Fumagalli, Anita

An Examination of the Validity of the Concealed Information Test

An Examination of the Validity of the Concealed Information Test

ANITA FUMAGALLI

A thesis submitted to the University of Huddersfield in partial fulfilment of the requirements for the degree of Doctor of Philosophy.

August 2020

ABSTRACT

This study investigates the validity of the Concealed Information Test (CIT), considered as the most accurate polygraph test able to discriminate guilty from innocent suspects (e.g. Ekman & O'Sullivan, 1991; Bull et al., 2004; Grubin & Madsen, 2005). Despite scientific support for the use of the CIT, the test is currently rarely employed, due to the lack of demonstration for its applicability within real forensic contexts (Kraphol, 2011). This thesis aims to provide further evidence of the CIT mechanisms and demonstrate new potential applicability as an investigative and preventative tool. This study addresses these aims by evaluating certain aspects of the validity of the CIT which have received less attention in literature: the effect of visual stimulation, the ability to detect criminal intentions and the application of the CIT in a group setting.

In order to conduct the analysis, the study recorded and analysed different physiological measurements: EDA (electrodermal activity), breathing and heart rate. Analyses were conducted to determine the accuracy of the test and the effects of the stimulations designed. Study I explored the basic mechanisms of the CIT and the impact of relevant information on participants' physiological responses. For this purpose, the study used a standard CIT examination, which involves only verbal stimulation. Results from this study demonstrated the efficacy of the standard procedure of the CIT. In fact, the test was able to detect relevant information from guilty participants (SCR $M=2.85$; RLL $M= -1.41$; Cardio $M= -0.5$) and identify correctly innocent participants (SCR $M=0.03$; RLL $M= 0.02$; Cardio $M= -0.07$), showing a significant difference between the two groups (d=3.03; p= <.005). In addition, it was found that stimulation characteristics could affect the process of recollection, determining the way in which certain relevant information was acquired by the participants. These initial but key conclusions have been used as a basis for developing the design of the later studies.

In order to demonstrate the applicability of the CIT in forensic settings, Study II investigated the effectiveness of an alternative stimulation method: visual stimulation. Despite the amount of photographic evidence available in forensic cases, not many researchers have explored the potential of visually stimulating examinees during a polygraph examination. Results

suggested that visual stimulation could increase the accuracy rates of the test. Specifically, the combination of verbal and visual stimulation had a stronger impact on participants' physiological responses when compared with Study I. Results also presented a significant difference (d=2.05; p= <.005) between the experimental (SCR $M=0.95$; RLL $M= -0.72$; Cardio $M= -0.54$) and the control group (SCR $M=0.02$; RLL $M= 0.04$; Cardio $M= -0.05$). Therefore, this study demonstrates the effectiveness of visual stimulation, proposing real potential applicability of the test in forensic settings with the use of photographic evidence.

With the aim of suggesting a new applicability of the CIT in forensic settings, Study III evaluated the effectiveness of the CIT in a more realistic setting by attempting to detect intentions to commit a crime. There is a lack of research around the potential applications of the CIT as a tool for preventing crimes, especially regarding suspect prioritization or planning prevention prior to a crime being committed. For this reason, Study III investigated the accuracy of the CIT when trying to detect criminal intentions after having planned a crime. Results from this study showed that the CIT was able to detect crime-related information from the experimental group (SCR $M=7.39$; RLL $M= -1.41$; Cardio $M= -2.43$), reporting a significant difference in the detection rate (d=7.35; p= <.005) compared to the control group (SCR $M=2.22$; RLL $M= -1.27$; Cardio $M= 2.58$). Compared to other mock crime designs (e.g. MacLaren, 2001; Meixner & Rosenfeld, 2011), this study was able to combine a mock crime design with visual stimulation and detection of crime-related intentions. This new design provided more information regarding the external validity of the test, but also highlighted the effectiveness of the CIT, particularly when detecting intentions to commit a crime. The innovative aspect of this method has implications related to crime prevention activities.

Finally, Study IV retained the methodological elements of Study III, but applied them within a group setting. Previous literature on deception (e.g. Granhag et al., 2015) exclusively focused on solo offenders. Therefore, there is a need to expand the research and develop new technologies for the prevention of groups of criminals as well. The purpose of this study was to examine whether crime-related information and intentions to commit a crime could be detected from a group of mock suspects. The results showed that the CIT was able to detect crime-related information from the mock suspects (SCR $M=1.15$; RLL $M= -0.73$; Cardio $M= -0.1$) with a significant difference (d=11.72; p= <.005) from the control group (SCR $M=2.22$; RLL $M= -1.27$; Cardio $M= 2.58$). The combination of a group setting and the application of visual stimulation to detect criminal intentions can be considered as an innovative addition to the standard setting of CIT examinations. This analysis suggested that the CIT can be effective as a preventative tool for groups of criminals.

A combination of SCR (Skin Conductance Response) scores, RLLs (Respiration Line Length) and heart rate was used for the analysis. Collective results from the studies concluded that the physiological measure with the highest detection rate was the SCR, compared to heart rate and RLL. This supports the current literature, which suggests that SCR is the most robust and accurate physiological measure to be used with the CIT (e.g. Bradley et al., 2011; Council, 2003b; Holden, 2000). However, conclusive arguments from the thesis took into account the combination of the three measures together, which generated larger validity rates for the test compared to the best scores from any single physiological measure.

Various factors limited the full accuracy rate of the test, causing the occurrence of false positive and false negative rates. Although some of the results were not optimal, this study provides a deeper understanding of the CIT mechanisms and proposes new potential applications. In conclusion, this study proposes the use of the CIT as a preventative tool within forensic settings; however, more research is required around the accuracy of the CIT in real settings, especially if applied to groups of examinees. With the benefit of future research, the CIT could support police investigations by detecting criminal information, prioritising and identifying potential suspects involved in solo or group criminal activities.

AKNOWLEDGEMENTS

This thesis was made possible by all the help and support I received from many persons. With this little section, I would like to acknowledge those who have been with me on this long journey.

I want to thank my supervisors Dr John Synnott and Dr Maria Ioannou for having supported me during these years. I am grateful for their professional and psychological support along the PhD journey. Their help was always accompanied by guidance, significant suggestions and a lot of encouragement. Both of them shaped my skills and my development as a researcher.

I want to thank my friend Calliope Tzani-Pepelasi, my PhD-sister during these last 3 years. She has been my spiritual companion, always ready to help me reacting towards dark moments and always ready to encourage me during these long British winters. Thanks for always supporting and encouraging me with a smile and infinite Greek humour.

I want to thank the community of Cafe Nero, in Huddersfield. Every single customer in that coffee shop experienced part of this journey. I have shared progress, concerns, failures and good news with most of them. Although some of them could not appreciate the importance or the meaning of my project, this little community supported me beyond expectations. Even though I did not appreciate working there, I will always keep with me the stolen conversations between a coffee and another one. Small words and shy smiles comforted and inspired me during these three years. Some of the souls from this coffee shop will always be with me.

I want to thank my little family in Huddersfield, which adopted and supported me during these years. My PhD journey was not the easiest one, but it was full of people that encouraged and inspired me along the way: Franci, Alice, Alessandro S., Stefano, Gabriele, Alessandro F., Katheek, Elena, Chiara, Ibra, Alan, Saul, Antonella and Maria. Every person contributed to the progression of this project, some with kind words, and others with delicious food.

I want to thank my new "family" at Coventry University. Although they faced the last phase, but not less difficult, of this journey, I received intense, daily, psychologically support. I am so glad I am working with this amazing team, which is helping me to refine my skills as a researcher, as a lecturer but also as a person. Although my PhD journey is almost finished, this team constantly encourages me to grow the enthusiasm for the new future journeys.

I want to thank my family-community back in Palermo. Thanks for having believed in me during these long years. Thanks for having me always with you, although the distance keeps me away from your daily lives. I miss you and I wish one day I would be able to come back home for good, with a big backpack full of amazing experiences to share.

I want to thank Alexandros who witnessed the making of this thesis. He supported me without reserves, beyond any patience possible. Every phase of this PhD had its own psychological breakdowns. Although Alex did not personally experience all of them, he tried to understand every difficulty and hard moment I had to face. I am so grateful he was with me during this part of the journey, the most difficult, frustrating and emotional of all. He was the best support and friend I could ask for.

Lastly, I want to thank my family, my mum, my dad and my brother, for their genuine, infinite and immense support during my experience in the U.K., particularly during my PhD journey. They always believed in me, and this gave me the strength to push myself beyond my fears, my comfort zone, and my limits. Thanks for your unconditional love. Thanks for your psychological and emotional support. Thanks for being always there when I need you. Nothing of this would have been possible if it was not for you and your love.

Finally, I want to thank my Grandma Pina who was physically with me only for half of this journey, but spiritually in every single word of this thesis. Although the distance was a heavyweight for her, she always appreciated the effort of my life project. For this reason, she always encouraged me to take all I could get from this experience. Although we do not live in the same world anymore, I am sure she helped me to arrive at the end of this journey, “safe and sound”. Thanks for your hugs, warm food and immense love after every return. I hope you are proud of me. This thesis is for you.

A new journey is waiting for me, and I look forward to it!

Anita Fumagalli

Coventry, 29/10/2018

Table of Contents

List of Tables

List of Figures

To “Nonna” Pina

CHAPTER 1: INTRODUCTION

1.1 Background of the project

The polygraph is one of the most discussed and debated tools of the last century. Although at the beginning of its use the polygraph was hailed as a revolutionary tool able to detect deception, many doubts were raised due to the lack of scientific research that could support its reliability (OTA, 1983). Polygraph supporters believe that this tool can detect lies through changes in physiological reactions; opponents think, instead, that the polygraph is not a reliable tool, as the results produced by the examination can be easily controlled and manipulated by the application of simple countermeasures. Although the limits of this machine are widely known, the polygraph is used across 70 countries (Barland, 1999). The USA, for instance, is the biggest supporter of polygraph technology; in fact, many federal and local law enforcement agencies use the polygraph in criminal investigations (Beardsley, 1999). Furthermore, the application of the polygraph is also used for pre-employment screening of police applicants to select the most suitable ones. It is also utilized in the private sector, where employers can request a polygraph examination for security reasons (Segrave, 2004). Usually, screening tests are used for investigating the applicant's history of involvement in a range of possible activities that might concern the employer and, therefore, affect the hiring process (Handler, Honts, Krapohl, Nelson, & Griffin, 2009).

In the last decade, the polygraph has also been used in the treatment and supervision of community-based sex offenders (Holden, 2000). In fact, in a forensic-clinical setting, the polygraph has been used during surveillance practices in post-conviction stages, in order to monitor sex offender activity; in addition, it can be helpful in the prevention and detection of reoffending if or when it occurs (Grubin, Madsen, Parsons, Sosnowski, & Warberg, 2004). In July 2013, new legislation was introduced in the U.K. which regulates the use of the polygraph on sex offenders for determining their risk of reoffending.

In legal settings, the admissibility of polygraph evidence varies between different countries, for example in the U.K. results from polygraph examinations cannot be considered admissible as evidence, with the only exception being certain serious sex offenders' cases. Conversely, polygraph evidence can be admissible in court in 19 American states. However, such evidence can be accepted at the discretion of the trial judge or when both parties agree to admissibility before the polygraph tests being conducted (Daniels, 2002).

In 1966 the American Polygraph Association (APA) was established, which is the largest professional association for polygraphy in the world. This body provides a code of ethics, standards of practice and procedural guidelines. In the USA, in order to be accredited, polygraph examiners must be registered with the APA.

Polygraph administration includes the application of different tests. The most widely used method is the Control Question Test (CQT; see Chapter 2, section 2.2.5) (Iacono, 1991). However, the Concealed Information Test (CIT; see Chapter 2, section 2.5.6) finds systematic use in Japan and in Israel, where it is considered the most reliable test among the others (Elaad, Ginton, & Jungman, 1992; Matsuda, Nittono, & Allen, 2012). Although the CQT is extensively used in a number of different contexts around the world, the CIT is the polygraph test most supported by the scientific community, due to higher validity supported by theories and research. For this reason, much research has been conducted on the validity of the CIT, trying to demonstrate its validity and reliability.

Only a few countries have adopted this test for forensic or clinical purposes, such as Japan. The use of the CIT was introduced into Japanese police agencies in 1956; since 1970, more than 5,000 CITs have been performed each year (Matsuda, Nittono, & Allen, 2012; Osugi, 2011). In Japan the CIT is used for criminal investigations, especially when trying to determine if a suspect possesses crime-related information. Although the CIT is usually used for testing suspects, the polygraph is also used as a preventative tool for false accusations (Osugi, 2011).

Although the CIT is considered one of the most valid and accurate tests (Ben-Shakhar, 2011), it has received less consideration compared to other polygraph tests. It has been suggested that one of the main reasons why Western countries choose not to adopt the CIT is related to the misbelief that the CQT is equally reliable and easier to apply in a practical context (Verschuere, Ben-Shakhar, & Meijer, 2011).

For this reason, the current project aims to evaluate the validity of the CIT by providing an additional scientific contribution to the existing literature. Due to the lack of awareness about the efficacy of the CIT and the lack of communication between the research and practitioner community, this study will try to explore if the CIT is a reliable tool that can be applied to many purposes.

1.2 Thesis structure

The first chapter included in this thesis (Chapter 2) will present the general theoretical background of this project. Specifically, it will evaluate the different existing polygraph tests: the Relevant/Irrelevant Test, the Control Question Test and the Concealed Information Test (CIT). However, specific attention will be dedicated to the evaluation of the CIT, which represents the focus of this research. Therefore, an accurate analysis of the mechanisms of the CIT, its theoretical assumptions and any practical and theoretical limitations associated with it will be discussed and critically evaluated. This analysis is necessary as it will be crucial for estimating the validity of the CIT. Different issues associated with the application of the polygraph will be covered, addressing specific limitations and also the generalizability of laboratory studies applied to real-life scenarios. One of the sections within this chapter will provide details about the theoretical framework of the processes underlying the mechanism of the examination; this will include cognitive and physiological processes associated with deception, habituation and the orienting response. Furthermore, factors influencing physiological reactions and factors that can affect the detectability of the test will be discussed. Finally, the last section of the chapter will evaluate concepts of memory associated with deception. The CIT's mechanism is based on a recognition process, which involves different memory components. Therefore, it is essential for this thesis to provide an overview of the most crucial memory processes involved when trying to store, recollect and hide specific information that the CIT is trying to investigate. In this way, it will be possible for the reader to have a better understanding of the cognitive mechanisms associated when trying to undertake deception. At the end of this chapter, in section 2.4, a specific discussion around the purpose of the thesis will be provided, presenting the aims of the project and of the single studies.

Chapter 3 will provide detailed information about the design of this project and the methodology that was applied to the studies conducted. This chapter presents an inspection of the methodological strategies and procedures implemented when performing polygraph examinations. As every study presents specific differences in the process and some aspects of the methodology, every study chapter (Chapters 4, 5, 6, and 7) will give background regarding the purpose of the experiment, information about the sample used, specification regarding the method used, and a results and discussion section.

Finally, a general conclusive chapter will address if the purpose of this study has been achieved and how; this will help the reader to have an overview of the aims achieved within this study. Limitations that might have occurred will be highlighted and discussed; finally, the chapter will consider the implications of this study and suggestions for future research.

CHAPTER 2: DECEPTION AND POLYGRAPH

2.1 Lies and Deception

"First there was the Word. And then there was the lie".

(Lynch, 2009)

Lying and deception have always been part of human behaviour. Lying represents one of the most typical features categorising human behaviour since the beginning of time, from the oldest tale in the world (Adam and Eve) to the Watergate scandal. Entire stages of human history were based on lies. For example, Nazi propaganda, which succeeded in convincing the German people that Jews were evil; or the Piltdown Man, a skull which was believed to be related to an extinct human ancestor that, almost a century later, was revealed to be from an orangutan (Walsh, 1996). Some of the most important narrative poems celebrated lies and deception, like in Homer's Odyssey with the Greek's ingenious Trojan horse, or the extraordinary description of Malacoda in Dante's Inferno in the Divine Comedy.

Lying is a natural act, a social strategy that individuals apply in their everyday life. There is no universal definition of lying; it is a term that has always been subject to different interpretations. Generally, a lie is a deliberately false statement that the speaker states to be true. Therefore, any lie violates an implicit guarantee that what someone says is true. An individual, thus, intends to lie when inviting others to trust and rely on what they say by ensuring its truth; at the same time, the intention is to betray that trust by making a false statement (Carson, 2009). On the other hand, the concept of deception is slightly different: it is considered as the intent to cause someone to have false beliefs. Deception implies success; in fact, a statement that is not believed by others could still be considered a lie (Carson, 2009). However, while a lie must be a false statement, deception does not necessarily involve making false statements. As such, true statements can be deceptive (Lynch, 2009).

