin, 1998, p. 3)

Alice wishes to send a message to Bob. The message can either contain an embedded message or simply be covertext, in which case she is acting either actively or inactively respectively.

While acting inactively Alice sends a covertext, denoted as C, from the distribution P_C (Cachin, 1998). Covertext is generated from a source known only to Alice, and contains no hidden message.

When Alice is acting actively she sends a stegotext, denoted as S, which is generated from the distribution P_S. It contains a hidden message E, embedded using the embedding function F. E is a random message, and forms part of the message space . The embedding function requires Alice and Bob to share a secret key K and a private random source R. Following Kerckhoff's principle it can be presumed that F, P_c and P_s are all known to Eve, but K must remain secret (Cachin, 1998).

In Figure 2.4 the switch before the public channel determines whether Alice is active or inactive. Alice is inactive when the switch is in position 0, and sends only the cover text C across the public channel. Alice is active when the switch is in position 1. Here message E is embedded in covertext C using the function F based on the shared key K and the private random source R. The product of this is the stegotext S that is sent across the

public channel. Bob is able to extract the message E using the extraction function G and the shared key K (Cachin, 1998).

It is worth noting that this representation of a steganographic system has a flaw. For the system to work, Bob must know whether Alice is active or inactive. To know this Bob would require information that is not transmitted in the message (Cachin, 1998). However this can be ignored when considering the possibility that Bob may simply attempt to use the extraction function G on every message received, and the fact that doing so does not make the system any weaker to detection from Eve (Cachin, 1998).

Alice may send a message from one of two probability distributions, P_C or P_S, depending on her being active or inactive. If Eve is to detect a hidden message, she must perform a technique called Hypothesis Testing in order to determine which distribution a message falls into (Cachin, 1998).

Relative entropy is the difference between two distributions. We can quantify the security of a stegosystem in terms of its relative entropy, $D(Pc||Ps)$ (Cachin, 1998). In a perfect stegosystem $D(Pc||Ps) = 0$ (Cachin, 1998).

Hypothesis Testing

Hypothesis testing is the act of evaluating which of several hypotheses are true (Cachin, 1998). It forms a part of statistical analysis (Provos & Honeyman, 2003).Hypotheses H_0 and H_1 for example attempt to explain an observed statistical measurement from a text, Q (Cachin, 1998). This results in two different probability distributions that may have produced Q, namely P_{Q0} and P_{Q1} which correspond to the hypotheses H_0 and H_1. If a hypothesis is correct, then Q was generated from the corresponding probability distribution.

A mathematical approach to hypothesis testing is not relevant to the cheating problem. Statistical analysis cannot be easily done on the non-digital mediums discussed in Section 3.3, but the understanding of the approaches used to detect cheating is important to consider.

Chapter 3

Investigation

3.1 Defining cheating

Many have attempted to understand cheating by studying the reasoning for, justifications of, or attitude toward cheating provided by candidates. However this provides understanding of the candidate, not of the act of cheating. This investigation will attempt to define cheating in a way that is relevant to theoretical examinations, create categories for the classification of cheating, and apply concepts from information theory and steganography to these categories.

Typical theoretical assessment tests a candidate's knowledge at the time of assessment, through unseen questions in an environment that constrains their access to external information. Candidates are allowed time to prepare themselves for the assessment by reviewing the relevant sections of a subject, and occasionally by preparing a response to a seen question or a document containing information they are allowed to take into the examination. Assessments have a fixed time period and a fixed number of achievable marks for each question. The mark allocation for a question is relative to the amount of time it should take or its difficulty.

The type of assessment for a subject changes depending on the classification of its discipline. As discussed earlier hard pure discipline assesses regularly, with a larger number

of short questions aimed at testing atomic pieces of knowledge. Soft pure disciplines in contrast ask long questions, usually in the form of essays, to assess a candidates knowledge of the whole subject at the end of the teaching period. Hard and soft applied disciplines differ from their pure counterparts by having practical assessments that test skills and practices, usually in addition to theoretical exams. Owing to the fact that practical assessments differ greatly from each other and theoretical exams, they will not be a focus for investigation.

The question in a theoretical examination assesses a candidate's knowledge by requiring the candidate to produce solicited information. Examinations do not require production or recall of data only information, as everything required by an examination has meaning associated with it. This information can come in the form of single atomic units of information, such as an answer to a multiple choice question or a question that requires a candidate to name or list something, or a coherent or logical collection of information, such as a mathematical equation or an essay. Regardless of the form of the information, this information must be produced by the knowledge of the candidate.

Assessment of atomic units of information can for convenience be considered as testing the ability of a candidate to recall information. Recalling information requires the candidate to have the understanding capable of producing the specific piece of information that the question relates to. It can be thought of as a basic form of knowledge. In this case, once the information has been produced it is sufficient to answer the question. In order for a candidate to skew their assessment for such a question, the candidate simply needs to access that piece of information.

For questions requiring a collection of information, knowledge is tested on a far deeper level than simple information recall. In a hard pure discipline a candidate will be required to apply their knowledge to unseen problems, requiring for multiple bits of connected knowledge to be produced. Candidates in a soft pure discipline have their own opinion and overall understanding assessed, requiring them to produce a coherent collection of information that adequately represents their knowledge relating to the question. Owing to the fact that this is a collection of information, and more subjective and personal in soft

pure disciplines, it is unrealistic for a candidate to gain access to the entire collection of information. It is more likely that a candidate who cheats would gain access to individual pieces of information capable of creating the understanding needed to answer the question. Should a candidate be able to access an entire collection of information, this can be considered the same as accessing a single piece of information sufficient to answering a question.

In both cases information is external. The origin of this information falls into one of two categories. The first category is information stored in a medium prior to the beginning of the examination. This medium can be sanctioned, in the form of a cheat sheet' authorised by the examiner, or can be brought in to the examination illegally for the purpose of cheating. It has been produced by the knowledge of another, and does not reflect upon the candidate. The second category is information that is produced during the examination by other knowledge of other candidates.

From these points of origin, we can establish three categories for cheating:

1. information illegally stored in a medium brought into the examination: Information Smuggling.
2. information gathered from the answers of another: Information Theft.
3. information communicated from one candidate to another: Communicated Information.

Information theft will not be investigated by this paper, as it does not involve the creation of a system for cheating prior to the commencement of the examination.

Cheating can be defined as access to external information that could not be recalled or created by the candidate, that is either sufficient to answer a question or capable of creating the knowledge that would allow a candidate to do so, thus skewing the assessment of that candidate's knowledge.

3.2 Theoretical Explorations

3.2.1 Alice's History Examination

Alice is currently enrolled in a History course at her university, along with a number of other subjects. Her history examination will assess her knowledge of the French Revolution, which is divided up into three sections. To prepare for the upcoming examination period her History lecturer has given her the following questions for each of the sections, explaining that knowing how to answer these questions fully will prepare her for the actual examination.

The section on the "March on Versailles" requires Alice to know:

- Who marched on Versailles?
- Why did they march?
- What was the outcome of the march?

The section on the "Storming of the Bastille" requires Alice to know:

- What lead to the storming?
- What is the 'Declaration of the Rights of Man and of the Citizen'?

The section on the "Reign of Terror" requires Alice to know:

- What was the significance of the Guillotine?
- What was the Revolutionary Tribunal?

Alice does not consider history important to her degree as it is not one of her majors. She wishes to do the minimum amount of work required in order to pass the subject. This paper will use Alice's situation as an example of a typical approach to cheating.

3.2.2 Motivation for cheating and choice of technique

A candidate, like Alice, wishing to cheat does so to skew their assessment in their favour, in the hopes that an incorrect assessment will be of benefit to them. A candidate might do this for one of two reasons, either they lack the cognitive ability to create or retain the knowledge required to perform well in the assessment or they are not willing to commit the time required to sufficiently internalise the knowledge. A lack of cognitive ability may only extend to the knowledge required, and should not be considered a reflection of a candidate's overall cognitive ability. For example a candidate may struggle with Mathematics, but excel at English.

A candidate that lacks the cognitive ability to retain the required knowledge, or is not willing to invest the time required to internalise the information, is faced with a far simpler option for cheating. Presuming that the candidate is capable of utilising the knowledge required to perform in the assessment, they simply require access to external information that will allow them to create it. All three cheating techniques make this possible, with information smuggling being the most feasible technique as it can be performed individually and requires the least effort.

A candidate lacking the cognitive ability to create the knowledge required inside of an examination is posed with a more interesting problem. Presuming once again that the candidate is capable of utilising the knowledge that would be required to answer an unseen question, the issue becomes how to create this required knowledge. Smuggling external information which would be used to create this knowledge is not possible, as the knowledge required is not known before the commencement of the examination. Only Information Stealing and Communicated Information are feasible as a result. As previously established, information stealing is an unreliable technique which makes Communicated Information the preferable choice. However, Communicated Information entails a number of issues that make it more complex than information smuggling, such as the need for a partner.

The reason for a candidate cheating is a key factor in the feasibility of the different cheating techniques. As a general rule, a candidate who does not wish to invest the effort

required to prepare for an examination should make use of information smuggling, while a candidate who lacks the cognitive capabilities to perform in an examination should rely on the information communication. Information theft should be a last resort, owing to its lack of reliability.

3.2.3 Time to master

It is important to consider the amount of time required to create either a system for communication or a stegosystem that will be used to cheat. As previously stated, one of the motivations for cheating is not being willing to commit the time required to fully internalise required knowledge.

Candidates should consider the following:

- S - Amount of time available for preparation
- I - Amount of time required to internalise knowledge
- C - The amount of time required to create a communication or steganographic system
- D - The difference between I and C such that $D = I - C$

There are two situations when a candidate may believe that cheating is their best course of action. The first is when $S < I$ and $S \geq C$. Here the only option that will allow them to fully prepare for the examination is to develop a cheating system in the limited time available. The second is when the difference D is great enough to justify the risk, this is purely subjective on the behalf of the candidate.

3.2.4 Recall, retention and knowledge

The definition of knowledge in this paper states that it exists solely in the mind of the knower. The implication of this definition is that all knowledge exists in the same state, in

which it accessible by the knower without requiring conscious effort or external information to recall. This is not true, as knowledge can be said to have a level of internalisation that reflects the amount of effort or information required to recall it.

Fully internalised knowledge does not require effort to recall, and exists in a state of unconscious accessibility. Alice possesses knowledge about History. She knows that the events of History took place before she became aware of them. Alice has fully internalised this piece of knowledge to the extent that it is part of the way she thinks; anything she learns in history she knows happened before she learnt it, without having to consciously think about it.

However a piece of knowledge that has been recently acquired requires some effort to recall, as it has not been fully internalised. While reading her History textbook Alice reads information about French women marching on Versailles because of a scarcity of bread. This information creates the knowledge that economic problems in France were causing dissatisfaction with the ruling authority, King Louis the XVI. A few days later when reading about another section of the French Revolution, Alice is not conscious of her recently acquired knowledge and does not recall the economic dissatisfaction of the French people until she is reminded about the March on Versailles. This idea has only been partially internalised, and requires external information to recall. Conscious mental effort may also allow Alice to recall this information. Thinking about the French Revolution may allow her to make mental associations to topics that remind her of the March on Versailles, which in turn helps her recall knowledge about the people's economic dissatisfaction.

When considering knowledge, we must consider the degree to which it has been internalised. At the highest degree of internalisation information has been fully internalised and can be unconsciously recalled. As the degree of internalisation decreases, the amount of conscious effort or external information required to recall that information increases.

3.2.5 Information Lists

Alice is faced with the challenge of internalising the knowledge required to answer the questions given to her by her lecturer. Fully internalising each piece of knowledge will require more time than Alice is willing to invest. Instead she decides that a lower degree of internalisation will be sufficient, if she can find a way of easily recalling each piece of knowledge. To do this she will require a technique to help her recall the information.

One popular recall technique, which requires mental effort, is mnemonics. Mnemonics translate information into a form that can be remembered more easily, and has been shown to aid in memorisation (Bellezza, 1981, Pandey & Zimitat, 2007). Most commonly a phrase or word is used to represent the pieces of information central to a piece of knowledge. The first letter of each word in the phrase or each letter in the word corresponds to the first letter of a piece of information. For example the phrase "**M**en **S**tarted **R**ioting" is a mnemonic for the key sections in French Revolution.