Saint Augustine, one of the most revolutionary philosophers, tried to investigate if the definition of a lie includes the intention to deceive; in fact, he stated that speaking a falsehood with the intention to deceive is considered a lie. However, he was not sure that in the case where there is a lack of intention, the statement should still be considered a lie (Augustine, 1952).

Saint Thomas Aquinas improves upon Saint Augustine's philosophy on lying, affirming that evil acts like lying are always considered immoral, regardless of the intentions or the circumstances (Aquinas, 1972). While Plato suggested the concept of the "noble lie" as a justifiable deception that benefits the collective, Kant considered all forms of lying inexcusable (Martin, 2009).

Philosophers and psychologists are still debating today how to define these two concepts. However, the impulse to deceive others seems to be an innate instinct that is also common in animals (Mitchell & Thompson, 1986). Contrarily to humans, deception in animals might not occur as a conscious act, and it can be implied at different levels of cognitive ability. Some animals have developed a structural method of deception by appearing like something that they are not, such as the environment around them. Certain species of animals, like iguanas or chameleons, are able to mimic and camouflage by using a combination of materials and colours to help conceal themselves by making it hard for others to see them (Ford, 1996). In other anti-predator strategies, animals like lizards, birds and sharks may also feign death in order to escape predation or reduce the probability of attacks (Mitchell & Thompson, 1986). Furthermore, great deceptive behaviours have been observed in monkeys, which are able to utilise vocal forms of deception: for example, they can alert the other members of the group with vocal sounds and trick them into thinking that there is a predator sighting. They can then take advantage of their distraction to steal their food (Wheeler, 2009). Therefore, the intention to mislead others seems to be a universal behaviour.

Across history, humanity has tried to understand and explain the deep mechanisms that push individuals to lie and deceive each other. Although there are no universal explanations for this archaic behaviour, the following sections will outline the historical overview of lie deception and some of the most relevant theories that have tried to explore deceptive behaviours.

2.1.1 Theories of Deception

The process of lying has been described as a deliberate choice to mislead a person. However, the target person must be unaware of the intention to mislead, and must not have consented to be deceived (Ekman, 1992). Deception and lying are an active part of our daily life. According to a survey conducted by Patterson and Kim (1991), it was found that 90% of people in North America admitted to having lied in their life. Diary studies employed to record people's every day deceptive communication showed that about 20% to 33% of our daily interactions are false (DePaulo, Kashy, Kirkendol, Wyer, & Epstein, 1996a; Hancock, Thorn-Santelli, & Ritchie, 2004). Even though deception is a standard feature, some researchers have indicated that people do not lie to the same degree. In fact, in a study conducted by Serota, Levine and Boster (2010) it was shown that on average participants told lies once or twice a day.

The frequency of lies differs depending on the type of relationship you share with the receiver. In fact, it was demonstrated that people lie less to close partners compared to casual acquaintances and strangers (DePaulo, Kirkendol, Tang, & O'Brien, 1988). However, in a study based on adultery, it was shown that between two-thirds to three-quarters of married people in a North American and British sample had lied to their partners about their extramarital affairs (Lawson, 1988).

Deception is significant because of the negative consequences that may result from it; nonetheless, some authors have considered the process of lying as a necessary social dynamic, like a "social lubricant" (Vrij, 2008). Different reasons push people to lie to each other. People may lie to gain material advantage or to avoid materialistic loss; others might lie for fear of punishment (Vrij, 2008) or to protect themselves, some merely to amuse the sender (Ekman, 1988). For example, studies have demonstrated that people usually lie when applying for jobs or promotions; they lie as a strategy for creating collaborations or resolving conflict in the workplace (Kashy & DePaulo, 1996). In other cases, lying can be used to develop social relationships. Lies can be created to avoid embarrassment for themselves or others (Kashy & DePaulo, 1996).

2.1.2 The self–presentation theory of deception

The majority of lies are not planned in advance. According to the *self-presentational theory of deception* (DePaulo et al., 2003) lies arise from spontaneous interactions as a result of immediate demands of the conversation. Lies can have different shapes and forms; they differ in content, in type and importance. For example, *white lies* are told to protect other people's feelings; this refers to the kind of deception that is usually acceptable, as it is used by individuals to protect others (Sullivan, 2001); on the other hand, outright lies are used to avoid telling the truth.

In deception literature, it has been found that people are rarely able to distinguish lies from honesty. However, a meta-analysis by Bond and DePaulo (2006) reported that although discrimination accuracy is inferior, people can detect deception at a small but significant level of accuracy. The same study has also demonstrated that one of the causes of low detection accuracy might be related to problems with ecological validity. In fact, research found that more motivated liars were detected significantly more often than unmotivated liars (Vrij, 2008).

2.1.3 The non-verbal leakage theory

When trying to detect liars, verbal and non-verbal cues can play a central role. The most common non-verbal cues include gaze aversion, increased movement, more frequent and longer pauses, slower speech, hesitations and speech errors (DePaulo et al., 2003; Ekman, 1992). These type of behaviours are commonly thought to be indicative of anxiety; in line with this belief, research has demonstrated that liars tend to speak in a higher-tone voice, displaying longer pauses during conversations (Vrij, 2000). There is substantial literature on verbal and non-verbal behaviours associated with lying. Ekman and Fresen's *non-verbal leakage theory* argued, for example, that when liars are highly motivated to deceive, they experience anxiety, higher arousal levels, shame and excitement (Ekman & Friesen, 1969). They also claimed that the face is the easiest part to control because it provides high levels of feedback. On the other hand, the body and limbs are considered harder to control, therefore they are most likely to show more non-verbal signs of arousal. In fact, the attempt to control these behaviours could be extremely difficult, resulting in over controlling (e.g. fewer movements of the body, and shorter response durations) (Vrij, 2000). Nonetheless, there is very little evidence that supports these claims.

2.1.4 The motivational impairment effect

Another theory that tries to explain how physiological cues are associated with deception is the *motivational impairment effect* (DePaulo et al., 1988): this approach suggests that when liars are highly motivated, they will experience a reduced control over leakage cues, increased behavioural rigidity and improved verbal performance. Research indicates that this physiological change will increase the likelihood of detection.

It has also been found that deception is a more cognitively demanding task than truth-telling (Anolli & Ciceri, 1997; Vrij, Akehurst, & Morris, 1997). In fact, extra cognitive resources are necessary for the construction of lies, remembering details to ensure coherence, monitoring others' reactions, suppressing the truth while they are lying (Vrij, 2008). This cognitive process will be discussed in more details in section 2.6.2.2.

As mentioned in the previous section, research indicates that humans are not particularly skilled lie detectors. According to many researchers (DePaulo & Rosenthal, 1979; Vrij, 2000), the ability of the average person to catch a liar is typically at the level of chance. Even police officers, who should be trained to be more sensitive to deceptive behaviour, performed no better than at chance levels (Vrij, 2008). In support of this, a study demonstrated that customs officials were no more accurate than university students during mock crime examinations (Kraut & Poe, 1980). Similarly, no difference was reported between federal law enforcement officers and students in detecting deception (DePaulo & Pfeifer, 1986). A study was conducted by Ekman and O'Sullivan (1991) to investigate the skills of lie catchers. They studied nurses that were instructed to conceal their negative or positive emotions after a pleasant and a non-pleasant movie. Once the nurses had been interviewed, video recordings of the interviews were shown to police officers, court judges, secret service agents, polygraph examiners, and finally to non-specified professionals. The study found that professional lie detectors, with the exception of secret service agents, succeeded in detecting deception just slightly better than chance, and no better than non-professional lie detectors.

According to the motivational impairment effect previously discussed, the ability to catch liars might also be related to the likelihood of verbal and non-verbal cues being presented. In fact, it was suggested (DePaulo et al., 2003) that highly motivated liars might be expected to be more detectable than those with little or no motivation to avoid detection, only because they might tend to show more deceptive cues. Results from a meta-analysis, where the motivation of liars was manipulated, found that highly motivated liars were more often detected than unmotivated liars. Also, in studies where an incentive was given, such as money, more non-verbal cues were reported (Bond & DePaulo, 2006). Therefore, according to the available research, the more motivated liars are in order to avoid being caught, the more their behaviour will reveal their deception.

Deception represents one of the oldest social behaviours adopted by individuals. Theories have extensively studied this phenomenon, and they all agree that deception is an intention that sometimes manifests through the body. Therefore, in certain cases, it is extremely difficult to control or conceal emotions. Many theories have tried to explain how this process occurs, and which cues are more commonly associated with deception. These theories represent the theoretical grounds of this research, which explains why the polygraph mechanism was developed.

2.2 The Polygraph

2.2.1 The Origins of the Polygraph

Since the earliest times, lying has always been part of the human behaviour. This aspect of human action has always attracted the interests of people, who have tried to develop methods and techniques that can detect a liar. In order to detect liars, ancient societies developed creative procedures based on torture or other forms of trials. Most of them were daily used in ancient Greece, pre-Christian Scandinavia, Iceland, Polynesia, Japan and Africa. In Africa, for instance, honest men were expected to be protected by God, and therefore be able to hold their arm in boiling water for longer than a liar (Grubin & Madsen, 2005).

In Scandinavia, people were using a specific type of test for sexual misconduct. For example, if a woman was accused of adultery, she was forced to hold a red-hot iron for a short time: if her hands burnt, she was considered guilty (Sullivan, 2001). In northwest Africa, a fang from a snake was inserted under the eyelid of a suspect. People believed that a truthful person would be able to eject it by rolling the eye. Another interesting method of lie detection was used in Israel, where a donkey's tail was coloured with black ink and placed in a darkened room. The suspect had to enter the room and pull the donkey's tail. Once the suspect returned, his or her hands were checked; the suspect would be considered guilty if he came out of the room with clean hands (Larson, 1932).

An unusual combat procedure was used in Germany in the Middle Ages to settle allegations of infidelity. The husband would be placed waist deep in the ground with a club in his right hand. The wife would wear a long-sleeved shirt with a rock sewed in the inside one of the sleeves. The wife would then attempt to hit the restrained husband, while avoiding his attempt to beat her. The one able to survive the combat was considered faithful (Larson, 1932).

Other detecting deception techniques were focused more on the stress imposed on the individual than the psychological impact of the task. For instance, in China, it was common to make suspect liars chew rice powder while asking questions related to the issue investigated. Subsequently, the rice was examined. They believed that salivation was linked to the emotional state of the individual. Therefore if the powder was dry, the person was judged to have been lying (Sullivan, 2001).

Another notable example relates to Romans: people used to apply a specific method of selection of the bodyguard candidates, asking provocative questions. Those who blushed were selected for the job. The concept behind this technique was that if a person blushes in response to provocative questions, most likely the candidate is not participating in conspiracies. A more rigid method was used in Sparta, where young men were forced to stand on the edge of a cliff and were asked if they were afraid. Greeks believed that pale men were lying, so they were pushed from the cliff (Grubin & Madsen, 2005).

The idea that emotion is associated with deception has existed since the Middle Ages. In fact, the measure of the pulse was used for detecting a wife's faithfulness to the husband. In order to uncover the wife's affair, an advisor needed to measure her pulse while a name of a potential lover was presented (Trovillo, 1939). This method is one aspect on which modern lie detector tests are based. At the beginning of 1900, an Italian criminologist called Cesare Lombroso developed a more contemporary approach, as he was the first scientist who adapted the plethysmograph to monitor changes in blood volume during investigations. In 1914, an Austrian psychologist called Benussi started to investigate the relationship between lying and multiple physiological measures. The results of his studies stated that a lie was accompanied by a change in the ratio of exhalation to inhalation (Larson, 1932). This represented a remarkable discovery, as for the first time real scientific evaluations were provided to prove the relationship between physiological responses and emotions.

2.2.2 The origins of the polygraph

The first American pioneer of lie detection was Hugo Munsterberg, who in 1908 published a book that focused on the importance of investigating the physiological correlations of deception. In 1915, a student of his, William Marston, who was credited with coining the term *lie detector* (Bond & DePaulo, 2006), devised the systolic blood pressure deception test. The equipment included the intermittent recording of a suspect's systolic blood pressure during questioning (Grubin & Madsen, 2005). Marston was one of the most influential promoters of the polygraph. In fact, he promoted it through different advertisements and aids such as the Gillette Blade, a new type of razor blade that became one of the most successful and famous aids in which the polygraph was inserted. He also advertised the polygraph in Look magazine, highlighting the practical value of the lie detector in marriage guidance counselling (Bond & DePaulo, 2006).

Furthermore, Marston supported the use of his lie detector in criminal investigations and court cases, and he was the first person to attempt to submit the results from a lie detector test as official evidence in a court case (Waid, Orne, Cook, & Orne, 1981). The case in question relates to James Frye, who was accused of robbery and murder. According to Marston (1938), Frye initially denied the offences he committed; however, a few days later, he withdrew the confession admitting to the crimes. After utilising a polygraph, Marston concluded that Frye was telling the truth. Nonetheless, the judge refused to accept Marston's results as testimony for that case, because it was stated that the lie detector test had not gained sufficient acceptance within the scientific community (OTA, 1983). This case, thus, became a precedent known as the Frye Test in 1923, which became the first barrier to the inclusion of polygraph evidence in American courts.

In 1921, John Larson created the first modern polygraph instrument for use in criminal investigations. Although this work was influenced by Marston's studies, the instrument included records of pulse rate, blood pressure, and also respiration. By applying that device, he developed a new interviewing technique named the Relevant/Irrelevant Test (RIT) (Larson, 1932). Despite the success of this device, Larson was developed himself some serious doubts about the accuracy of the polygraph. In fact, he published one of the first studies on polygraph evaluation, in which he concluded, “The deception test alone should never be used as court evidence” (Bradley & Ainsworth, 1984, p.896).

After Larson, Leonarde Keeler represented the second most prominent polygraph examiner. In 1939, he developed two significant improvements to Larson’s polygraph: metal bellows, to improve the recording of changes in blood pressure, pulse and respiration, and a kymograph, which allowed chart paper to be pulled under the recording pens at a constant speed. Furthermore, Keeler added a third physiological component that was able to measure changes in a person’s skin resistance, the Psychogalvanometer. He justified this addition stating that a deceptive person would sweat more than a truthful person. This physiological process would cause a decrease in skin resistance, due to a higher concentration of negatively charged chloride ions on the surface of the skin (Grubin & Madsen, 2005). Due to the importance of his contribution, Keeler’s device was the first instrument adopted by the Federal Bureau of Investigation (FBI). In addition to that, he established the first polygraph unit with the Chicago Police department. That followed the bloody gang war between the gangs lead respectively by Al Capone and George Bugs Moran, which forced the police to add new techniques to fight crime (O'Toole, Yuille, Patrick, & Iacono, 1994). Lastly, Keeler is credited with developing the Concealed Information Test (CIT; see Chapter 2, section 2.2.6).

2.2.3 The development of a successful application

In 1938, collaboration was established between Keeler and John Reid, where both men founded the first polygraph schools in America, which were proving to be a very lucrative business. In 1942, a standard polygraph course became available, and consisted of a two weeks’ training. A few years later the training was increased to six weeks (Sullivan, 2001). Today, the American Polygraph Association’s (APA) accredited course lasts eight weeks. Reid developed the Reid Polygraph, an equipment that also included the recording of muscular activity in the forearms, thighs and feet, through metal bellows placed under the arms and the chair’s seat (Grubin & Madsen, 2005).

During the 1940’s, the use of the polygraph by law enforcement agencies increased dramatically. For example, in 1945, the American government utilised the polygraph to screen

German candidates to create a police force in post-war Germany. The polygraph program, conducted by Keeler, succeeded in identifying many Nazi members, sympathisers and different criminal activities. However, the program was subsequently abandoned (Sullivan, 2001). Another screening program was conducted at the Oak Ridge facility, the place where the atomic bomb was created. This program identified various security breaches, disclosures of tool thefts and unauthorised disclosure of classified information (Matte, 1996). Despite the fact that the RIT had become the standard questioning procedure in the early 50', different criticisms emerged; in fact, this new technique was subsequently considered not adequately standardised, thus unreliable (Sullivan, 2001). In order to overcome the limitations of this test, Reid developed a new procedure in which comparison questions were included. This new technique was named Control Question Test (CQT). Later on, it was changed to Comparison Question Test (CQT), as the "control" question was not considered as a control in the true scientific sense (Lykken, 1988).

The contribution of Reid also consisted in highlighting the importance of the behavioural features when trying to determine the truthfulness of a suspect. In fact, he described some of the deceptive behaviours to be noticed and included in the examination, such as appearing nervous, angry or avoiding eye contact (Sullivan, 2001). In addition, Keeler and Reid emphasised the role of the examiner as the real lie detector, guided by his or her background and training (Reid & Inbau, 1977). However, this approach was criticised not just for its lack of standardisation, but mainly because of the assumptions that specific behaviours are uniquely associated with deception (Backster, 1963a). To overcome the problem, Backster (1963b) developed the first numerical system for evaluating polygraph charts, which made the polygraph examinations' results more scientific and standardised.