Men - **M**arch on Versailles

Started - **S**torming of Bastille

Rioting - **R**eign of Terror

Here each piece of knowledge is easily reduced to a key piece of information, that can be recalled using a relevant phrase. It requires mental effort to memorise the phrase and the association to the pieces of information. This means that mnemonics are not a cheating technique, as no external information is required.

Continuing with the concept of reducing knowledge into key pieces of information, Alice decides to make a list of containing all the information required to answer the seven questions given. She represents her list as follows:

1. March on Versailles
    (a) women

   (b) bread

   (c) louis/paris

2. Storming of Bastille

   (a) economy

   (b) human rights

3. Reign of Terror

   (a) national razor

   (b) dictatorial Power

This shall be referred to as an Information List, as it contains all the information required to recall knowledge. It contains the same information that she would use in her mnemonic, but does not require mental effort to memorise or recall through association. It is an example of external information, and using such a list in an examination is cheating.

It should be noted that in the example above, the knowledge recalled through the information provided has a low level of detail. Expanding on this knowledge by providing additional information for details increase the amount of information that needs to be either listed or memorised. For example the "March on Versailles" can be expanded as follows:

1. Women

   (a) wives

   (b) market

   (c) paris

2. Bread

   (a) overpriced

   (b) scarce

3. Louis/Paris

    (a) return of ruling authority

    (b) end of independent authority

Increasing the amount of information required will increase the amount of mental effort required when using the mnemonic technique, as increased information requires additional or more complex mnemonics. It will also cause an increase in the length of the Information List. Alice decides to use an Information List as opposed to a mnemonic as it requires less mental effort.

The main advantages of creating an Information List to represent required knowledge are: the reduction of knowledge to single piece of atomic information, and the ability of an information list to be stored and transmitted. Reducing knowledge to single pieces of information has the effect of increasing the channel capacity of either the medium or communication channel being used owing to the reduced amount of information being sent (Lelewer & Hirschberg, 1987). Information that can be stored or represented symbolically can be used in the encoding systems that will be explained in Section 3.3.2 and Section 3.4.4.

Information Lists will be used in this paper as a practical example of how information used for recall can be represented.

3.3 Information Smuggling

A candidate can skew their assessment by using information created prior to entering the examination venue. This information does not exist as part of the candidate's knowledge at the time of the examination, and is brought into the venue illegally on a physical medium.

This medium may be authorised, overlooked or unauthorised. An authorised medium is permitted inside the venue by the examination restrictions, such as an eraser or form of

identification. An overlooked medium is one that is conventionally allowed into the venue despite it containing or having the potential to contain information, such as a water bottle with a label or the cover for an eraser. An unauthorised medium is one that is not permitted inside the venue by examination restrictions, such as a mobile phone or piece of paper.

It is assumed that a candidate will wish to avoid punishment for having unauthorised information. It is therefore ideal for the candidate that the medium is overlooked. An unauthorised medium, such as a cell phone, will be detected upon inspection by an examiner. An overlooked medium's information is able to avoid detection when under inspection.

The amount of inspection required to detect information is a relevant concern to a cheating candidate. Ideally, for the candidate, it should be impossible to detect the information, though it is unlikely that this is will be the case. It is more plausible that an examiner will overlook a representation of information on a medium.

This section will discuss the concept of examiner oversight, explaining how the decisions made by examiners allow for the creation of stegosystem for smuggling information. Following this, we will devise a practical example of such a system and draw attention to the shortcomings of this approach.

3.3.1 Examiner Oversight

Examination environments differ in their strictness, even when the restrictions are specific. However a fair number of examination restrictions are vague, leaving it up to the examiner to detect a medium that contains illegal information. This approach is fairly effective, as for the majority of situations detecting illegal information in a medium is fairly simple.

The easiest form of detection of illegal information is the detection of the medium itself. A piece of paper could easily be identified as having the potential to contain illegal information, either because it is specified that no loose paper may be brought into the

examination or because the examiner decides that it may contain information relevant to the examination. Likewise a programmable digital device, such as an music player; cell phone or programmable calculator, should draw the attention of an examiner.

Such mediums are not appropriate for a strategic attempt to cheat. They depend on luck and timing, with a candidate relying on having the time to reveal the medium from its hiding place for long enough to find the relevant information without being seen. There is very little reliability in such an approach. A more Steganographic approach would focus on finding a better cover medium that could avoid detection or be able to withstand the attention of an examiner. This medium would not need to be concealed.

A distinction must be drawn between a medium being detected and drawing attention. While in a number of cases, such as a digital device, drawing attention is equivalent to being detected, there are some situation where an examiner may make the decision that a medium does not contain illegal information even though it may draw attention.

One example of this may be a drawing on a candidate's ruler or eraser, such as an impression of primitive cave paintings. These drawing should draw the attention of an examiner, but upon examination of the drawing an examiner would be faced with a dilemma. The drawing, while against examination guidelines, does not appear to contain any information allowing a candidate to cheat. Examination guidelines would most likely require the candidate to be removed from the examination, but doing so may punish an innocent candidate. An examiner may decide that the drawings can be overlooked in order to not unnecessarily inconvenience a potentially innocent candidate. Removal of a medium required by the candidate, such as stationery, may also disadvantage a candidate.

In this oversight lies the potential for Information Smuggling. It creates the potential of a practical steganographic system to be used in an examination environment where the limitations of the environment is determined by decisions made by examiners.

3.3.2 Creating a Steganographic system

A steganographic system should be concerned with the medium it makes use of and probability distribution the covertext used by that medium, the information it wishes to embed and system for embedding that information. A basic example system will be created as a demonstration of this process.

Medium

Alice requires a medium that will not draw attention if present in an examination, and should avoid mediums that have previously been associated with cheating. Pieces of stationery are appropriate mediums as they are required inside an examination and will imply that the medium has been with Alice for an extended period of time. This presents Alice with both a reasonable excuse as to why there may be writing on the medium (e.g. she doodled while in class) and the necessity of the medium (she requires stationery to complete her examination) discourages an examiner from removing it from the examination. Any medium that fits these criteria is viable, but due to the limitations of the environment stationery is one the most preferable.

This paper will develop a steganographic system that uses a ruler as the cover medium, chosen as it offers a flat surface to write upon.

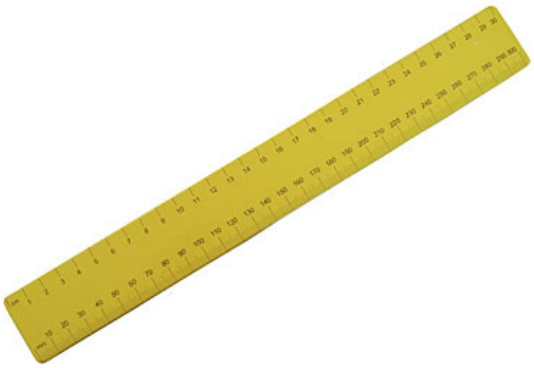

Figure 3.1: A feasible medium

Covertext

It is essential that the covertext used to embed information is overlooked by an examiner. If this is to happen the covertext must appear to exist on the medium for no reason. If it appears deliberate, it will lead to detection. There should be a plausible reason for the symbols existing that an examiner can easily discover, for example that the candidate doodled shapes on their ruler while sitting in class. The shapes are simple and common enough symbols that they will not appear deliberate. Relations or patterns between symbols may appear to be artistically motivated. This should allow a symbol set based of basic shapes to escape the attention of an examiner. For example consider the circle, rectangle and triangle represented in Figure 3.3.2.

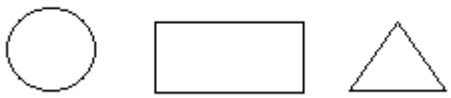

Figure 3.2: Basic Symbol Set

These symbols themselves are generic enough to escape notice, as they do not appear to represent information. Any set of symbols that may conceivably be overlooked can be used.

Together the symbol set makes up the covertext alphabet β that will be used to embed information into the medium. It is imporant to note that this alphabet is being created for the medium, rather than the medium dictating the alphabet that can be used (past the practicality of writing the alphabet on the medium).

$$\beta = \{circle, rectangle, triangle\} \tag{3.1}$$

Stegotext

The external information that Alice wishes to bring into the examination is the stegotext message set α for the stegosystem she will be using. Using the example of the French

Revolution, in particular the "March on Versailles"

$$\alpha = \{women, bread, paris\} \tag{3.2}$$

Encoding

The system for encoding our stegotext α into our covertext β should consider the fact that encoding and decoding will be done by a human being, and not an information system. One option for overcoming this problem is to create an encoding that is built upon associations between the covertext symbol and the stegotext message. Using Alice's example, a circle forms part of the symbols used to represent bothgenders, thus allowing for a connection to women. A rectangle is a similar shape to a loaf of bread. A triangle is a similar shape to the Eiffel tower in Paris. The example encoding E would appear as follows

$$\begin{aligned} E &= \alpha \rightarrow \beta \\ E &= \{women \rightarrow circle, bread \rightarrow rectangle, \\ &\quad paris \rightarrow triangle\} \end{aligned} \tag{3.3}$$

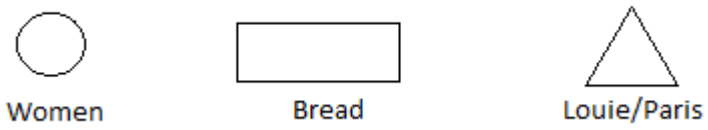

Figure 3.3: Associated Encoding

This system can be extended for larger covertexts and stegotexts.

3.3.3 Shortcomings

Several shortcoming are revealed when attempting to apply Steganography to cheating, both with the systems created and the applicability of Steganography.

Statistical Analysis

Modern Steganography is concerned with embedding into a digital medium. As such detection comes from a statistical analysis of the medium, attempting to discover changes in the probability distribution of the medium from those that could be expected of a medium that has no embedded message. This form of detection is not compatible with the Information Smuggling. The result of this is that many of the principles of Modern Steganography are not applicable to cheating.

Time to Master

As established in Section 3.2.2 one motivation for the use of an Information Smuggling system is the reduced time required to create the system when compared with taking the time to internalise the knowledge required. The system created however requires a similar process to internalisation. Both require you to identify important information and internalise a way of associating that information so that it can be recalled. The only advantage offered by this system would appear to be that only the association need to internalised, and not the information itself. For example when trying to internalise that it was the women who marched on Versailles, one must internalise the information that it was the women and the association with the march. If one wished to smuggle the information, the only advantage would be that the information about the women would not need to be internalised, while a different association (that between the circle symbol and the women) would still need to be internalised. It should be considered however that the identifying important information has the side effect of the associated knowledge being internalised. The act of creating the system may render the system obsolete.

3.4 Communicated Information

The act of communicating information is one that has undergone much study. This section will apply the widely accepted concepts of Information Theory to the act of communicating information illegally in an examination environment. It will discuss the dynamics of a cheating partnership, the relevance of numeral systems, and devise a practical approach for developing communication systems that are capable of operating inside examinations.

3.4.1 Partner

Communicating requires the involvement of at least two parties, where one operates as an information source and the other an information destination. This means that for Alice to cheat she requires the assistance of another candidate, Bob. While it is possible that Bob may not be a candidate of the examination, or may be an examiner, this possibility will not be considered as it exists outside of what can be generalised about the examination environment. It is presumed that examiners will operate in the interest of assessment, as the integrity of the examination has a correlation with their own integrity. If the assessment is compromised, so is their reputation. An accomplice who is not present in the examination environment falls outside of the scope of this paper. Techniques for cheating that allow Bob to communicate from outside the examination venue are dependent on the structure of the venue, and may not be possible in a different one. It can be assumed that accomplices Alice and Bob are both candidates for assessment.

The motivation for communication must also be considered. As was previously established in Section 3.3.2, a candidate who lacks the cognitive ability to perform in an examination can rely on the knowledge of another candidate for handling unseen questions. In this situation there is the implication that one candidate is dependent on the other, as one has the ability to perform and the other does not. The concern for Alice, our incapable candidate, is that Bob, our capable candidate, has no motivation to cheat. The transfer of information will be unidirectional. However it is equally possible that Alice may have the ability to answer questions that Bob cannot, or has knowledge that Bob lacks without

sufficient time to impart it on Bob. In this situation, Alice and Bob are co-dependent. A communication system devised by them would need to be bidirectional.