During the 1950's, an increase in the polygraph use was reported, due to the widespread use not just in the United States (US), but also in countries like Japan, Israel and Korea (Matte, 1996). By the beginning of the 1960's, the US federal agencies were conducting around 19,000 polygraph examinations per year (OTA, 1983). On top of that, the polygraph started to be used by the private sector. Specifically, the polygraph was used to examine whether applicants were telling the truth on their applications or if they were suitable for a job basing on their answers (OTA, 1983). Also, the polygraph was used to periodically screen employees to investigate if they were stealing from the company or if they were involved in some criminal activity (Lykken, 1998).

2.2.4 The first doubts about the validity of the polygraph

In 1965, the Committee on Government Operation evaluated the use of the polygraph for screening. They reported that there was no scientific evidence to support that type of application, and they considered the polygraph inaccurate and inadequate (United States Congress, 1965). Despite different efforts were made to restrict or ban the polygraph, by the 1970's, employee screening became a multi-million dollar business (Lykken, 1974); the majority of federal agencies established a staff of polygraph examiners, conducting around 23,000 polygraph tests (OTA, 1983). Post-conviction polygraph testing was informally initiated, as the value of the polygraph use was also recognised in the supervision and management of probationers (Holden, 2000). Because of the increase in the use of the test, the Committee of Government Operations in the US House of Representatives requested to the Office of Technology Assessment (OTA) an evaluation of the scientific evidence for the polygraph. It was reported that the accuracy of the polygraph was still unclear; therefore, its utilisation in screening was not supported. In addition, the report highlighted the polygraph's limitations to countermeasures (strategies used to pass a polygraph test) and percentages of false positive and negative (examinees erroneously identified as truthful or deceptive) (OTA, 1983).

During the 1980's, general concern about the use of the polygraph started to spread across the scientific community. Specifically, the case of Floys Buzz Fay generated negative feelings, providing an example of the real practical and theoretical limitations of the polygraph. Fay was arrested after an armed robbery in which an office worker was killed. Although Fay was accused and arrested, the prosecution was unable to provide any corroborative piece of evidence. Therefore, they proposed him a deal: the charges would be dismissed if he could pass a polygraph exam. Fay accepted the proposal; however, he failed the test with the consequence of being convicted to a life sentence for aggravated murder. After two years, a man confessed he had committed the robbery, so Fay was released (Sullivan, 2001). The results from the OTA report combined with the concerns generated by Fay's case and by the scientific community, pushed President Reagan to enforce the 1988 Employee Polygraph Protection Act. This legislation restricted the use of the polygraph mainly to the private sector; however, federal agencies continued to increase the use of the polygraph and expanded the Department of Defence's polygraph program; that allowed the screening of federal employees to detect potential spies (Council, 2003b). It was possible to see the beneficial effects of this legislation in the case of Harold Nicholson, a Russian spy that was found deceptive after a polygraph test and, consequently considered an official spy (Segrave, 2004).

In 1999, because of a significant breach at the Los Alamos National Laboratory, reportedly committed by a Chinese spy, polygraph tests were ordered for all the laboratory's nuclear weapons scientists (Segrave, 2004). That measure created a new debate on the accuracy of the tests, and specifically on whether federal agencies should use them on their employees. Consequently, the Department of Energy requested to the National Academies of Science (NAS) to conduct an assessment on the validity of the polygraph when used in personnel security screening settings. The NAS's evaluation reported polygraph's accuracy rates falling between 81% and 91%. In addition, it was suggested that the polygraph tests were able to discriminate deceptive and truth tellers at rates well above chance. This conclusion caused significant concerns around the application of the polygraph to screening settings (Grubin & Madsen, 2005).

In 1993, the Frye test was displaced by the US Supreme Court's decision. From this decision, general guidelines were provided regarding the admissibility of contentious scientific evidence, the reliability and the general acceptance of the technique. This measure allowed courts to make decisions on the admissibility of polygraph evidence on a case-by-case basis (OTA, 1983).

In the United Kingdom (UK), the polygraph and its use have not been officially accepted. However, different evaluations were conducted during the 1980's (Bull, Baron, Gudjonsson, Hampson, Rippon, & Vrij, 2004). Similarly to the US, the British Government requested an assessment of the polygraph use after an episode in which the machine was able to discover the identity of a spy (Segrave, 2004). Therefore, the British Psychological Society (2004) investigated the reliability and validity of the polygraph, also taking into consideration the ethical concerns associated with the test. The report stated that at that time the polygraph tests were not adequately standardised, thus considered scientific tests. The lack of empirical evidence that might support the polygraph's accuracy was also highlighted. In addition, the report stressed how some of the methods applied during the polygraph administration were contrary to the principle of the British Psychological Society's code of conduct, such as provoking stress in examinees and misleading them about the effectiveness of the procedure. Although the review expressed negative opinions about the reliability of the CQT and the

Relevant/Irrelevant Technique (R/I), the CIT received some scientific support, based on the theoretical framework. However, the British Government decided not to accept the polygraph for personnel screening and investigations.

Despite the polygraph is widely used in some countries, it has always received criticism within the scientific community. However, the main critiques are always directed to the CQT (Ben-Shakhar, 2002; Furedy & Ben-Shakhar, 1991; Lykken, 1974): this debate will be discussed in more details in Chapter 2, section 2.5.5.2, dedicated to the validity of the CQT.

2.2.5 The study of the Validity of the Polygraph

The polygraph is commonly known as "lie detector" because of the misleading assumption that it is able to detect lies. However, the underlying theoretical premise is that the polygraph is able only to detect arousals, which are assumed to be related to the process of lying (Bashore & Rapp, 1993; Bull et al, 2004; Saxe, 1994). The role of the polygraph, thus, is to try to detect deception through the measurement of physiological reactions. This indirect process is necessary due to the impossibility to establish a physiological pattern related to deception (Saxe, 1994).

The theoretical assumption upon which the polygraph tests are generally based is that deceptive individuals will likely experience different emotions than a truthful one (Patterson & Kim, 1991). Consequently, the power of these emotions is strongly correlated to the likelihood of detecting deceptive elements in physiological output. Therefore, emotions such as fear and stress are the most common examples associated with the concern of getting caught, usually resulting in different physiological changes in the body (Synnott, Dietzel, & Ioannou, 2015). However, the relationship between the emotions and physiological reactions are more complex than a simple group of a pattern associated with deception. Therefore, it has always been difficult to provide precise levels of accuracy when evaluating polygraph tests, due to the complex mechanisms involved during the examinations.

Determining the accuracy of the polygraph has always been a critical debate in the scientific community. This occurs because many studies claim a high, almost perfect, accuracy rates; while others have rejected these conclusions (Iacono & Patrick, 1997). Trying to evaluate the validity of the polygraph test consists in investigating whether it can provide information about deceptive examinees that cannot be obtained from other techniques (Council, 2003b).

Before evaluating the scientific evidence of the polygraph, it is essential to define some key concepts that are important in the analysis. The idea of *validity* refers to whether an instrument measures what it is supposed to measure (Bryant, Grimm, & Yarnold, 2000). In this specific case, therefore, whether the polygraph is able to discriminate between physiological responses. However, different forms of validity need to be distinguished. The most important is the *construct validity*, which indicates the theory of how and why a specific test works. The construct validity is very crucial, especially in psychology research, as it provides the degree of how well explanatory theories and concepts account for the performance of a test (Bryant et al., 2000). Regarding the polygraph, for example, when lying provokes psychological arousal, which creates physiological changes. The polygraph measures the physiological responses that correspond with galvanic skin response, respiration, heart rate and blood pressure. These measurements are then processed, combined and scored to compute an overall index, which is used to make a final judgement about examinee's truthfulness. The validity of psychophysiological detection of deception depends on validity all along this chain (Council, 2003b). However, in polygraph settings many elements from the theory reported above could limit the construct validity; for example, psychological changes that correlate to psychological arousal vary across individuals. In the same way, there is a lack of scientific evidence that supports the theory that deception has consistent psychological significance for every individual (Council, 2003b).

Another form of validity is *criterion validity*, which refers to how well a measure matches a phenomenon that the test is intended to obtain. In the case of the polygraph, the actual deceptiveness of truthfulness of examinees on the relevant questions in the test (Council, 2003b). The term "accuracy" is typically used as a nontechnical synonym for criterion validity, and it will be used like that in this thesis.

In order to measure criterion validity, it is necessary to choose the appropriate criterion, which depends on whether the polygraph is being used for event-specific investigation, employee screening of pre-employment screening (Council, 2003b). In the case of the present project, the polygraph test needs to determine the examinee's truthfulness about a specific event or task. The validity of the polygraph test is the equivalent of the test outcome compared to the actual truthfulness. However, in some cases, the measurement of the accuracy of specific events or information investigated can be difficult, as there might not be available information regarding what occurred. Although, when trying to evaluate the Concealed Information Test (CIT) the measurement should be easy to define, as for the nature of the test itself, the examiner is usually aware of the real relevant information investigated.

2.2.5.1 The concept of validity within the polygraph

The polygraph can be considered as a diagnostic test; therefore, it results in binary judgment (yes or no). Typically, a polygraph test provides two categories of decision: significant response and no significant response. The only intermediate category can be addressed as *inconclusive response* (Council, 2003b). There are two aspects to accuracy: *sensitivity*, which, specifically in this case, refers to how well the test detects deceptive examinees (Swets, 1996); a perfect sensitive indicator of deception is one that present positive result whenever deception occurs. The more significant the proportion of deceptive examinees detected in the test, the more sensitive the test is to deception (Council, 2003b). The other important aspect is *specificity*, which, in this case, refers to how well the test identifies truth tellers (Swets, 1996). An indicator, which is perfectly specific to deception, always shows negative when deception is absent; it is positive only when deception is present. The greater the proportion of truthful examinees who appear truthful on the test, the more specific the test is (Council, 2003b).

In line with these measures, it is essential to specify that two closely related measures of test performance play a crucial role in polygraph screening decisions: the *False Positive Index* (FPI), defined as the ratio of false positives to true positives. This indicates how many truthful examinees will be wrongly detected. Finally, the *positive predictive value* (PPV) provides the probability that an examinee with a deceptive polygraph result is being considered deceptive. These two are inversely related: the lower the PPV, the higher the FPI (Council, 2003b).

The term *validity*, in polygraph testing, is usually referred to the psychophysiological test. Typically, the polygraph is considered valid only if deception is strongly associated with a specific pattern in the record of physiological responses made from polygraph (Council, 2003b). Therefore, in order to be considered as a valid indicator of deception, the test must perform better against an appropriate criterion of truth than indicators that have no validity. It is, thus, important to define the points of comparison: one could involve the level of performance that would be obtained by random guessing about the examinee's truthfulness or deceptiveness on relevant questions. In this analysis, the predictive validity of the test consists in the difference between its predictive value and the random guessing. However, this reference point produces a minimal comparison, which is too weak for practical uses (Council, 2003b).

Another reference point can be related to the *incremental validity* which is used for evaluating whether a new test will increase the predictive ability provided by an existing method of assessment (Swets, 1996). In polygraph testing, it refers to the degree to which deception is accurately detected by other techniques usually used in investigations, such as backgrounds checks or questionnaires. Concerning the validity, this method reports being more effective (Council, 2003b).

A third reference point for the evaluation of the validity of the polygraph consists in a comparison of an experimental group, where the examinees are purposely lying regarding the relevant questions, and a control group, where the examinees have not being involved in the experimental stage. In this way, this comparison might help to determine the extent to which the effectiveness of the test can be associated with its validity (Council, 2003b).

Also, when considering studies on the validity of the polygraph, it is essential to examine the *internal validity*. It refers to whether a study or experiment have adequately considered any extraneous variables that might affect the accuracy of the test (Bryant et al., 2000). Typically, this can be evaluated through the insertion of a control group that will be, then, compared to the experimental group. Finally, the *external validity*, which refers to whether the results from a study or experiment can be generalised to other contexts (Bryant et al., 2000).

2.2.6 Studies about the accuracy of the Polygraph

As discussed, evaluating the accuracy or the criterion validity of the polygraph is challenging. This is because the research literature, especially regarding the CQT, reflects two distinct conclusions that are still debated: one that entirely supports the validity of the polygraph testing, and the other one that refuses its application, because of its lack of validity. However, one of the reasons why these two conclusions are so different can be related to the implementation of two different methodologies used for investigating the validity of the polygraph

2.2.6.1 Laboratory Studies

Laboratory studies involve the simulation of real methods of polygraph testing in a controlled situation. The polygraph is usually used to investigate "mock" crimes created by the researcher or the actual simulation of small crimes (Patrick & Iacono, 1989). The main problem with laboratory studies is that participants do not feel threatened by punishment or by any legal consequence. Usually, the only motivation to lie is enhanced by monetary rewards. The sample usually does not involve real criminals, but students or members of staff aware of their participation in a research. In fact, Elaad and Ben-Shakhar (1989) suggested that the motivation to lie could be crucial. In their study, examinees were motivated by a financial stimulus to pass the examination. Results showed that motivated participants were more likely to be detected compared to those that were not motivated to pass the polygraph test (Elaad & Ben-Shakhar, 1989). Although, the strong element of this approach is that the criterion validity is guaranteed and internal validity is high (Forman & McCauley, 1986). The main criticism consists in the lack of ecological validity: in a laboratory setting, deceptive examinees are aware that the consequences of lying are not severe so, despite the result of the polygraph test, they will be free to leave the mock crime setting without further involvement. Conversely, in a real-life scenario, deceptive examinees are likely to be concerned about the polygraph results. Therefore, the emotional involvement in the two settings is extremely different (Fiedler, Schmid, & Stahl, 2002).

In a study conducted by Elaad, Ginton and Jungman (1992), Respiration Line Length (RLL) and Electrodermal response (EDA) were measured within a real-life setting. It was found that levels of EDA were higher in the real-life context, compared to the reported laboratory results; this can be explained due to the fear of conviction; guilty examinees' arousal levels were higher.

However, some researchers suggest that the difference between real-life settings and laboratory is not as crucial as it is claimed (Kircher & Raskin, 1988; Pollina, Dollins, Senter, Krapohl, & Ryan, 2004). According to these studies, if laboratory experiments provide a representative population, realistic testing procedures and participants are appropriately motivated to deceive a blind examiner, then it is possible to obtain useful information regarding the estimation of field accuracy. Further studies have shown that the accuracy of the polygraph in laboratory settings is high, with a sensitivity level that ranged between 88% and 93% (Forman & McCauley, 1986; Patrick & Iacono, 1989).

2.2.6.2 Field Studies

Field studies are considered as the "real-life" polygraph tests conducted on real criminal suspects. One of the main limitations of this approach in polygraph testing consists of identifying who is lying, to be able to compare the polygraph results. Similarly, in a real context, it is difficult to control all the factors and variables that might influence the test. For example, participants cannot be assigned to conditions of deceptiveness or truthfulness. In this setting, the accuracy of the polygraph can be determined just by comparing polygraph results with judicial decisions or by a legal panel of experts that can assess the confession of the suspect (OTA, 1983). Field studies are considered highly problematic because the methodologies applied for determining the truth have different limitations. Firstly, individuals can be wrongly sentenced or imprisoned for various charges. Secondly, the panel decisions are not always accurate, unable to establish a criterion validity (Dohm & Iacono, 1993). Moreover, the use of elicited confession can be also difficult, as deceptive examinees that pass the polygraph (false negatives) are excluded from the research. In the same way, even innocent individuals' wrongly detected (false positives) are excluded from the study, because of the lack of confession (Council, 2003b). Therefore, it is not recommended to rely on examinees confession as a criterion for validity.

To overcome these limitations, a critical study was conducted by Patrick and Iacono (1991a) with the Royal Mounted Police in Canada. In total, 402 polygraph cases were evaluated: the researchers discarded those cases where confessions or panel decisions were the only factors for determining accuracy. Instead, they searched for evidence of their guilt or innocence after they had completed the polygraph test with the support of additional evidence. This analysis found that in only one case it was possible to verify the guilt of a suspect independently. The results of this experiment report how inaccurate the polygraph was during investigations. In fact, in some cases, suspects were asked to undertake the polygraph test even if there were not enough evidence for a conviction.

Another issue related to field cases involves the internal validity and how factors that affect the polygraph results cannot be controlled. In fact, in many cases, the examiner's decision can be influenced by other information related to the suspect; this might include victim statements, medical evidence or records from previous convictions (Saxe, Dougherty, & Cross, 1985).

However, there are some clear advantages of field studies. The sample, which should represent the population and the strategy applied for creating it. In studies with the polygraph, the cases chosen for the study should be selected randomly, without having prior information regarding the physiological data. Another advantage consists in the polygraph evaluations; the physiological data should entirely determine the outcome of the polygraph. Finally, with field cases, the truthfulness of examinees must be validated by a criterion independent from the outcome of the polygraph test (Honts & Quick, 1995).

In according to the National Research of Council (2003b), only seven polygraph field studies (with overall datasets including between 25 and 122 polygraph tests) passed the minimal criteria for being reviewed; this included six studies on the CQT and just one on the CIT. The inclusion criteria required factors such as questions, physiological measures, instrumentation and scoring method used.

Similarly, the Organisation of Technology Assessment(1983) reviewed just ten field studies which included different cases of criminal investigations. However, none of the cases studied met all the aforementioned criteria.

2.2.7 Factors affecting Polygraph's Validity

In polygraph testing, the validity is a complex and critical construct. This characteristic makes the accuracy and its evaluation difficult and a number of different factors can influence it.

2.2.7.1 Physiology

As the polygraph results are based on physiological measures, it is very vital to understand how the individual differences can have an impact on final results. There are differences in response systems, such as electrodermal, cardiovascular, endocrine and central nervous system activity (Council, 2003b). One of the earliest reported individual difference in psychophysiological measures is related to electrodermal lability, considered as the frequency of responses that are obtained in the absence of any external stimulation (Crider & Lunn, 1971). In one of the few studies which tried to evaluate whether this physiological variable can affect the validity of polygraph results, it was found that stabile examinees with spontaneous electrodermal responses were less likely detected in a concealed information task, compared to labile examinees. On the other hand, innocent labile examinees were more likely to be wrongly detected as deceptive (Waid & Orne, 1980). Although these results are interesting, no robust study that investigated in depth the difference in electrodermal responses, as the other physiological measures. Therefore, it is not clear how much the validity of the polygraph can be affected by individual physiological differences.

2.2.7.2 Personality

Different studies have tried to evaluate whether the validity of the polygraph can be affected by the personality traits and individual characteristics of examinees. Some personality traits like anxious or conscious have been investigated. Usually, anxious subjects tended to respond strongly to stimulation, compared to non-anxious participants on the CIT. According to theory, their anxiety affected mainly their skin conductance responsivity (Giesen & Rollison, 1980). Similarly, it was found that people having a high level of Machiavellianism traits and self-monitoring show greater physiological responsivity (Bradley, 1987; Bradley, 1992). However, other studies failed in finding any particular relation between personality traits and variation in polygraph validity.

Gudjonsson (1982) found no relevant correlations between personality traits assessed by personality inventories and detection of deception with CIT. Research on abnormal personality traits such as personality disorders and psychopathy demonstrated that there is no concrete relation between psychopathy and accuracy of the polygraph (Honts, Raskin, & Kircher, 1985; Raskin, Barland, & Podlesny, 1976). Furthermore, in a study conducted on psychopaths and non-psychopathic prisoners, it was reported that there was no difference in responses (Raskin & Hare, 1978). Similarly, Patrick and Iacono (1989) found that socialisation or psychopathy do not affect the validity of the polygraph. Although the limited literature presented here has found no relevant impact of personality on polygraph validity, it is currently put forward that further studies are needed to test this conclusively.

2.2.7.3 Demographic Variables

There is a lack of research that has been conducted to evaluate the relation between polygraph testing and demographic variables (such as ethnic background, education or intelligence). It might appear obvious that intelligence might affect polygraph testing, as more intelligent examinees tend to educate and train themselves about how to use countermeasures. However, it has been shown that intelligence and education level does not affect the validity of the polygraph (Barland & Raskin, 1975a).

Similar results were provided by research regarding the impact of race and gender. Just a few studies have tried to evaluate race factors (Buckley & Senese, 1991; Reed, 1993) and all of them reported no significant effects of examiner's race or examinee's race. However, this was possible because of the technique of blind re-scoring of polygraph charts, where it is unlikely that the race might influence the examiner. Further support is found in a study be Reed (1993) who reported that in a sample of 375 polygraph tests there were no statistical differences in accuracy between Caucasian and African American examinees.

Non-significant results have also been reported for gender also. Two studies (Bradley & Cullen, 1993; Matte & Reuss, 1992) have investigated gender differences for specific physiological responses, during polygraph tests. These studies reported that there was no statistical difference in detecting deception between males and females. Furthermore, it was found that age could be differentiated through false-positives and real deceptive (Patrick & Iacono, 1989). Young participants were more likely to be classified as false positives. However, these demographic factors should be analysed more in-depth.

2.2.7.4 Countermeasures

One of the most relevant problems with the use of the polygraph is the application of certain countermeasures used in order to decrease the test's accuracy, producing a truthful outcome (Council, 2003b). The reason why guilty examinees apply such countermeasures is related to the attempt to influence the examinations to reduce the likelihood that their deception will be detected. In addition, these countermeasures can be also used by the truthful examinee, who try to reduce the possibility of false-positive results (Council, 2003b). However, studies have shown that some countermeasures used by innocent examinees could, instead, increase their chances of being detected as deceptive (Honts, Amato, & Gordon, 2001) Countermeasures are considered a distinct threat to the validity and utility of the polygraph (OTA, 1983). There is a wide range of potential countermeasures, such as drugs and alcohol use in addition to some mental and physical strategies that can be employed. For example, Mascke and Scalabrini (2000) provided suggestions on how to mimic the physiological responses of a deceptive examinee. Generally, the theoretical base underpinning the strategies consists of in increasing responses to the comparison questions (when referring to CQT); this can occur by inducing pain while responding to these questions, for instance, by pressing on a tact placed in the shoe or biting one's tongue.

Mental countermeasures consist of a range of techniques for increasing physiological responses to comparison questions. For example, it was suggested by Maschke and Scalabrini, (2000) that the use of arousing mental imagery (e.g. thinking of being hit by a car) or attention focusing strategies (counting backwards) during comparison questions, could create stronger physiological responses. Although many researchers have investigated the impact of these strategies on the overall effects of countermeasures, they have not analysed the specificity of these measures. For example, Honts, Raskin and Kircker (1994) conducted such a study where 80 subjects were trained in using mental and physical countermeasures. The results of this study reported that almost 50% of the sample was able to pass the CQT by using these physical techniques. Therefore, mental and physical countermeasures were "equally" effective; in fact, it was impossible to detect the single countermeasure through observation. Same findings were obtained by other studies (Honts et al., 1985; Honts, Raskin, & Kircher, 1987; Honts, Devitt, Winbush, & Kircher, 1996), which equally supported the efficiency of the application of countermeasures in the attempt to decrease the likelihood of detect deceptive subjects in specific polygraph tests.

However, there are many limitations related to the laboratory settings and mock crimes scenarios associated with the use of countermeasures. The majority of the studies have been conducted with volunteers in controlled settings; therefore, it is reasonable to question the generalizability to real-life contexts. In general, the examiner is experienced, and trained and examinees are not always well prepared. The psychological factors, such as fear and anxiety, play an essential role in the performance during the test. Finally, it has not been investigated whether specific countermeasures have homogenous effects on all of the chart readings obtained during a polygraph test (Council, 2003b).

Many studies have produced results that do not support the influence of mental and physical results. Physical measures refer to physical activities used by examinees to avoid being detected, such as self-inflicting pain with the tongue or by controlling the breathing in according to the type of question asked. Mental measures consist of psychological activity performed by examinees to create misleading physiological responses (Ben-Shakhar, 2011). It was stated that not all mental countermeasures are effective; instead, some of them are able to reduce the accuracy of certain polygraph tests, such as the I/R test and the CQT (Kubis, 1962). Diversely, it was found that the same countermeasures applied in other tests were not effective with the CIT (Ben-Shakhar, 2011).

2.2.7.5 Drugs and Alcohol

Conflicting research can be found in the literature in regards to the effect of drug and alcohol during polygraph testing. For example, it was found that anxiolytics had a detrimental impact on a CIT (Waid et al., 1981). A similar experiment was replicated, where the effects of meprobamate, diazepam and propranolol were compared on detection of guilt with CIT (Iacono et al., 1984). In contrast, their results found that none of the drug conditions analysed had a significant effect for detecting deception. This last conclusion was also supported by Iacono, Cerri, Patrick and Flemming (1992), who obtained similar results, stating that these substances did not affect their sample.

Regarding the effect of alcohol on polygraph testing, just a few studies tried to investigate this issue. In one of the studies (Bradley & Ainsworth, 1984) it was found that being intoxicated at the time of the test had no relevant effect on the validity of the test. Instead, being intoxicated reduced the accuracy of deceptive examinees. Similarly, O'Toole, Yuille, Patrick and Iacono (1994) found that being intoxicated had no particular effect on the validity.

Overall, the available evidence reported above suggested that drugs and drinking have only a limited impact on polygraph testing. However, just a few studies tried to evaluate the effects of drug and alcohol on the accuracy of the polygraph. Therefore, the evidence obtained are slightly limited by the laboratory setting and by the fact that all subjects were volunteers. Clearly, this aspect should be investigated further for getting more relevant results.

2.2.7.6 Expectation Effects

Few social psychology theories can explain some of the examinees' behaviour while they are being tested during a polygraph procedure. It was suggested that in some cases the examiners' expectations of a guilt examinee might have an influence not only on examiner's judgments when analysing the charts but also on examinees' physiological responses during the examination (Council, 2003b). In social psychology, this is defined as Pygmalion Effect (Terence & Daniels, 2003), which occurs when higher expectations lead to an increase in performance. In regards to this, two forms of biases have been distinguished: the *cognitive confirmation effect* and *the behavioural confirmation effect* (Darley & Gross, 1983).

The cognitive confirmation effect describes the tendency to influence the interpretation of ambiguous information by the examiner's beliefs and expectations. Therefore, if an examiner thinks that the examinee is guilty, he might interpret information in a way that confirms this belief. In a study conducted by Barland and Raskin (1975b) this tendency was also confirmed; 17 polygraph examiners out of 10 produced evaluations on their examinee's truthfulness based on information they were provided before the test. After the administration of the test, these evaluations were maintained. Although there is a small body of research regarding this effect, it has been shown by other researchers (Elaad, Ginton, & Ben-Shakhar, 1994) that it is possible to bias the examiners on the outcome of a polygraph examination. The behavioural confirmation effect describes the influence of expectations on individuals' behaviour towards other (Darley & Gross, 1983). In a polygraph setting, an examiner might be motivated to manipulate the results in a way that they will match with his beliefs or feeling towards the examinee. This approach has been considered unethical and inappropriate (Honts & Perry, 1992).

2.2.8 Factors affecting the Polygraph's Utility

A small distinction exists between the concept of validity and utility of the polygraph test. Keeping in mind the concept of validity discussed in section 2.2.5 of the current chapter, the efficiency of the polygraph refers to the practical value of the test combined with a given degree of accuracy in specific decision-making context (Council, 2003b). As for the validity of the polygraph, different factors can affect the utility of the test.

2.2.8.1 Deterrent

One of the main effects associated with the use of the polygraph, especially for screening, is deterrence. The threat of polygraph testing can work as a deterrent: if an individual believes the polygraph can detect a deceptive examinee with high probability, and the consequences of being identified are not desirable, then the examinee might tend to avoid the criminal behaviour that is being monitored (McGuire, 1969). Deterrence is different from the concept of validity of the polygraph because the test can be an effective deterrent even if it does not provide valid information about deception. However, these two concepts are related, as the utility of the polygraph depends on the beliefs that individuals have about the validity and about how the results will be used (Council, 2003b). If examinees possess these beliefs, they will be deterred from engaging in behaviours they believe the polygraph can detect.

2.2.8.2 Admissions and Confessions

Another utilisation of the polygraph consists in facilitating interrogations (Davis, 1961). Supporters of the polygraph suggest that individuals are more likely to confess or providing disclosed information about criminal behaviour if they believe that the test will be able to detect them (Council, 2003b). A study on sex offenders (Sigurdsson & Gudjonsson, 1996) suggested that there are three main reasons why suspects confess their crimes. Firstly, suspects' perceptions of the strength of the evidence against them, secondly external pressure such as fear of custody and finally internal pressure, such as personal thoughts. All these factors can be present when a suspect decides to confess. However, the perception of the evidence against the suspect can represent one of the most influential factors.

There is no enough research supporting the ability of the polygraph to elicit confessions; however, many studies supported this theory with the use of the "bogus pipeline" technique (Jones and Sigall, 1971; Quigley-Fernandez and Tedeschi, 1978; Tourangeau, Smith, and Rasinski, 1997). This paradigm consists in attaching subjects to a series of fake sensors as well as to a device presented as a lie detector, in truth, the machine is non-functional. It was showed that with this setting, examinees are more likely to admit personal beliefs and facts than similar examinees not connected to the bogus lie detector (Council, 2003b). Different studies confirmed the effectiveness of this procedure, in a study conducted by Jones and Sigall (1971) students were asked to provide some personal information regarding their racial attitudes while attached to the bogus machine. Those students connected to the machine disclosed more politically incorrect responses than those from the control group, as they felt it was in their interest to tell the truth rather than be caught out lying. Similarly, even admissions occur with the same mechanism. Examinees tend to confess acts or behaviour that had not previously been disclosed in a bogus pipeline setting; this has been considered as a strong support for the utility and validity of the polygraph (Council, 2003b).

However, there are cases in which admissions and confessions obtained from interrogation through the polygraph are false. It was suggested that false confessions are even more common than truthful ones (Kassin, 1997). Although there is no robust literature about this issue to draw on, one study showed that 17% of examinees on a bogus pipeline setting showed strong physiological responses to a question regarding a minor crime which was admitted even if they never committed the crime before (Meyer & Youngjohn, 1991). Therefore, even if this technique can be useful for certain confessions, its application needs to be carefully treated.

In conclusion, admissions and confessions during polygraph testing can support the validity of the polygraph just in a few circumstances; in some cases, the admission can even negatively affect the assessment of the validity, as in real settings the admission occurs at the end of the evaluation of the administration of the polygraph (Council, 2003b).

The appropriate criteria for determining the validity of the polygraph test can be different and can vary depending on the purpose of the examinations and the contexts. In fact, this section showed how many factors can influence the polygraph accuracy during an examination and how these elements can affect the results. Although, it has been showed that there is a general lack of literature around the actual effects of such factors concerning the application for the polygraph. Therefore, it would be difficult to consider some of these factors as relevant and crucial for the present project. The overview of field and laboratory studies showed that the accuracy of the polygraph could be variable. As matter of fact, field studies cannot guarantee criterion validity and have many issues with internal validity. Diversely, the main problem with laboratory studies relates to the external validity, but a strong criterion validity can be investigated. This variation in the accuracy rates reported by all the studies discussed result in inconclusive definite opinions regarding the validity of the polygraph (Council, 2003b).

2.2.9 Polygraph Tests

The previous section provided a general overview of the polygraph validity, highlighting which factors represent the main limitations to its accuracy. However, it is necessary to provide a detailed analysis of the theoretical assumptions that underpin the different polygraph tests.

Different tests can be used within polygraph examinations. Many factors cause the differences between the polygraph tests, such as the levels of validity, the theoretical assumptions, the structure and type of questions used and finally the way in which the physiological reactions are triggered. The most popular tests used are as follows:

- Relevant/Irrelevant Technique
- Control Question Test
- Concealed Information Test

2.2.9.1 The Relevant/Irrelevant Technique

In 1932, Larson developed one of the first polygraph procedures: the Relevant/Irrelevant Technique (RIT). This polygraph test contains two types of questions:

- Crime-Relevant Questions: these questions relate to the crime under the investigation. For instance, "Did you kill Miss X last night in his apartment?"
- Irrelevant Questions: these questions are not associated with the crime; therefore, their contents are neutral. For instance, "Is it Friday today?"

The theoretical assumption of this test it is based on the idea that physiological responses from both types of questions need to be compared. If the examinee produces greater responses to relevant crime-related questions than to the irrelevant questions, it indicates that deception occurred while answering to the relevant questions (Bull et al, 2004). In line with this principle, innocent examinees should have an equal response to all questions presented; they should be emotionally disinterested to the crime-related questions (OTA, 1983). In those cases in which results appear to be similar from both questions, the conclusions suggest that the examinee was truthful (Podlesny & Raskin, 1977).

However, there are many issues related to the accuracy of this test. Because of the direct nature of the questions, the crime-related questions are more likely to provoke physiological reactions to examinees, even if they are non-deceptive individuals (Saxe, 1994). However, strong physiological responses can also occur if innocent examinees are afraid of not being believed or if the questions provoke surprise or misunderstanding (Bull et al, 2004).

In 2011 a committee of the American Polygraph Association (APA) conducted a literature review regarding the scientific evidence of various polygraph techniques. The committee found that one of the only important study to investigate the accuracy of the RIT to be conducted by Krapohl, Senter and Stern (2005). However, this represented the only available study that satisfied the minimal requirements for being recognised from the APA; therefore, research about the validity of the RIT results are not satisfactory.

Due to the lack of extensive literature regarding the RIT, the test was considered to be not valid (Association, 2011). As consequence of this neglect, Krapohl and Rosales (2014) decided to conduct a study to estimate the accuracy of the RIT. The results of this study showed that although the test was able to detect deception in the majority of cases, the level of false positive was significantly high. Therefore, it has been suggested that due to its apparent limitations in detecting truthfulness, the RIT cannot be used as a stand-alone method. Its application can be restricted to settings where there is a high tolerance for false positives errors. As matter of fact, the RIT is used only for security screening purposes as part of the employee selection process by American government agencies and law enforcement organisations (Krapohl et al., 2005).