When considering a unidirectional partnership in which one candidate has nothing to gain from the other, and is merely inconveniencing themself, it should be considered that alternative motivation for assistance may exist. An emotional relationship between partners, such as a friendship or romantic relationship, or an agreement between partners, where roles will switch in a different assessment, are understandable motivations for partnership.

3.4.2 Numeral System

A communication system transfers information in the form of messages. Messages form part of a possibility distribution, which contains all possible messages to send and their likelihoods. In normal verbal or text communication, human beings generally communicate using sentences constructed from complete words. If Alice wishes to know what concept is relevant to a question in her examination, the normal form of communication may be a question such as "What should I talk about in question 3?". However communication is prohibited in an examination environment, and evidence of communication may lead to disciplinary action. It is in the interest of a candidate to communicate in as few messages as possible because of this. The shorter the transmission time of a message and the time it exists in the communication channel, for example the shorter the time a hand signal is used, the lower the chance of being detected. Coding theory provides us with compression algorithms that make such a practice possible.

We have already touched on the topic of reducing knowledge to atomic units of information in section 3.2.6. These atomic pieces of information have the ability to create the knowledge required to answer a question. If Alice requires a piece of knowledge to answer a question, she simply requires information produced by Bob's knowledge to create her own.

Using the example of the French Revolution, the following is the alphabet of information

that must be mapped onto a more appropriate alphabet

$$\alpha = \{women, bread, Louis, economy, humanrights, \\ nationalrazor, dictatorialpower\} \tag{3.4}$$

What constitutes a more appropriate alphabet? An alphabet should

- use messages that are as short as possible
- be easily translatable into its original meaning

.

Common representations of information are the bit, ban (or decimal digit) which are capable of representing 2 and 10 different possibilities with a single digit. A bit has the alphabet $\beta = \{0, 1\}$ and the alphabet of a ban is $\theta = \{0,1,2,3,4,5,6,7,8,9\}$. If one wishes to represent α in either of these alphabets the results would be the following:

$$\begin{aligned} women &\rightarrow 000 \\ bread &\rightarrow 001 \\ Louis &\rightarrow 010 \\ economy &\rightarrow 011 \\ humanrights &\rightarrow 100 \\ nationalrazor &\rightarrow 101 \\ dictatorialpower &\rightarrow 110 \end{aligned} \tag{3.5}$$

and

$$\begin{aligned} women &\rightarrow 0 \\ bread &\rightarrow 1 \\ Louis &\rightarrow 2 \\ economy &\rightarrow 3 \\ humanrights &\rightarrow 4 \\ nationalrazor &\rightarrow 5 \\ dictatorialpower &\rightarrow 6 \end{aligned} \tag{3.6}$$

Unseen questions result in the probabilities of the different information pieces being unknown, meaning that fixed-length encoding is the more appropriate than variable-length. The results is that the β results in 3 bit message and θ in 1 ban messages. Considering our first criterion it appears that a ban encoding would be preferable. Any message could be conveyed using a single transmission, which is preferable to three individual transmission for the bit representation. A ban also appears preferable considering our second criterion, as collections of bits may prove challenging for the human mind to translate into singular facts. Recalling that the number '3' decodes to the information "economy" is more natural to human thought than decoding '011'. It should be noted that the time spent to internalise the mapping of an encoding system is not of particular relevance, but the difficulty of decoding is of great importance. The more uniquely identifiable the digit, the easier the mapping.

There is theoretically no limitation on the size of the alphabet that can be used for the numeral system. A numeral system's alphabet could have a 4 digit alphabet as easily as it could have a 10 digit alphabet. The conclusions found by our criteria imply that the best size for a numeral system is the exact number of pieces of information that we wish to represent. This system will transfer only a single digit for each piece of information, and will be the easiest to decode.

However this results in a practical problem. An alphabet of 32 different concepts would

require 32 unique ways of representing information. When information representation cannot take the form of human language or numerical digits, the number of unique representations becomes more challenging. Take for example representation on the human hand, which has five fingers (conveniently referred to as digits). The alphabet size for a single human hand is 5 unique digits, and ten for two hands. Each combination of extended digit is unique, and can represent a letter in an alphabet. However, attempting to hold extend your thumb, middle and pinkie finger will demonstrate the difficulty of using the human hand to represent all 5 digits. A third criterion must be considered: the practicality of representing the size of the alphabet. This criterion opposed the ideals of the first two, resulting in the ideal alphabet being the largest one that can be easily represented.

Considerations for representation

Consider a visual communication system built upon a physical medium where the states of that medium are used to represent different letters of the encoded alphabet β. A typical medium remains constant throughout communication, with the nature or state of the medium (its positioning, shape, orientation) changing according to the letter that is being communicated. Owing to its physical nature, a medium has a limit to the number of messages it can reasonably represent. The capacity of a medium C is the number of states it can reasonably represent. For example a pen may have 4 distinctly different states based on its orientation to the front of a desk S_o where

$$S_o = \{north, south, east, west\} \tag{3.7}$$

The state of a pen may also be represented by its position on the desk in relation to the candidate S_p where

$$S_p = \{left, right\} \tag{3.8}$$

These physical mediums have a very limited number of states that can be used to represent each letter in the required alphabet, $|S_o| = 4$ and $|S_p| = 2$. We can attempt to increase the number of states a medium can represent, though in doing so we may decrease the distinctiveness of each state. Consider S_{oo}

$$\begin{aligned} S_{Po_2} = \{&north, north - east, east, \\ &south - east, south, south - west, \\ &west, north - west, north\} \\ |S_{Po_2}| &= 8 \end{aligned} \tag{3.9}$$

The number of states has been doubled by including an orientation between existing orientations. However in doing so we have decreased the distinctness between states. A reduction in the distinctivness of messages may result in the distortion of the message being transmitted. This noise should ideally be minimised by making states as distinct as possible. The number of states used to represent a medium should be minimised to a level where each message is distinct.

As established, a single letter of an alphabet should ideally be capable of communicating an entire unique piece of information, resulting in $|\alpha| = |\beta|$. We can achieve this using a collection of distinct states from one or more mediums. Each letter in the alphabet can be represented by a combination of different state classifiers for the mediums being used, without decreasing the distinctiveness of each state. Consider

$$\begin{aligned}
\beta &= S_o \times S_p \\
&= \{north, south, east, west\} \times \{left, right\} \\
&= \{\{north, left\}, \{north, right\}, \{south, left\}, \{south, right\}, \\
&\quad \{east, left\}, \{east, right\}, \{west, left\}, \{west, right\}\} \\
|\beta| &= |S_o| \times |S_p| \\
&= 8
\end{aligned} \tag{3.10}$$

where a combination of two distinct state sets results in a larger alphabet. Combinations of state sets present a convenient way of creating larger alphabets without significantly increasing the possibility of misreading a symbol.

An appropriate alphabet

When considering the appropriateness of an alphabet one should consider the established critera should

1. minimise the number of letters required to transfer a message
2. be easily translatable
3. be practical

where a practical alphabet should ideally make use of a combination of distinct state sets.

3.4.3 Message Set

The creation of a practical means for communication inside an examination should consider first the message set that it wishes to communicate. In addition to the identified

atomic pieces of information in an information list, the message set should contain control messages such as “Request”,“Begin”,“End” and “Repeat”. These messages will enable candidates to request the information they require, neatly encapsulate collections of information and request a retransmission respectively. It should also be possible to transfer messages that contain numbers, so that candidate may identify for which question their accomplice is requesting information. Examples of expected message strings are ”Request,Begin,1,0,End” which requests the information relevant to question 10, and ”Begin,1st-numeral,3rd-numeral,End” which conveys information pieces one and three in response to this request.

Another consideration is the whether communication of information is bidirectional or unidirectional. In bidirectional communication both candidates require information from one another. Their message sets must be symmetric, as both require the same functionality of the other. In a unidirectional system only the incapable candidate requires information. Two message sets can exist in this case, one for requesting information and one for transmitting it.

We consider the following message set for bidirectional communication:

$$\begin{aligned}\alpha = \{&request, begin, end, repeat,\\ &0, 1, ..., 9, i_0, ..., i_{n-1}\}\end{aligned} \tag{3.11}$$

where n is the total number of information pieces in the information list. The result of this is that $|\alpha| = 4 + 10 + n$. Generalised we find that

$$|\alpha| = A + D + n \tag{3.12}$$

where T is the number of additional messages and D is the number of digits to transfer.

Considering a unidirectional system for communication, the message set can be broken into two messages set that overlap. R is the message set of the receiver of information and T is the message set of the information transmitter.

$$\begin{aligned} R = \{&request, begin, end, repeat, \\ &0, 1, ..., 9\} \\ T = \{&begin, end, repeat, \\ &i_0, ..., i_{n-1}\} \end{aligned} \tag{3.13}$$

$|T| = t + n$ and $|R| = r + D$ where t is the number of additional messages needed by the transmitter, and r is the number of additional messages required by the receiver.

Alice and Bob's Message Set

Alice's information list contains 10 pieces of information, such that $n_a = 10$. Alice and Bob wish to communicate bidirectionally, and have agreed on the additional messages $A = \{request, begin, end, repeat\}$. The message set that they will be using is

Additional Messages	Information	Digits
request	versailles	0
begin	women	1
end	bread	2
repeat	paris	3
	bastille	4
	economy	5
	rights	6
	terror	7
	razor	8
	dictatorship	9

This results in the alphabet

$$\begin{aligned}\alpha = \{&request, begin, end, repeat\\ &versailles, women, bread, paris, bastille,\\ &economy, rights, terror, razor, dictatorship,\\ &0, ..., 9\}\end{aligned} \tag{3.14}$$

resulting in

$$\begin{aligned}|\alpha| &= A + D + n\\ &= 4 + 10 + 10\\ &= 24\end{aligned} \tag{3.15}$$

α will be used in the creation of practical examples of visual and verbal communication systems.

3.4.4 A Visual Communication System

This section will propose a practical means for solving Alice's problem by using a physical communication medium. It will establish the encoded alphabet β and give examples of conversations between Alice and Bob using this alphabet.

Medium

The mediums used by this example are a hand h, a pen p and a highlighter l. Their state sets are as follows:

$$
\begin{aligned}
S_h &= \{open, closed, none\} \\
S_p &= \{penNorth, penSouth, penEast, penWest\} \\
S_l &= \{highlighterNorth, highlighterSouth, highlighterEast, highlighterWest\}
\end{aligned} \tag{3.16}
$$

These states can be seen in the figures below.

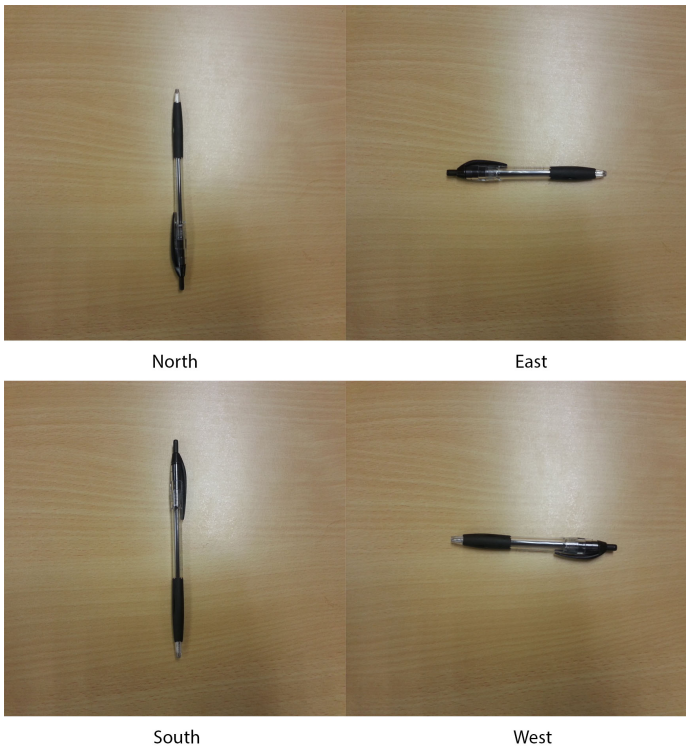

Figure 3.4: States of the pen medium

The β alphabet can be created by combining the S_h with either S_p or S_l.