Is has been suggested, thus, that the RIT should not be used for polygraph examinations, due to the lack of validity that the test presents, considering it as an improper technique (Honts, 1991; Lykken, 1998; Saxe, 1994).

2.2.9.2 The Control Question Test

The Control Question Test (CQT; also known as the Comparison Question Test), was developed by John Reid in the late 1940s. At the beginning, the CQT was developed as an improvement for the limitations related to the R/I technique; however, it rapidly became one of the most known and used polygraph tests (Iacono et al., 1992; Lawson, 1988). This test compares responses produced from three types of questions:

- *Relevant Questions*: these questions directly refer to the crime under investigation. For instance, “Did you kill Miss X during the night of the 2th of March?”
- *Control or Comparison Questions*: these questions indicate acts that are indirectly related to the crime under investigation, so they do not refer to any crime specifically. These questions are purposively generic and vague. For instance, “Have you ever committed an illegal act?” The role of the control question is to embarrass both guilty and innocent examinees in order to provoke physiological responses. These types of questions are constructed with the purpose to give no other choices but lie when answering. Therefore, the resulted denials from these questions are considered deceptive. In normal situations, some individuals might admit to have committed some illicit actions (expressed in control questions). However, in a polygraph examination the admitting of this would cause the examiner to realize that the examinee might be the type of person that would commit the crime in questions, considering him/she guilty (Bull et al, 2004).
- *Irrelevant Questions*: these questions are not connected to the crime under investigation or to general crimes. For instance, “Is your name Anna?”

The basic assumption of the CQT is that control questions should generate greater physiological reactions than relevant questions within innocent examinees; this occurs because innocent examinees will be more concern about their answers related to the control question because they know they are deceptive to these questions and truthful to the relevant questions (Bull et al, 2004).

An examinee cannot be found guilty for having committed a crime by giving deceptive answers to control questions, as they are not directly related to the offence under investigation. For this reason, the control questions are purposefully misleading (Lawson, 1988). In other words, the control questions intend to divert the *psychological set* of truthful examinees away from the relevant questions (Vrij et al., 1997). Alternatively, guilty suspects give deceptive answers to both types of questions, producing similar physiological responses. However, the most severe threat to examinees consists in the relevant questions, which should provoke a greater physiological reaction than control questions (Bull et al, 2004).

Different criticisms have been levelled against this test (Kashy & DePaulo, 1996); however, the core of the critique concerns the CIT's theoretical assumptions. Although this test is considered as an improvement of the RIT, it is not considered theoretically robust enough so that that relevant questions produce higher physiological reactions because of deception (Bond & DePaulo, 2006; Ekman & Friesen, 1969). In fact, there could be other reasons that can trig arousal, such as the fear of not being believed (Bull et al, 2004). In addition, the test is not considered standardised, as the control questions depend on the type of crime under investigation. Moreover, in order to provoke the appropriate physiological reaction, control questions as "have you ever committed an illegal act?" can only be asked to examinees that have already committed such type of crimes (Bull et al, 2004).

In addition, they purposely intents to mislead examinees, which is a crucial element in the process of creating arousal during the answering of control questions, has been considered unethical (DePaulo & Rosenthal, 1979). In truth, this procedure contrasts with the general guidelines applied on testing, which highlight the importance to provide precise information on examinees (Bull et al, 2004) Even Lykken (1998) criticised this aspect of the test, considering the built-in test bias against innocent suspects. The reason why this represents a massive failure in the test is that it is impossible to know to what extent the control questions can provide an adequate psychological counter to the emotional impact of a false accusation made by the relevant questions (Kraut & Poe, 1980). Consequences of a false allegation due to wrongful CQT results can have a consistent emotional impact on innocent examinees.

The other important debate regarding the use of the CQT is related to the incongruent results obtained from studies regarding the validity of the test. In fact, for guilty examinees, accuracy ranges from 75% (Ekman & O'Sullivan, 1991) to 87% (DePaulo & Pfeifer, 1986), with a false-negative rates that vary from 1% to 13% (Swets, 1996); while for innocent examinees, the accuracy of the CQT ranges from 59% (Ekman & O'Sullivan, 1991) to 83% (Bryant et al., 2000), with a false-positive rates that range from 10%to 23% (Swets, 1996). Due to the difference in the results obtained by different studies, the use of the CQT in court and other settings is still debated (Podlesny & Raskin, 1977).

However, the CQT remains the most used and preferred approach in the United States and many other countries. Although scientific evidence and arguments regarding the limitations of the CQT have been reported during the last decades, field examiners report that this technique is almost infallible (Vrij et al., 1997). An explanation for this strong position could be related to the feedbacks that field examiners receive regarding the accuracy of decisions they make in polygraph test cases, which appear to be selective and systematically biased (Vrij et al., 1997). In confirmation of this, Patrick and Iacono conducted an important field study, analysing all cases tested by the police polygraph unit in a major Canadian city over a five-year period. They found different systematic biases, such as receiving feedback regarding their decisions from post-test confessions, or that the outcome of the polygraph tests systematically influenced the nature of the feedback that examiners received. As a consequence, police polygraph examiners almost never received post-examination feedbacks that could disconfirm their test decisions (Vrij et al., 1997); this means that police polygraph examiners rarely had the opportunity to face the fact that their decisions were found wrong. Probably, the constant systematic biases and the resistance regarding the establishment of credible accuracy estimates in real-life cases represent the main reasons why the validity of the CQT is still debated (Patrick, 2011).

The promise of both R/I and CQT tests is that guilty examinees will be more aroused when answering to "relevant" questions than when answering to "control" questions, due to the fear of being detected (Bull et al, 2004). However, there are different issues with this theoretical assumption, as innocent examinees might also physiologically react to "relevant" questions for different reasons: in cases when these questions are emotion evoking, such as when the question contents are related to a person or to an event that is emotionally significant for the innocent examinee. Furthermore, the innocent examinee might experience fear during the examination because of the concern of not being believed. In order to resolve this limitation, the Concealed Information Test was developed, placing its premise on a different assumption: examinees will be physiologically responsive once they recognise crucial details of a crime (Bull et al, 2004).

2.2.10 The Concealed Information Test (CIT)

In 1959, David Lykken developed the Concealed Information Test (CIT; also known as Guilty Knowledge Test). The aim of the test is to investigate whether an examinee possesses knowledge about a particular crime that they deny knowing (Krapohl, McCloughan & Senter, 2006).

2.2.10.1 Description of the CIT

The CIT tries to determine whether a suspect possesses knowledge about a crime that could exist only if the examinee is guilty. The CIT consists of a series of multiple-choice questions, developed by using information from the offence under investigation that only the perpetrator and the police could be expected to know. Each alternative needs to appear equally plausible to the innocent examinee, but one of which would be clearly identifiable to the guilty, e.g. investigating the type of weapon used during an assault (Patrick, 2011). The relevant information would be the knife, while the others will be plausible weapons that can be used during an assault. This test contains three types of questions:

- *Irrelevant Questions:* these questions do not include any information regarding the crime under investigation. For instance, "Are you a student from University?"
- *Relevant Questions:* these questions contain the relevant information that is used in order to trig physiological responses into the guilty examinee. For instance, "Did you kill Miss X with a Knife?"
- *Control Questions:* these questions do not contain the relevant information; however, they should have the same nature of the relevant one. For instance, "Did you kill Miss X with a gun?"

Breathing rate, blood pressure and electrodermal activity (EDA) are measured while the questions are asked. The guilty examinee should produce his largest physiological response with the relevant question, being able to recognise the familiar detail among other irrelevant options. Conversely, the innocent examinee should react in the same way to all the alternatives (Krapohl, McCloughan & Senter, 2006).

Due to different structures and mechanisms upon which the test is constructed, it was stated that the CIT does not measure deception or guiltiness. In fact, initially this test was referred as the guilty knowledge test. However, this description is not entirely appropriate: the test is not able to prove whether a suspect has guilty knowledge (Patrick, 2011). On the contrary, the test operates as a *recognition test*. In other words, it is considered more like a memory detection test, able to identify whether an individual recognises information familiar to him/her (Patrick, 2011). (More information regarding the memory processes involved in this mechanism were discussed in section 2.4.4)

2.2.10.2 Theoretical assumption of the CIT

In order to understand the theoretical assumptions of this test, it is necessary to analyse the CIT's underlying mechanisms. Different theories and approaches are involved in the psychophysiological mechanisms of the test, which can be classified in two different categories. The *Emotional-Motivational Approaches*, which seem to be unnecessary for the detection of concealed information, but could enable discriminability; and *Cognitive Approaches*, which provide a more in-depth explanation of the relationship between cognitive and physiological mechanisms during a polygraph examination (Verschuere et al., 2011).

2.2.10.2.1 Emotional-Motivational Approaches

One of the theories that tries to explain the physiological reactions to the relevant stimulus is the *Conditioned Response Theory*, based on principles of learning psychology (Pavlov, 1927). In line with the classical conditioning theory, it is stated that an intrinsically neutral stimulus can impact behaviour through association with a stimulus that evokes that behaviour (Davis, 1961). In the context of a polygraph examination, it is assumed that the criminal act and the crime scene function as unconditioned stimuli, while the crime-related items are the conditioned stimuli. In this way, the relevant questions, which contain crime-related information, can provoke physiological reactions in examinees (Verschuere et al., 2011). For example, in the context of police interrogations, this approach can be applied when trying to investigate an experience with strong emotions (e.g. showing pictures of a victim to a murder suspect). However, although the theory provides an interesting interpretation of the CIT mechanism, there is no empirical research that has directly tested this theory (Verschuere & Ben-Shakhar, 2011). The theory presents different limitations: it cannot be applied to all types of concealed information testing due to the lack of information that can work to create neutral conditions (Verschuere et al., 2011). In addition, a laboratory study (Ben-Shakhar & Elaad, 2003) has shown that when participants were asked to hide recognition of certain code words or a chosen playing card, strong physiological responses where produced. Therefore, such trivial stimuli do not seem to evoke strong emotions like the theory claims, but they are able to elicit strong physiological responses. For this reason, other theories have been considered more applicable for explaining the mechanisms of the CIT. However, more research is necessary in order to clarify whether a conditioned mechanism is involved in some of the CIT examinations.

The *Punishment Theory* (Davis, 1981) provides another interesting interpretation. According to Davis, the enhanced responses evoked by the relevant items are the physiological consequences of the fear of failure when deceiving. Another theory needs to be mentioned as it is strongly related to this approach, and that is the *Defensive Reflex Theory* (DR; Sokolov, 1963). According to this theory, the defensive reflex is elicited by unpleasant stimuli. In fact, this theory provides protection to the organism from aversive stimuli. In CIT settings, the questions asked elicit defensive responding, because of their threat value. Therefore, for the guilty examinees, relevant items should be more threatening, provoking stronger defensive responding; while for innocent examinees, all the items have the same threat value, thus all the questions should elicit similar responses (Verschuere et al., 2011). Physiologically, fearful responding is characterized by respiratory hyperventilation (Van Diest et al., 2001), heart rate acceleration (Graham & Clifton, 1966), and startle potentiation (Lang, Bradley, & Cuthbert, 1990). However, studies have shown that in CIT settings, hyperventilation does not usually occur; instead, guilty examinees typically respond with greater respiratory suppression to the relevant items compared to the control items (e.g. Verschuere, Crombez, De Clercq, & Koster, 2004; Gamer, Rill, Vossel, & Godert, 2006). Similarly, other studies have shown that in CIT settings, the heart rate in guilty examinees shows an initial acceleration followed by a secondary deceleration (Gamer et al., 2006; Verschuere, Crombez, De Clercq, & Koster, 2005). Although this initial acceleration seems to support the punishment/DR theory, it has been shown that this mechanism is also found when examinees are exposed to control items. However, the second deceleration is found only when examinees are exposed to the relevant items (Verschuere & Ben-Shakhar, 2011). Therefore, the physiological response pattern evoked by relevant items does not support the fear system described by the Punishment/DR theory. In support of this, the mechanism fails to explain the high detection rates that occur in laboratory settings, where there are no severe consequences for the examinees (Ben-Shakhar & Elaad, 2003). Although it is necessary to consider the impact of a real-life situation and how the circumstances can affect the individuals' state of mind, this theoretical framework cannot provide a stand-alone explanation of the psychological processes involved in the CIT mechanism.

Another theory that focuses on emotional factors is the *Emotional Conflict Theory* (Davis, 1981). According to this approach, two opposite reaction tendencies occurring at the same time produce a larger physiological reaction than the reaction to either alone. The natural tendency of answering questions directly creates one reaction tendency; instead, the situation in which an examinee is motivated to deny the truth would create an incompatible reaction tendency. In other words, enhanced physiological reactions evoked by the relevant questions reflect the internal emotional conflict, which is the result of the urge to tell the truth and the need to deceive to avoid detection (Verschuere et al., 2011). For this reason, a stronger physiological reaction would be aroused when answering "no" to the relevant question. However, studies have shown contrasting results; in fact, some of them found clear differences between the responses to relevant and control items (e.g. Elaad & Ben-Shakhar, 1989; Furedy & Ben-Shakhar, 1991; Horneman & O'Gorman, 1985; Gustafson & Orne, 1965a); others found no relevant results (Verschuere et al., 2009). Although there is conflicting evidence for this theoretical approach, other authors have suggested a different interpretation of this mechanism (Spence et al., 2004; Vrij, Fisher, Mann, & Leal, 2006). They suggest that this conflict might not be purely "emotional" but mainly "cognitive". In fact, it is possible that the increased response to the relevant item results from a denial that can be considered cognitively more demanding than a silent or a truthful answer (Vrij et al., 2006). Conversely to this theoretical approach, it has been suggested that clear denial might play an important role, but is not necessary for successful concealed information detection (Verschuere & Ben-Shakhar, 2011). In fact, a study demonstrated that the clear physiological responses could also be evident in cases where examinees responded "yes" to the questions or remained silent (Vrij et al., 2006). However, it is not clear if this is caused by an emotional conflict or perhaps as the result of a cognitive process involving recognition/familiarisation of information. In any case, a purely "emotional" theoretical framework is not able to provide conclusive evidence for explaining this mechanism.

The motivation to lie is a factor that plays a crucial role during a polygraph examination. For this reason, this emotional aspect has been widely investigated, with the result that a *motivation impairment effect* has been found in a few studies (Burgoon & Floyd, 2000; DePaulo, Kirkendol, Tang, & O'Brien, 1988). This effect indicates that the more motivated the liar, the more likely the lie is to be expressed in non-verbal behaviour. This approach is strongly related to the punishment/DR theory because the consequences of the polygraph test could increase the motivation. Gustafson and Orne (1963) conducted a series of experiments in which they manipulated the participants' motivation to examine the relationship between motivation and the degree of psychophysiological differentiation in the CIT. Results showed that motivational instructions enhance CIT detection performance. However, the meta-analysis conducted by Ben-Shakhar and Elaad (2003) demonstrated that successful detection rates were also found under low motivation conditions. Therefore, as for obvious deception, motivation can help to increase the CIT detection efficiency; however, it does not represent a necessary condition for detection (Verschuere et al., 2011). This factor might represent a big limitation for the laboratory studies, with the risk of decreasing the validity of the results. Although it has been shown that motivation is not a factor that needs to be necessarily manipulated in order to obtain effective detection (Verschuere et al., 2011); however, it is important to keep participants highly motivated, so that minimal levels of physiological reactions can be produced during examinations.

In conclusion, these findings indicate that emotional-motivational mechanisms are crucial in the process of lying during a CIT; however, their theoretical limitations suggest that other processes might also be involved in the description of the CIT dynamics. The present thesis will take into consideration some of these theoretical assumptions, which will help to strengthen the theoretical framework of the current project. The above information relates directly to aims 1 and 2 (For more information about the aims of the project, see section 2.4).

2.2.10.2.2 Cognitive Approaches

The most influential theory that tries to explain the underlying mechanisms of the CIT is the *Orienting Response Theory*. The idea of the orienting reflex was first introduced by Pavlov (1927) and then developed by Sokolov (1963), describing it as a complex reaction provoked by a novel stimulus or a change in the environment. Therefore, when repeated presentations of the same stimulus occur, the response magnitude will gradually decline, producing an effect called "habituation". Sokolov assumed that repeated presentations of a given stimulus produce an internal representation that contains all the parameters of that stimulus, a defined "neuronal model". All the input information received from the environment is compared with the existing neuronal models; a mismatch between stimulus input and the internal models results in an Orienting Response (OR). Conversely, if the input matches with the existing model, the OR will be inhibited and habituation will occur.