$$
\begin{aligned}
\beta &= S_h \times (S_p \cup S_l) \\
\beta &= \{open, closed, none\} \times \{penNorth, penSouth, \\
&\quad ..., highlighterEast, highlighterWest\} \\
\beta &= \{\{open, penNorth\}, \{open, highlighterNorth\}, \\
&\quad ..., \{none, penWest\}, \{none, highlighterWest\}\}
\end{aligned} \tag{3.17}
$$

Figure 3.5: States of the highlighter medium

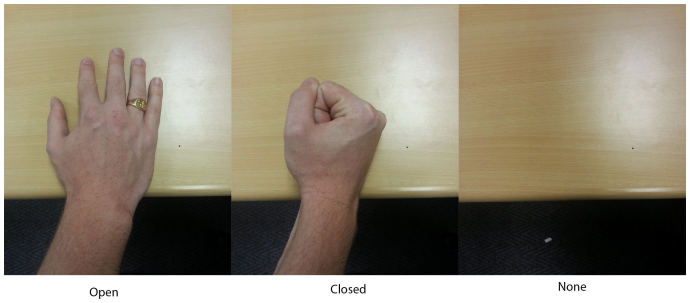

Figure 3.6: States of the hand medium

Encoded Alphabet

α can be directly encoded onto β in this case:

α Message	β Message	α Message	β message
request	open,penNorth	razor	closed,highlighterNorth
begin	open,penSouth	dictatorship	closed,highlighterSouth
end	open,penEast	0	closed,highlighterEast
repeat	open,penWest	1	closed,highlighterWest
versailles	open,highlighterNorth	2	none,penNorth
women	open,highlighterSouth	3	none,penSouth
bread	open,highlighterEast	4	none,penEast
paris	open,highlighterWest	5	none,penWest
bastille	closed,penNorth	6	none,highlighterNorth
economy	closed,penSouth	7	none,highlighterSouth
rights	closed,penEast	8	none,highlighterEast
terror	closed,penWest	9	none,highlighterWest

Communication Examples

The following are examples of communication between Bob and Alice using this system.

Request Info Alice wishes to know what information she must apply to question nine, to request this she uses the message sequence m, which can be seen in Figure 3.7

$$m = \{request, begin, nine, end\} \tag{3.18}$$

Bob wishes to let Alice to know that the information 'terror', and 'dictatorship' are relevant to that question. He can do so with the message set n shown in Figure 3.8

$$n = \{begin, terror, dictatorship, end\} \tag{3.19}$$

Figure 3.7: Alice requests information relevant to Question Nine

Figure 3.8: Bob responds to Alice's request

Repeat If Bob does not understand which question Alice has requested information for, he can respond with message set r shown in Figure 3.9

$$r = \{repeat\} \tag{3.20}$$

The same response can be used by Alice, if she does not understand the messages she receives from Bob.

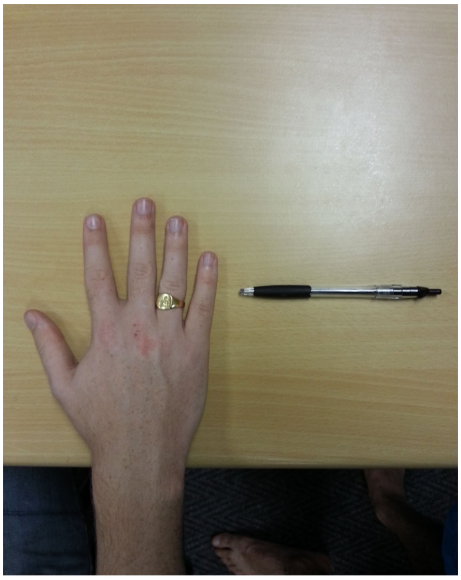

Figure 3.9: Bob asks Alice to repeat her request

Chapter 4

Conclusion

4.1 Importance of Definitions

The definitions of knowledge and information were of great importance to this paper. The understanding of the cheating problem that they provided allowed for an information based definition of cheating. From this definition we can establish categories of cheating and the motivation for their use. They also highlighted the importance of the information used to create knowledge in the mind, and the use of this information to increase the extent to which knowledge has been internalised.

4.2 Applicability of Steganography

The concepts explored by Steganography relate to those posed by the cheating problem. Both deal with the problem of communicating without detection, however Modern Steganography and Traditional Steganographic problems differ from the cheating problem.

Modern Steganography is primarily concerned with the avoiding detection by statistical analysis. It is no longer concerned with evading detection by human scrutiny. The result

of this is that most of the techniques discussed and explored in Modern Steganography are not applicable to the cheating problem.

The problems explored by Traditional Steganography, such as the Prisoners' Dilemma, are concerned with embedding information into a known medium, with a known probability distribution. They follow Kerckhoff's principle which is not compatible with the cheating problem. Cheating is concerned with the creation of its own probability distribution for its own medium, and the amount of suspicion that would be raised by that medium.

Thus while Steganography and cheating have the same objective their problems differ on a fundamental level.

4.3 Applicability Of Information Theory

The basic concepts of information theory proved applicable to the cheating problem. Coding theory is particularly applicable, as it allows for information to be encoded into a form that can avoid detection, such as a symbol set used in Information Smuggling or a language used by candidates. Other concepts such as Relative Entropy and the structure of a communication system were also relevant to the cheating problem.

4.4 Prevention

This paper has provided examples of systems that may be created using an information-theoretic understanding of the cheating problem based on of assumptions about the environment that assessment will take place in. These assumptions, such as examiner oversight and the proximity of partners, are the weakness of the systems being created.

Prevention of Information Smuggling is simply a case of combating examiner oversight. An examiner's awareness of the potential to store information in a seemingly information free medium would result in stricter action against suspicious mediums. To avoid

the complication of disadvantaging altogether, examinations could provide standardised stationery that if given to candidates as they enter the examination venue.