In respect to this, it has been found that when elicited by the relevant items, it is possible to observe a specific physiological response pattern associated with the OR. Lynn (1966) has described these typical physiological features as decreased heart rate, increased sensitivity of the sense organs, increased skin conductance, general muscle tonus, pupil dilation, vasoconstriction in the limbs and low-voltage electrical activity in the brain. In a CIT setting, Lykken (1974) applied a connection between the OR and the differential responsivity to the relevant questions of the test. Consequently, for a guilty participant, the relevant item presented will have an exceptional significance, which should produce a stronger orienting response compared to the non-relevant or control questions. Conversely, for an innocent participant, the relevant item should not have any particular significance, thus all items presented should evoke the same ORs. According to this approach (Verschuere et al., 2011), it might not even be necessary for an examinee's response, as the OR would occur regardless of whether the response were negative or positive. This assumption helps to explain some of the limiting aspects of the *Emotional Conflict Theory*, confirming that an overt denial might not be necessary for an effective physiological change. Furthermore, the OR serves to allow more elaborate processing of the OR-eliciting stimulus (Kahneman, 1973; Ohman, 1992). In fact, research demonstrated that there is a positive correlation between OR and later recall of the stimuli (Corteen, 1969). In support of this view, a study conducted by Waid et al. (1978) showed that detection was positively correlated with the number of words that were recalled after a CIT examination. In addition, recalled items were more likely to provoke an electrodermal response than non-recalled items. In conclusion, all these studies demonstrated that detection efficiency has a positive association with the recall process, which means that examinees tend to remember more details from the information contained in the relevant questions (Verschuere et al., 2007). In contrast to the emotional-motivational approaches, Lykken (1974) stated that emotional factors are irrelevant; despite the examinee's psychological-emotional state concerning the test, it is still possible to expect a greater physiological reaction when recognition occurs. For this reason, the lack of threat or psychological pressure typical of a real-life situation is not considered a limiting factor for laboratory settings. In fact, the present study will demonstrate that although the studies will be developed in a laboratory setting, the cognitive mechanisms involved in the recollection and recognition of information will be activated and will produce relevant results.

Another theory that is strongly related to the OR mechanism is the *Dichotomization Theory*, identified by Lieblich et al. (1970), and later extended by Ben-Shakhar (1980). The theory postulates that stimuli are represented in two distinct categories: relevant and irrelevant. It is assumed that during a CIT examinee pay attention to whether these stimuli belong to one category or the other one; the other aspects of the stimuli, such as distinguishing features between the two stimuli, are ignored. According to Sokolov's theoretical perspective, for each stimulus category a single neuronal model is formed. Therefore, habituation generalises within each category. For an innocent individual, all stimuli belong to the irrelevant category, and responses will quickly habituate (Sokolov, 1963). This theory is seen as an additive component rather than in competition with the previous approaches (Verschuere & Ben-Shakhar, 2011). Therefore, it mostly emphasises the relevance factor, based on the signal value of the stimulus. However, this theory does not take into consideration the effect of emotional-motivational factors on differential responding to significant stimuli, which could be relevant when trying to provide an accurate explanation of the mechanisms involved.

The last approach to note is the *Feature-Matching Theory* developed by Gati and Ben-Shakhar (1990), which tried to supplement the OR theory and expand part of Sokolov's theory (1963). This theory postulates that the competing process of stimulus inputs with stimulus representation is comparable to the process involved in the comparison of two stimuli to make similar judgments. It was assumed that sets of features could characterize both stimulus input and neuronal model (or stimulus representation). As previously mentioned, two factors can produce the OR elicitation: novelty and significance of the stimulus. Each factor is analysed with a different feature-matching process. The degree of mismatch between features of the coming input and the activated neuronal models determines novelty; contrastingly, the degree of match between features of the input and the representations of past significant experiences determines significance. The outcomes of the two matching processes are integrated to produce an OR. Contrary to the dichotomisation theory, which considers the stimulus significance as a dichotomy, this approach sees the significance as a continuum. In addition, this theory affirms that the OR generalises to stimuli that share common characteristics with the relevant stimulus (Ben-Shakhar & Dolev, 1996). In fact, in the stage of developing irrelevant items, it is important that they are as distinct as possible from the relevant item; this aspect gives robustness to the test against changes in the modality of presentation of the items e.g. verbal versus visual presentation of the relevant items; these elements have been considered essential when building up a CIT (Verschuere et al., 2011). This represents a crucial element for the construction of the CIT items, which will be taken into consideration during the development of the CIT of the present study.

There are many alternative explanations for the validity of the CIT. However, such explanations beyond the orienting response mostly played presented weaker arguments or theoretical frameworks that have yet to be fully empirically supported. The strongest support for this theory has been provided by the electrodermal response mechanism, which matches the typical properties theorised for an orienting response. In fact, the response magnitude of the orienting reflex can be regarded as reflected in the electrodermal response amplitude (Rosenfield, 2018). Although electrodermal responses are considered as the main indicators for the occurrence of the orienting response, there are other physiological responses in the CIT related to certain mental processes associated with deceiving or concealing information (see Section 2.3), which will be assessed in this thesis. The above information relates directly to aims 1 and 2 (For more information about the aims of the project, see section 2.4).

2.2.10.2.3 Memory Detection

As we will see in the next section (section 2.3), memory plays an important role in the process of lying. Specifically, the process of encoding, storing and recollecting information in the long-term memory can be crucial for the effectiveness of certain polygraph tests, such as the CIT. In fact, this test aims to detect the presence or absence of relevant information in the examinee's memory, involving cognitive processes such as recognition and familiarisation (see section 2.3). For the nature of the basic assumption on which the test is based, the CIT is considered as a memory detection test, which can be used to reveal whether or not an examinee possesses certain information. However, the difficult aspect of this mechanism is that there are various types of criminals and different type of crimes committed in different circumstances. Consequently, when using the CIT, police investigators need to analyse the crime from different angles. With this principle, the test can be used not only to detect offenders or innocent suspects, but also potential suspects. Although the CIT could clarify if a suspect possesses certain relevant information regarding past actions, it has been hypothesised whether the test is also able to detect future actions. According to a study conducted by Meijer, Verschuere and Merckleback (2010a), the CIT is able to detect not only memories or past actions but also intentions of committing an action. Theoretically, this can be possible because the memory mechanisms involved in the process of recollecting real actions is the same as the one involved when trying to recollect an action that an individual is willing to make. When comparing both scenarios, the authors suggested that an event that has been committed could be remembered better than simple intentions, because intentions might be based on scripts. However, such scripts should be remembered anyway, as a result of familiarisation of information. Due to the lack of any real occurrence of the action, the significance of memory intentions can be low. Although the study conducted by Meiyer et al. (2010a) demonstrated that the CIT was effective in both conditions: committing a mock crime and planning a mock crime. This new application was also supported by Maixner and Rosenfeld (2011), who showed how the CIT was able to detect criminal intentions from a group of suspects. Although these results are promising, no further studies have been conducted around the use of the CIT for predicting criminal intentions, leaving a big gap in the literature. Many limitations have been raised from these studies, such as the limited available information about an action that has not happened yet or the level of information that a suspect could possess, due to his/her role in a criminal plan. For this reason, the present study will provide additional contributions around the use of the CIT as a tool for preventing criminal acts. The above information relates directly to aims 1,5 and 8 of this thesis (For more information about the aims of the project, see section 2.4).

2.2.10.3 Validity of the CIT

The CIT is considered one of the most valid methods for differentiating between guilty and innocent examinees (Elaad, 2011; Holden, 2000; Iacono & Patrick, 1997). Several reviews have been conducted to define the accuracy of the CIT. Ben-Shakhar and Elaad (2003) conducted a meta-analysis of CIT laboratory experiments and demonstrated that the CIT is highly accurate. Iacono and Lykken (1997) conducted two surveys about the validity of the methods of detection. They found that 77% of the members of The Society for Psychophysiological Research and 72% of the members of the American Psychological Association considered the CIT to be based on sound psychophysiological theories.

Ben-Shakhar and Furedy (1990) conducted a review of 10 CIT laboratory studies, showing that 83.9% of 248 guilty examinees and 94.2% of 208 innocent examinees were correctly identified. Similarly, Elaad (1998) reviewed 15 mock crime studies, aiming to determine the validity of the CIT. Results suggested that the accuracy rates of the test accounted for between 80.6% and 95.9% respectively. In addition, in 11 of these 15 studies no false-positive errors occurred. Vrij (2008) presented different reviews which summarise the accuracy rates of the

CIT in laboratory studies, reporting that the correct mean rates ranged between 78% and 86% for guilty subjects, and between 94% and 99% for innocents.

Although many studies have provided optimistic results regarding the validity of the CIT, its practical application remains limited (see Section 1.1), leaving space for the application of less reliable tests, such as the Control Question Test. For these reasons, Ben-Shakhar et al. (2011) have suggested that a more extensive application of the CIT for practical contexts can be obtained only if further research upon the validity and the effectiveness of the CIT is conducted. In fact, the literature provides information regarding the general validity of the CIT, but not around its real effectiveness in practical contexts, causing a general lack of communication between the scientific and the practitioner communities. Therefore, it is necessary to provide a greater awareness of utilisation of the CIT as a practical tool.

2.2.10.3.1 Strengths of the CIT

One of the strengths of the CIT is the examiner control of levels of the false positive effect. The false positive effect occurs when the polygraph detects truthful examinees as deceptive. Two factors can determine this effect: firstly, the test's properties can have a massive impact on results; false positive results are inversely related to the number of questions in the test. Secondly, the examiner can set the minimum number of correct alternatives the suspect is required to answer before being considered deceitful. The control over this mechanism is related to the low probability that this will happen by chance. Therefore, this control over false-positive probability has essential implications. This represents the main reason why it is suggested that the CIT can certainly be useful within investigations (Meijer & Verschuere, 2010).

2.2.10.3.2 Limitations of the CIT

Although the CIT is considered to be valid compared to other tests, different limitations are an obstacle to an accurate investigation of the test's mechanisms and its complete effectiveness. One such obstacle is represented by the limitations of the laboratory studies, which do not provide optimal results concerning ecological validity. Unfortunately, the only two studies that tried to estimate the accuracy of the CIT in real contexts were conducted in the early 1990s (Elaad, 1990; Elaad et al., 1992); these involved CIT examinations conducted from 1979 to 1991. Both studies agreed that the Electrodermal Response (EDA) is a good indicator of concealed information as well as the Respiration Line Length (RLL), which appeared to be efficient for the discrimination between guilty and innocent participants in CITs. More specifically, they reported 27% correct detection of crime-related questions to guilty examinees and 91% correct detection of questions to innocents. Hira and Furumitsu (2002) conducted another field study in Japan based on confessions and other physical evidence. Results showed that 74% of deception in guilty examinees was detected; furthermore, it obtained a perfect detection rate for innocent examinees. However, several uncontrolled factors play a crucial role when dealing with real-life CITs. The above information relates directly to aims 1, 2 and 3 (For more information about the aims of the project, see section 2.4).

2.2.10.3.2.1 Field studies Vs. Laboratory studies

The first factor relates to the time elapsed from the actual crime to the test. Most of the studies about the CIT are conducted with mock crimes, and this presents a limitation, as during examinations in a laboratory setting examinees are usually tested immediately. However, in a real-life context, a suspect might be examined after a more extended period of months or even years, with the consequence that information related to the crime might decay (Elaad, 1990). Elaad (1992) conducted a study in which participants were tested between two days and a week after committing the mock crime. The outcomes from the study showed that many participants could not remember some of the critical information used during the mock crime. Furthermore, in a study conducted by Carmel et al. (2003), they manipulated the encoding of the relevant information and the time of the test, including two type of items, peripheral and central. Results show that central items were better recalled after one week, evoking greater EDAs than peripheral items.

Perception and retention of information in a real-life situation are more complex, as more elements can influence these processes, such as time pressure, personal interest in the information, exposure to complex scenes and contexts (Elaad, 2011). For this reason, it might be even more difficult to be able to remember certain details of a crime that should be used as central information in a CIT examination. In respect to this concern, Verschuere et al. (2007)

stated that research findings obtained from mock crimes might not represent the real-life situations fully. For this reason, they conducted a study which compared the results obtained from real-life studies and mock crimes. They suggested that real-life situations are more stressful than laboratory contexts; therefore, they considered the laboratory studies as not extremely representative of real life, lacking eco-validity.

Finally, it has been suggested that CIT results from laboratory settings better generalise field conditions compared to the other polygraph studies (Rosenfield, 2018). In fact, as the CIT is based on detecting memory instead of deception, the major factors that can affect the reliability of the laboratory tests are those that can impact the encoding and consolidation of information, such as stress levels. Researchers have found that when using experimental designs, applying incidental instead of explicit encoding (Carmel et al., 2003), or enhancing arousal in participants (Rosenfield, 2018), the CIT can show high levels of validity even in laboratory settings. Therefore, results can be generalised in real-life settings.

This difference has crucial consequences on the process of lying. In fact, in a mock crime a fake deception is created in an artificial environment which in reality is not really threatened; therefore, the motivation and the consequences that deception produces during an examination are more deeply felt in real-life situations than in a laboratory setting (Verschuere et al., 2011). Consequently, this causes weaker physiological responses to questions during the test. Elaad, Ginton and Jungman (1992) conducted a study about the external validity of the CIT through the analysis of RLL and EDA in a real-life setting. They found that levels of arousal were higher in real-life settings. These results can be easily supported by the fact that during real examinations, examinees experience a real fear of conviction, an essential condition that does not occur in a laboratory setting. However, as discussed in Section 2.6.2.2, successful detection rates could also be found under low motivation conditions, which is usually more typical in a laboratory setting. Therefore, relevant studies about the validity of the CIT could also be obtained during an examination conducted in an artificial environment.

2.2.10.3.2.2 Leakage of information

It is important to discuss how certain limitations of the CIT can have a detrimental impact on the validity and applicability of the test. These limitations concern the questions design stage related to the issue of leakage of information. Sometimes it can be difficult to find crime-related information that would be known only by investigators and the offender (Krapohl, 2011). For designing an effective CIT, proper items need to be selected. Usually, suitable items relate to information about crimes that were kept secret from the media, unauthorized people and even from interrogators. In the same way, these items relate to details that guilty suspects are likely to recognise while committing the crime and recollect when undertaking the CIT (Bradley, Barefoot, & Arsenault, 2011). When a crime occurs, it is common to release some information to the media. In these situations, innocent people could be accidentally aware of some key details of a crime, but may not be able to explain why they know that information, thus potentially incriminating themselves (Bradley, Barefoot, & Arsenault, 2011).

Furthermore, guilty suspects could claim the fact that their knowledge of details is related to leakage of information by the media (Krapohl, 2011). Therefore, the leakage of critical information might put innocent suspects in great danger, as the knowledge of certain information might be sufficient for enhancing physiological responses (Bradley, Barefoot, & Arsenault, 2011). Consequently, this represents a determining factor that may produce false positive results in CIT. On the other hand, the leakage of incorrect information to guilty participants might be responsible for the distortion of their perception of the crime (Elaad, 2011). In the past, one of the most common police tactics for eliciting confessions consisted of introducing additional or incorrect information during an interrogation to intimidate suspects or to extract more information from their admissions (Leo, 1992). However, this incorrect or new information can compromise suspects' confidence in the information they possess about the crime. Therefore, in some cases, inaccurate information might be responsible for the large false negative results from the examinations (Elaad, 2011).

Consequently, this limitation can also have repercussions on the capacity of the test's applicability. In fact, Podlesny (2003) reviewed 758 criminal cases of the America Federal Bureau of Investigation (FBI) conducted with the CQT. The analysis suggested that because of the examinee's awareness of crime-related details, the CIT could not be used in 707 cases. The application of the CIT in real-life investigation, therefore, raised difficulties, especially because in order to protect crime-related information from suspects, the collaboration of all the investigators is required (Lykken, 1988). However, Podlesny's assessment was based on a retrospective review of criminal cases. This process underestimated the potential application of the CIT, as the information contained in the records was not selected with a CIT in mind (Meijer & Verschuere, 2010).

Lykken (1988) accused investigators of not being accurate enough when retrieving fresh information from the crime scenes. He suggested that all the investigators should collaborate in concealing certain details of a crime from suspects and the public. Bashore and Rapp (1993), supporting the validity of the CIT, suggested that examiners and investigators should not only avoid any information leakage, but also have access to an appropriate variety of details regarding the crime. Finally, Bradley et al. (2011) supported the validity of the CIT in real investigations, suggesting that the CIT can discriminate anyway between innocent suspects who are accidentally aware of certain crime details and guilty suspects.