Communication systems can be made infeasible by restricting the knowledge of candidates about the venue in which they will be assessed. A randomised seating arrangement will be sufficient for this. If candidates cannot reasonably presume that they will be seated in proximity to each other they may be less inclined to attempt to cheat in this way. If they decide to make use of a communication system regardless, they are still faced with the probability that they may be unable to communicate owing to their positioning.

While these prevention techniques are already found in some examination environments, their value is emphasised by the understandings of cheating provided in this paper.

4.5 Further Research

When creating an information list a candidate is also legitimately preparing for the examination. The act of identifying important pieces of knowledge aids in its internalisation. This indicates the potential for an information-theoretic analysis of knowledge internalisation. Such an approach would allow for the quantification of the value of information, allowing for identification of key information and the relationships between information through measures such as entropy, relative entropy, mutual information and conditional entropy.

The variation of typical steganographic problems also hold potential for futher research. In an age of where vasts amounts of information is shared in the form of images and short text strings the potential exists for a steganographic system that does not encode information into the bitwise representation of data, but into the data itself. This opens the possibility for steganographic systems which hide information in a seeminly unintelligble mediums, such as spam advertising text or images.

The more advanced concepts of Information Theory, such as Mutual Information and Joint Entropy, were out of the scope of this investigation. Further research could be done

into the applicability of these techniques to the cheating problem.

Bibliography

Ackoff, R. L. 1989. From data to wisdom. *Journal of Applied Systems Analysis*, **16**, 3–9.

Anderson, Ross, & Petitcolas, Fabien. 1998. On The Limits of Steganography. *IEEE Journal of Selected Areas in Communications*, **16**, 474–481.

Becher, Tony. 1994. The significance of disciplinary differences. *Studies in Higher Education*, **19**(2), 151–161.

Becker, Tony, & Trowler, Paul R. 1989. Academic tribes and territories: Intellectual inquiry and the culture of disciplines. *UK*.

Bellezza, Francis S. 1981. Mnemonic devices: Classification, characteristics, and criteria. *Review of Educational Research*, **51**(2), 247–275.

Bellinger, Gene, Castro, Durval, & Mills, Anthony. 2004. Data, information, knowledge, and wisdom. Online. Available from: `http://www.systems-thinking.org/dikw/dikw.htm`. Accessed on: 31 Oct 2013.

Bennett, Krista. 2004. Linguistic steganography: Survey, analysis, and robustness concerns for hiding information in text.

Bierbrauer, Jürgen, & Fridrich, Jessica. 2008. Constructing good covering codes for applications in steganography. *Pages 1–22 of: Transactions on data hiding and multimedia security III*. Springer.

Bjorklund, Mikaela, & Wenestam, Claes-Goran. 2000. Academic cheating: frequency, methods, and causes. *In: Proceedings of the 1999 European Conference on Educational Research*. Education-line.

Cachin, Christian. 1998. An Information-Theoretic Model for Steganography. *Pages 306–318 of: Information Hiding*. Lecture Notes in Computer Science, vol. 1525. Springer Berlin Heidelberg.

Cover, Thomas M, Thomas, Joy A, & Kieffer, John. 1994. Elements of information theory. *SIAM Review*, **36**(3), 509–510.

Descartes, René. 1967. Meditations on First Philosophy, 1641. Online. Available from: `http://kernsphilosophypage.com/Modern/ModernPhilosophyReadings1.doc`. Accessed on: 31 Oct 2013.

Joyce, Donald. 2002. Cheating, Outsourcing, Plagiarism: A Growing Problem? *Pages 245–247 of: Proceedings of 14th Annual Conference of the National Advisory Committee on Computing Qualifications*.

Lelewer, Debra A., & Hirschberg, Daniel S. 1987. Data compression. *ACM Comput. Surv.*, **19**(3), 261–296.

Lim, Vivien K. G., & See, Sean K. B. 2001. Attitudes Toward, and Intentions to Report, Academic Cheating Among Students in Singapore. *Ethics & Behavior*, **11**(3), 261–274.

MacKay, David JC. 2003. *Information theory, inference and learning algorithms*. Cambridge university press.

Memon, Aasma Ghani, Khawaja, Sumbul, & Shah, Asadullah. 2008. Steganography: A new horizon for safe communication through XML. *Journal of Theoretical and Applied Information Technology*, 187 – 202.

Munuera, C. 2007. Steganography and error-correcting codes. *Signal Processing*, **87**(6), 1528–1533.

Murphy, Kevin P. 1998. Elements of Information Theory. *In: Advances in Knowledge Discovery and Data Mining*. John Wiley & Sons.

Pandey, Priti, & Zimitat, Craig. 2007. Medical students' learning of anatomy: memorisation, understanding and visualisation. *Medical education*, **41**(1), 7–14.

Petryszak, Nicholas G. 1981. Tabula rasa–its origins and implications. *Journal of the History of the Behavioral Sciences*, **17**(1), 15–27.

Provos, Niels, & Honeyman, Peter. 2003. Hide and seek: An introduction to steganography. *Security & Privacy, IEEE*, **1**(3), 32–44.

Rhodes Academic Administration. 2013 (May). *Invigilation of Examinations*. Obtained by Personal Email: 4 Sep 2013.

Schunk, Dale H. 2003. *Learning Theories: An Educational Perspective*. 4th edn. Prentice Hall.

Shannon, Claude E, & Weaver, Warren. 1949. The mathematical theory of communication. *University of Illinois Press*, **19**(7), 1.

Simmons, Gustavus J. 1984. The prisoners problem and the subliminal channel. *Pages 51–67 of: Advances in Cryptology*. Springer.

Simmons, Gustavus J. 1985. The subliminal channel and digital signatures. *Pages 364–378 of: Advances in Cryptology*. Springer.

Toohey, Susan. 1999. Beliefs, values and ideologies in course design. *Designing courses for higher education*, 44–69.

Wang, Feng-Hsing, Pan, Jeng-Shyang, & Jain, Lakhmi C. 2009. Digital Watermarking Techniques. *Pages 11–26 of: Innovations in Digital Watermarking Techniques*. Studies in Computational Intelligence, vol. 232. Springer Berlin Heidelberg.

Zins, Chaim. 2007. Conceptual approaches for defining data, information, and knowledge. *Journal of the American Society for Information Science and Technology*, **58**(4), 479–493.