2.2.10.3.2.3Habituation effect

Another important limitation that has been demonstrated to affect the accuracy of the CIT is the *habituation effect.* As mentioned previously, this effect occurs when repeated presentations of a specific stimulus are given, producing an internal representation that contains all the information of that stimulus. When the stimulus is presented again, the input information received is compared with the internal representations formed, called a “neuronal model”. When a mismatch between stimulus input and internal models takes place, then the OR occurs; rather, if the input information matches the existing neuronal model, habituation occurs (Sokolov, 1963). In polygraph examination this means that examinees’ physiological reactions will habituate to the stimulation (even when it contains the critical item), decreasing their responses. This represents a key limitation of the validity of the test, as examiners usually repeat the questions at different times, for standardizing and validating their conclusions (Iacono, 1991). In other words, a deceitful examinee can start to show decreased physiological reactions to the relevant items presented already after the second series of questions (Ben-Shakhar & Elaad, 2002). However, it was suggested that this effect could be reduced by presenting the relevant items close together in the sequence of questions for an equal number of times (Podlesny & Raskin, 1977). On the other hand, as previously noted, the CIT cannot be explained only by the OR to novel stimuli. In fact, if the OR is considered in relation to personal significance of the stimulus, it should be logical to assume that the occurrence of habituation to a stimulus could be caused by the presentation of personally significant items

(Verschuere, Crombez, Koster, & De Clercq, 2007). However, the familiar significance that in some cases might cause habituation is also an important key of the mechanism underlying the measuring of physiological responses during a CIT examination. This is why Verschuere et al. (2007) considered the inhibition as a factor with double effects, which needs to be considered during the whole process of measuring the physiological responses. In fact, they suggested that deceitful examinees usually try to inhibit their emotional arousal when coming across an item that is relevant to them. However, this precise intention of attempting to inhibit the physiological response triggered by their memory (such as agitation) actually increases the physiological response. Moreover, after this initial intent, the inhibition of the response occurs naturally, due to the habituation to the relevant stimulation. Many authors confirmed this effect (Ben-Shakhar & Elaad, 2002; Podlesny & Raskin, 1977; Thompson & Amy, 2009); however, Ben-Shakhar and Gati's study (2003) investigated the impact of both verbal and visual stimulation on the occurrence of the habituation effect. They found that in both conditions the skin conductance tended to decrease by 25% in the signal value. Finally, it was also found that habituation tends to occur more often and with a stronger effect in laboratory settings, due to the lack of any real threat associated with the test (Thompson & Amy, 2009).

The habituation process is not considered a universal mechanism; therefore, it might occur in certain circumstances, with specific individuals and with specific test settings. Due to its complexity, this aspect represents one of the major limitations of the CIT, which requires additional theoretical contributions from research (Verschuere et al., 2011). For this reason, the present study will try to confirm the role of certain factors, such as the structure of the test, the nature of the questions, the length of the test and perhaps identify additional ones that could cause such limitation during the measurement of the physiological responses. By applying some of the guidelines provided by the literature for the development of the test, it will be possible to evaluate whether certain strategies are effective or not in decreasing this limitation. The above information relates directly to aims 2 and 3 (For more information about the aims of the project, see section 2.4).

2.2.10.4 The present Thesis

The above section presents a complete overview of the theoretical approaches that try to describe and explain the CIT mechanisms. The present thesis will use these assumptions and conclusions from past research to build a solid theoretical framework for this project, by using theories from both emotional-motivational and cognitive approaches. Although the literature shows that both theoretical categories present some limitations, the combination of both frameworks used in this project will help to provide a more robust theoretical explanation of the CIT. Furthermore, one of the main aims of this thesis is to demonstrate the applicability and validity of the CIT: this section evaluates different aspects of the CIT methodology that can be crucial for improving understanding around its validity. The review of these studies also raises a number of limitations, such as the habituation effect, limited availability of information and the level of information that suspects possess. The present thesis investigates, in depth, some of these limitations and potential new applications of the test, by providing additional contributions around the use of the CIT in forensic contexts. Finally, studies have suggested that the CIT is potentially able to detect intentions to commit crime. However, few studies have explored this application of the CIT, leaving a gap in the literature. For this reason, the thesis will demonstrate this new ability and application, and thus set out the implications of using the CIT as a preventative tool (for more information about the aims of the project, see section 2.4).

2.3 Memory and Deception

As introduced in the previous section, deception is an adaptive behaviour, a tendency that develops naturally during our life, employed constantly in our daily social interactions. Despite the natural tendency with which we produce lies, the process of deception is cognitively demanding and complex (Vrij, Fisher, Mann, & Lean, 2006). Lying requires the recollection of truth and the suppression of facts, the communication of an elaborate lie, appropriate information related to that lie and changes in behaviour (Ruffman, Murray, Halberstadt, & Vater, 2012). Hence, this process is strongly related to memory and its functions. We constantly rely on information we cognitively store, which is vital for our daily activities and daily relations with others. Therefore, memories are central to when individuals want to make false claims.

Different studies (e.g. El Haj, Antoine, & Nandrino, 2017; El Haj, Moroni, Luyat, Omigie, & Allain, 2014) have tried to determine the relationship between deception and memory, finding that it mainly involves episodic memory, which is responsible for containing autobiographical events (Tulving & Markowitsch, 1998). According to El Haj, Antoine and Nandrino (2017) a strong correlation exists between deception and the ability to remember to whom a piece of information had been told in order to avoid inconsistent versions. Therefore, it is important to investigate the main features of the memory, its functions and how its mechanisms are related to the deception process.

The neuroanatomy of the memory includes many anatomical structures in the brain which are responsible for different cognitive processes associated to memory functioning, essential for deception processes.

The sub-cortical anatomical brain structures involved in memory functioning are:

- Hippocampus - a complex brain structure situated in the temporal lobe. It is responsible for memory and learning tasks (Ward, 2009). Its main role relates to consolidation of information and spatial memory. In fact, it contains cognitive maps, which enable human navigation (Kolb & Whishaw, 2008). In addition, fMRI studies have found that the hippocampus is activated during the recognition of true/false events or during the process of pretending to know certain information (Abe, Okuda, Sasaki, Matsuda, Mori, Tsukada, & Fujii, 2008). Finally, the hippocampus is involved in the neuronal model process where novelty-sensitive and familiarity-sensitive neurons operate (Sokolov, 2001). Therefore, this component plays an important role, especially in the CIT mechanism, based on a recognition process.

- Cerebellum - an essential structure located in the rear of the brain, which is responsible for the learning of procedural memory and motor learning (Ward, 2009). It is involved in the cognitive aspects associated with short-term memory and decision making. For this reason, the cerebellum is one of the primary components involved in processes related to memory deception (the primary effect of lying and dishonesty) (Kozel et al., 2004; Ganis et al., 2003).

- Amygdala - a collection of nuclei situated in the temporal lobe. It is involved in emotional learning and memory. It is responsible for encoding emotional memories and enhancing them, such as traumas, anniversaries or a loss (Kolb & Whishaw, 2008). It has been demonstrated that the amygdala, and other brain regions, are responsible for emotional processing and social interaction during deceptive behaviours. In addition, studies suggest that emotional factors can affect the encoding and consolidation processes (McGaugh, 2000). This has crucial implications in the way in which certain information will be recollected during a polygraph examination (for more information, see section 2.3.4).

While the cortical anatomical brain structures involved in memory functioning are:

- Frontal Lobe - located in the cerebral cortex. It is mainly involved in the functions of the working memory, such as coordination of information, emotional expression and problem solving. It is responsible for selecting memories and organising them into a coherent memory trace. For this reason, it is involved in the cognitive process of lying, as this mainly involves managing and reorganising information which will be used for deceiving others (Ward, 2009).
- Parietal Lobe - located above the occipital lobe. It is responsible for constructing a spatial coordinates system to represent the world around us. It allows the activation of attention when required, provides spatial awareness and navigation skills. Similar to the frontal lobe, the parietal lobe is involved in the activation of certain deceptive mechanisms, helping to organise and coordinate information (Ward, 2009).

- Occipital Lobe - the smallest lobe in the human cerebral, located in the rearmost part of the skull. The primary function of the occipital lobe is vision. It processes everything it perceives as visual stimulation (Ward, 2009). This component is essential for processing visual information that can then be used as key information for recognition tasks, for example during CIT examinations.

2.3.1 Memory and polygraph

As was suggested by Tulving (2000), memory has different sub-memory systems, such as working memory, short- and long-term memory, visual working memory, etc. Each of these components performs different memory operations, and some of these are involved in lying and deceptive processes. For this reason, the table below briefly presents those memory components that play a crucial role in mechanisms of deception and particularly explains their role in the CIT.

Table 1. *Memory components and their function for polygraphing and CIT mechanisms*

Memory component	Description	Memory function for polygraphing and CIT mechanisms
Long-term memory	Compared to short-term memory, it can store a bigger amount of information for an unlimited duration (Atkinson & Shiffrin, 1968; Brady, Konkle, & Alvarez, 2011). It consists of sub-components: episodic, semantic and procedural memory. For more information, please see Appendix I.	Long-term memory allows individuals to retrieve information stored or activate relevant elements and make them available for the working memory. Based on this information and the liar's goal, individuals can decide whether to lie (Sporer, 2016). This mechanism can be easily applied in a polygraphing session, when individuals are asked to think about information about a case.
Working memory/Short-term memory	The ability to process multiple pieces of information, continually update memory contents with incoming stimuli, and recall the appropriate information (Lustig, May, & Hasher, 2001; Miyake, Friedman, Rettinger, Shah, & Hegarty, 2001). For more information, please see Appendix I.	Working memory is involved in lying processes due to its role of keeping multiple pieces of information together, such as details about an offence; this information can then be used again in future and manipulated, (Alloway, McCallum, & Alloway, 2015) For example, in a polygraph session, the interviewee would rely on working memory to update his or her responses with follow-up questions from the interviewer and shift between the real perspective and the fabricated reality that the interviewee had constructed in order to avoid detection. Working memory is responsible for processing multiple pieces of information involved in the actual lie, but not in the prior engaging in lie-telling behaviour (Alloway, McCallum, Alloway, & Hoicka, 2015).
Visual working memory	The visual working memory is a cognitive mechanism which stores a limited amount of visual information so that can it can be easily retrieved for ongoing tasks. The capacity of this memory depends on many individual characteristics, such as	Visual working memory would help to retrieve visual information. The potential of this memory can be used for retrieving certain information with the use of visual stimulation (e.g. pictures form a crime scene, forensic evidence). This can help to strengthen the process of recognition and familiarisation with certain information during a CIT session (this section is relevant for aims 1 and 4 of this thesis. For more information about the aims of the project, see section 2.4).

	intelligence and cognitive abilities (Dube et al., 2017). For more information, please see Appendix J.	
Visual recognition memory	This memory system is responsible for responding to novelty stimulation and recognising previously encountered events, objects or people (Medina, 2008). For more information, please see Appendix K.	As the CIT is considered a memory detection test (Meijer, Verschuere, & Merckleback, 2010), the mechanisms of recognition of information are crucial. In fact, the CIT aims to detect whether the individual possesses certain information via two processes involved in visual recognition memory functioning: recollection and familiarity. Recollection involves the retrieval of information associated with a previously experienced event (e.g. a crime). The second one relates to familiarity, which is the feeling that the event was already experienced, without recollection occurrence (Mandler, 1980) (e.g. in the case of the CIT, an individual can recognise information familiar to him/her, such as weapon or crime scene details. This section is relevant for aims 1 and 4 of this thesis. For more information about the aims of the project, see section 2.4).
Sensory memory	Sensory memory allows individuals to recall impressions of sensory information collected through different sensory channels once the original stimulus has occurred (Coltheart, 1980). For more information, please see Appendix L.	CIT mechanisms work around the power of recognition of certain information. As discussed in Appendix L, sensory information is stored with specific modality depending on the type of sensory channel used. Therefore, every store (e.g. visual, tactile, auditory) has slightly different durations (Persuh, Genzer, & Melara, 2012), affecting the strength of the priming process prior to the CIT and eventually the intensity of the physiological response. In other words, different elements/information that are present during a crime can be stored with different duration/intensity, depending on the nature of the information and the sensor channel used. This can have implications on how certain information will be retrieved and recognised during a CIT interview via physiological responses. Visual input creates the strongest recall value compared to the other senses, allowing the most extensive spectrum levels of processing modifiers (Intraub & Nicklos, 1985). This outcome can be explained by the fact that pictures and visual stimuli have an excellent potential for both physical and semantic richness. Equally, tactile memory representations create powerful recollections. Auditory stimuli follow conventional levels of processing rules, while odour memory has the weakest capacity of recall memories (Vaidya et al., 2002) (this section is relevant for aims 1, 2 and 4 of this thesis. For more information about the aims of the project, see section 2.4).

2.3.2 Memory and emotions

Emotions impregnate our daily lives; in most cases, our memory is triggered by emotional information, therefore emotions can enhance the ability to remember not just the occurrence of the event, but also where and when it occurred (Schmidt, Patnaik, & Kensinger, 2011).

Williams and Conway (Williams & Conway, 2008) have shown how the liveliest autobiographical memories tend to be related to emotional events, which have a higher probability of being recalled than neutral events (Williams & Conway, 2008). In support of this assumption, the binding hypothesis states that 'emotional reactions trigger binding mechanisms that link an emotional event to salient contextual features such as event location' (MacKay & Ahmetzanov, 2005, p. 26); in other words, high–arousal information is remembered with more contextual detail than lower-arousal information.

This mechanism can be explained by an analysis of the amygdala's role. As mentioned previously, the amygdala specialises in the processing of emotions (see section 2.3). During the encoding stage, a stimulus needs to be encountered. Although several factors can influence how well a stimulus is encoded, it has been demonstrated that emotion can affect attention from the stage of capturing until the processing stage, especially when attention is limited (Fox, Russo, Bowles, & Dotton, 2001).

During the second stage, which deals with the storing of the information, studies (McGaugh, 2000) suggest that the amygdala has a crucial influence. After the encoding stage, the information is very fragile for a small period of time, therefore it tends to decay. Thus the process of consolidation of these memories takes time. It has been found that one of the reasons for this slow consolidation process is to allow an emotional reaction to influence the storage of that event/information. Due to the event always being followed by an emotional response, such as arousal and the release of stress hormones, events that evoke emotional responses are also more likely to be remembered later (McGaugh, 2000). Finally, regarding the influence of amygdala in the long-term memory, it has been shown that there is a correlation between amygdala activity at the moment of encoding and later memory for emotional stimuli (Kirkpatrick & Cahill, 2003).

The influence of strong emotions has a significant impact on our memory, and this can be seen through research involving the study of factors that can influence memory in eyewitness testimony. Stress or trauma during a particular event can affect the encoding of the memory causing, in some cases, the repression of those memories from conscious awareness (Christianson & Loftus, 1990); this becomes exceptionally crucial when trying to conceal specific information that individuals also associate with strong feelings. However, difficulty in accessing specific memories can occur after, for instance, episodes of child sexual abuse or after being witness to or victim of a severe crime (Loftus, 1993).

Eyewitness research has shown that increased emotional arousal determines attentional narrowing (Christianson, 1992). This process causes superior encoding of central details, undermining the encoding of peripheral information. For example, one of the most stressful factors that has a substantial effect on the memory of the victim or eyewitness is the presence of a weapon during a crime. “Weapon focus refers to the visual attention that eyewitnesses give to a perpetrator’s weapon during a crime” (Steblay, 1992, p. 3). In other words, when an eyewitness focuses their central attention on the weapon, the ability to adequately encode and later recall peripheral details decreases. Remington, Johnston, and Yantis’s study (1992) suggests that weapon focus is an unintentional and “automatic capture”. However, other research recommends that this mechanism is avoidable, primarily if the attention is focused on something else (Yantis & Jonides, 1996). In addition, post-traumatic stress disorder can strongly influence explicit memory, causing severe difficulty in recalling specific traumatic events (Amir, Leiner, & Bomyea, 2010). Therefore, central details are better encoded than peripheral details (Meijer et al., 2011).

It is also possible to evoke emotional memories from sights, sounds and smells. In everyday life, individuals often associate sensory stimulation with emotions, for example the scent of gingerbread might evoke joyful Christmas feelings or the sight of a spider might trigger feelings of fear and panic. Therefore, sensory brain areas are not just involved in the detection of specific sensory information, but they are also included in processing emotional memories, especially those associated with particular fear and anxiety disorders (Sacco & Sacchetti, 2010). In one study conducted by Sacco and Sacchetti (2010), it was found that specific sensory information, for example a particular sound, is acquired and coupled with emotional information, such as the memory of fear or joy. This information, then, is stored in the auditory cortex as a bunch. Through this process, a specific sound acquires an emotional meaning. Therefore, the same sound can trigger the same emotional state experienced when it was acquired. One of the most typical examples can be found in the process that creates a trauma. Similarly, some of the polygraph tests (such as the Concealed Information Test) try to evoke physiological responses by proposing information that might be associated with emotional elements; when this is the case, it should be more difficult to hide physiological responses.

This overview provides essential information about the structure of the memory, and how relevant information is stored. It is clear how memory is directly involved in the cognitive processes underlying the intent of lying. It was discussed how an individual can save many different types of information, such as visual and auditory sensory information; all these elements are perceived and then stored in different ways. This section, therefore, shows that many and various factors could determine the way in which a piece of information can be memorised and then recalled. Polygraph tests are based on the use of memory for triggering physiological reactions during examinations; therefore, it is essential to understand when and why specific information can be perceived, stored, recognised and recalled after a stimulus. This process is crucial for explaining why an individual would physiologically react when stimulated by a piece of information during a polygraph examination. The above section is relevant for aims 1, 2, 3, and 8 of this thesis. (For more information about the aims of the project, see section 2.4).

2.3.3 The present Thesis

This section presents a brief overview of the role that memory has in processes related to deception. As one of the main aims of this thesis is to explore the validity of the CIT, this section helps to provide the relevant theoretical framework necessary for understanding the role of mechanisms such as encoding, storing and retrieving critical information before and after the CIT. This analysis will help to provide a clearer understanding of how certain events, objects, and experiences could impact memories, affecting the strength of certain physiological responses. With the use of concepts like the visual recognition memory system, it will be possible to reveal new information regarding how the recognition of accurate information during a CIT examination produces stronger physiological responses compared to information that differs in nature or typology (for more information about the aims of the project, see section 2.4).

2.4 Purpose of the Thesis

Different studies have found that the CIT can discriminate guilty from innocent suspects with a higher accuracy compared to the other polygraph tests (Ekman & O'Sullivan, 1991; Bull et al., 2004; Grubin & Madsen, 2005). However, CITs have not been extensively applied in real forensic contexts, due to the lack of demonstration of its applicability. As Ben-Shakhar et al. (2011) has suggested, a more extensive application of the CIT for practical contexts can be obtained only if further research upon the validity of the CIT is conducted. For this reason, this study will try to increase awareness and knowledge about the limitations and strengths of the test, proposing new practical applications.

In particular, this study will evaluate certain aspects of the validity of the CIT that received less attention: the effect of visual stimulation, the ability to detect criminal intentions and the application of the CIT in a group setting. Results from this study will provide further evidence of the applicability of the CIT as an investigative and preventative tool.

2.4.1 Aims of the Project

Validity and methodology:

In order to demonstrate the applicability of the CIT, further research upon the validity of the test needs to be conducted. For this reason, one of the first aims of this study concerns the evaluation of CIT's methodology. This thesis will investigate how to increase examinees' physiological reactions through the additional support of an alternative stimulation method: visual stimulation. Despite the amount of photographic evidence available in forensic cases, not many researchers have explored the potential of visually stimulating examinees during a polygraph examination. In addition, the literature shows contrasting opinions regarding the actual effectiveness of this methodology and a lack of application of this methodology in group settings. As the CIT is based on a simple mechanism of recognition of familiar information, photographic material from investigations could be used as central information contained in the test. For these reasons, it is necessary to develop additional studies that could help to clarify the effectiveness of visual stimulation, especially in group contexts.

Memory processes and detection of intentions:

To explore the validity of the CIT, it is also necessary to investigate the mechanisms involved in encoding, storage and retrieval of critical information before and after the test. Therefore, the purpose of this study focuses on assessing whether memory can be influenced by specific factors involved during a particular event/stimulation. Thus, the study will try to determine whether specific elements can have an impact on examinees' physiological reactions during a polygraph examination. Every experiment in this study will try to measure, analyse and control factors suggested by the literature, which could affect participants' memories. This will provide a clearer understanding of how certain events, objects, and experiences could impact memories, and thus why specific memories associated with particular information are stronger than others. Consequently, this could reveal new information regarding how the recognition of accurate information during a CIT examination produces stronger physiological responses compared to information that differs in nature or typology. In addition, results from this study will help to determine whether some of these cognitive processes can also be effective when trying to detect information about an intention to commit an action. In fact, the literature around this new application is very limited and not very extensive. For these reasons, this study will provide additional understanding of the effectiveness of memory mechanisms also involved in the recognition of intentions to commit an action. These results will certainly improve the understanding of the mechanism underpinning the CIT, providing greater awareness and new suggestions of the use of the CIT as a preventative tool in real forensic contexts.

Group Application:

The last main purpose of this study is focused on examining the effectiveness of the CIT to detect criminal intentions in a group setting. Almost all the studies conducted on the polygraph tests focused on solo offenders' features. However, in recent decades, the escalation of organised crimes has pushed the scientific community to concentrate more on new technologies for counter-terrorism and prevention of organised criminal activities (Roshdale, 2001). Only a few studies have tried to explore this new aspect of the validity of the polygraph. However, none of them has used the methodology proposed in the present study. For this reason, this present study will determine the validity of the CIT when trying to detect the concealment information from a group of participants. The investigation of a group setting opens new applications for the CIT, which can be used on potential suspects involved in criminal organizations or criminal groups. Therefore, the results from this study will provide an important contribution to the tiny portion of literature around this area. Furthermore, this study will try to offer new suggestions and evidence for the application of the CIT not only for solo offenders, but also for criminal groups.

The general aims of this thesis are:

1. Determining whether the CIT is a valid method for detecting relevant information
2. Determining which factors can affect participants' physiological reactions
3. Determining which are the main limitations of the CIT
4. Determining whether visual stimulation can improve the physiological reactions of guilty participants
5. Determining whether the CIT can detect criminal intentions
6. Determining whether the CIT can detect criminal intentions from a group of guilty participants
7. Determining the false positive levels on innocent participants
8. Determining which are the main implications of the results obtained from this study

2.4.2 Aims of the Studies and Hypotheses:

Study I:

This exploratory experiment will investigate the validity of the CIT by assessing the influence of sensory memory. The study will then try to trigger physiological reactions when trying to hide information during a CIT through a standard verbal stimulation.

Hypotheses of Study 1:

- Accuracy rates in the CIT will be above the level of chance
- A higher deception level in the experimental group will be found.
- A low false positive effect will be found.
- Habituation will reduce the accuracy rate of the test during the examination.
- Greater physiological responses during the relevant questions will be found
- A various in the accuracy rate between the questions series will be found

Study II:

This experiment will investigate the validity of the CIT by exploring the influence of sensory memory (similar to Study I). However, this study will try to trigger physiological reactions when trying to hide information during a CIT through a verbal stimulation combined with visual stimulation. Results will then be compared to Study I, to see any possible increments in the accuracy of the test. This study is unique compared to the others in this area because the literature has mostly ignored the potential effect of visual stimulation applied through the CIT.

Hypotheses of Study II:

- Accuracy rates in the CIT will be above the level of chance
- A higher deception level in the experimental group will be found
- A low false positive effect will be found
- Low levels of habituation effect will be found
- Greater physiological responses during the relevant questions will be found
- A various in the accuracy rate between the questions series will be found
- Habituation will reduce the accuracy rate of the test during the examination
- A difference between the Mannequin Task and the Egg Task will be found
- A higher overall accuracy rate will be found in Study II compared to Study I, due to the application of the visual stimulation

Study III:

This experiment will investigate the validity of the CIT in a mock crime scenario, in which participants will need to plan a criminal act. In this case, the study will try to determine whether the CIT is able to detect criminal intentions. This study is unique compared to the others in this area because little attention has been dedicated to investigating whether the CIT can be used as a preventative tool for criminal activities.

Hypotheses of Study III:

- Accuracy rates in the CIT will be above the level of chance
- A higher deception level in the experimental group will be found.
- A low false positive effect will be found.
- Low levels of habituation effect will be found.
- Greater physiological responses during the relevant questions will be found
- A various in the accuracy rate between the questions series will be found
- A difference between the City, the Location and the Weapon will be found.

Study IV:

This experiment will investigate the validity of the CIT in a mock crime scenario (similar to Study III). However, in this study, participants need to plan a criminal act in a group setting. In this case, the study will try to determine if the CIT is able to detect criminal intentions from a group of mock suspects. Therefore, results from this study will be compared to Study III, which will allow us to determine in which experimental setting the accuracy rate is higher. This study is unique compared to the others in this area because only a few studies have investigated whether the CIT is able to detect criminal intentions in a group setting. However, no previous studies have ever examined this criminal aspect with the methodology proposed in this study.

Hypotheses of Study IV:

- Accuracy rates in the CIT will be above the level of chance
- A higher deception level in the experimental group will be found.
- A low false positive effect will be found.
- Low levels of habituation effect will be found.
- Greater physiological responses during the relevant questions will be found
- A various in the accuracy rate between the questions series will be found
- A difference between the City, the Location and the Weapon will be found.
- A difference between the roles assigned to the participants will be found.
- An increase in the accuracy rate will be found, compared to Study III.

CHAPTER 3: METHODOLOGY

The main objective of this PhD thesis is to examine the validity of the CIT in different ways.

This chapter provides a summary of the methodology adopted to plan and design the experimental phase, together with the description of the data collection and analysis. The aim of this chapter is to provide a general comprehensive understanding of the experimental stages of this study. The choices made during each phase are motivated by critically presenting the existing literature review, highlighting potential and limitations of the experimental procedure.

The first section provides information regarding the recruitment methods used to select the participants for the experiments. Information about the participants involved in the study are included.

The second section contains the details about the equipment used, which includes the polygraph machine, a laptop, the five sensors used for the recording of physiological reactions and the software used to analyse the data. Sensors are described in detail, to help to understand how each physiological reaction has been recorded and then analysed. A brief sub-section provides an overview of the measurements used in this study: cardiovascular, respiratory and electrodermal measurements. More specifically, this section evaluates the physiological mechanisms at the base of the measurements. This overview covers the existing literature regarding the utilization of these measurements in studies involving the use of the polygraph.

The following section illustrates the questionnaire design: a detailed explanation of the questions involved in the test is presented. Understanding the structure of the questionnaire and the importance of every single question is essential for the methodological section. By describing the features of every question included in the CIT, it is possible to understand the basic mechanism behind the examination. In other words, this section describes how to make a question able to trigger physiological reactions in examinees.

Equally important is the process of developing tasks for this study, which have the purpose of priming participants prior the examination. A detailed description of the process involved in developing an effective context in which participants can lie in a standardized and controlled environment is presented. Therefore, it is possible to understand the difficulties and limitations behind the experimental process involved in the current study.

The last section focuses on the procedure adopted for the current study. It contains three sub-sections describing the three stages of the experimental process. The priming stage, which contains details about how participants were prepared to lie about information that were provided by the research team. The examination stage: which presents details regarding the pre-examination and the actual administration of the CIT and the post-examination phase. Finally, the scoring section, which covers the coding method and the analysing procedure applied for the data obtained from the examinations.

3.1 Participants and Recruitment methods

The total number of participants for this study is *N=325*. Undergraduate and master students that collaborated to the study as research assistants recruited the sample. As the present study was only interested in examining the validity of the CIT, no inclusion criteria were applied on participants. Diversely, the exclusion criteria were different: all the participants must have been over 18years old. As the polygraph examination procedure is based on testing participants by asking verbal questions and, in this specific case, by visual presenting information, participants who had physical disabilities, such as visual and/or hearing disabilities, were excluded from the administration of the test. In addition, participants with cardiovascular problems or blood pressure conditions were also excluded from the study, since the physiological reactions measured might be altered by such physical conditions.

As the present study was not interested in investigating how demographic information could affect the cognitive and memory mechanisms associated to deception, participants with a criminal background were not necessary. Although the implications of this study involve possible forensic applications on suspects and organised crimes, demographic studies conducted on terrorists' profiles found that the criminals associated to these types of crimes are usually well-educated, male or female, in their mid-twenties, from a middle-class background (Victoroff, 2005; Brannan, Eslerm and Strindberg, 2001; Canetti and Pedahzur, 2002). Furthermore, most identified/convicted terrorists came from middle or upper-middle class background and the majority had some college education (Crenshaw, 2000). For the above reasons, the sample collected for the current study (which includes members of staff and students from the University) could reflect some of the demographic characteristics of typical criminals, increasing the generalisability of the results.

Different recruitment methods were applied: at the first stage of the study, the experiment was advertised online through different social networks (Facebook, Twitter). The study was also advertised through emails, to reach students and members of the staff who were interested in taking part in the study. In addition, the study was advertised through flyers that were distributed around the University of Huddersfield and Huddersfield town centre. After the first Study 1, SONA System was also used for recruiting participants for the following ones. SONA is a research management System, which is used to monitor experiments and research in the Psychology Research Participant Scheme of the University. Through this system, it is possible to advertise a study by asking students to participate in research study in return for course credits.

3.2 Equipment

The physiological data was recorded with the *All in One Lafayette Polygraph* LX5000 *System* (Figure 1), owned by the department of Psychology at the University of Huddersfield. The equipment was purchased by the university from the official American Lafayette website (Lafayette Instrument Company, 2009). The Lafayette Polygraph is able to record, store and analyse physiological changes that occur during a polygraph examination. This system works at a transmission rate of up to 360 samples per second in all of the channels and 24-bit analog to digital conversion.

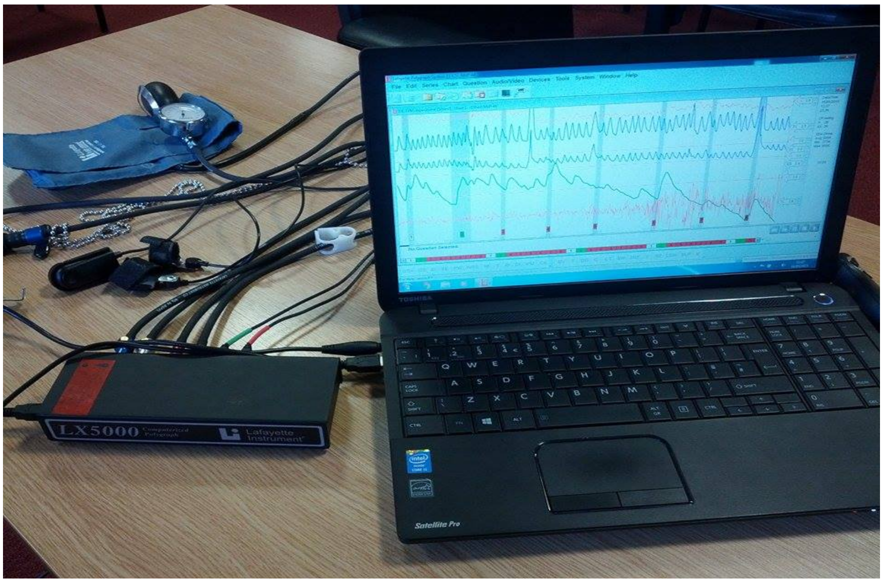

Figure 1. Example of a complete LX5000 Equipment

The Lafayette Polygraph System consists of three main components. The first one is a PC running the polygraph software, which is able to create and open polygraph files, enter and store information about the examinees, examine information, create and edit questions, record physiological data and create reports.

The second component is the Data Acquisition Subsystem (DAS). This system is used to record examination data by converting physiological signals into digital form.

The third component includes sensors, which consists of detection devices attached to the examinee in order to record his/her physiological reactions (Instrument, 2006). These devices are then connected to the DAS, which receives the input from the sensors attached to the subject and transfers it to the workstation. The sensors used within the Lafayette Polygraph LX5000 are:

- *Sphygmomanometer*: a normal blood pressure arm cuff and pump is used to record changes in the blood pressure and pulse rate.
- *Thoracic and abdominal Pneumatic:* upper and lower chest pneumatics, to measure the abdominal and thoracic respiration rates, through the variation of the pressure range.

- *Photoelectric Plethysmograph (PPG):* it is used to measure the volumetric changes occurring in pulse blood. This device utilizes a photosensitive cell to measure light reflected or passed through the tissue segment where the monitor is placed.
- *Snap Finger Electrodes(EDA)*: electrodes are placed in two different fingers, in order to record the electro dermal activity (Figure 2 and 3).

Figure 2. Example of sensors applied at the workstation.

At the top-right, it is possible to see both Thoracic and abdominal Pneumatics; at the top there is the Sphygmomanometer; at the extreme left Snap Finger Electrodes are place together; close to the electrodes, the Photoelectric Plethysmograph.

A polygraph chair is also part of the standard equipment. There are three sensors attached to the chair; two of them are within the seat padding, which are able to detect minor and major movements of an examinee during a polygraph examination. The third sensor (a small platform placed on the floor) records any movements from the feet of the examinee (Instrument, 2006). In order to test the efficacy of the movement sensors, a few trial examinations were conducted, and it was found that participants were struggling while sitting steady for the entire duration of the test, causing minor movements that were captured by the device, affecting the general results. Therefore, it was chosen to discard the use of any movement sensors for this specific study. The research objectives of this study were not focused on measuring and examining changes in movement of examinees during examination. The polygraph chair was anyway used to make the examination context more realistic.

In addition, as three sensors are attached to arms and hands, it has been necessary to provide a chair in which the examinee could stay steady and comfortable for the entire duration of the test.

Figure 3. An example of complete polygraph interview setting

3.2.1 Cardiovascular measures

Heart rate is an involuntary physiological function controlled by the autonomic nervous system. Specifically, it can be determined by counting the beats within a defined period of time (Schmidt & Segalowitz, 2008). The cardiovascular activity is measured through the cardio cuff, which is placed on the arm of the examinee. The machine is able to register the diastolic level of blood pressure during the examination. The examiner notes increase in the levels of blood pressure to each relevant question, which should cause a physiological change.

The utilisation of the heart rate in this field of psycho-physiological tests has a wide and varied history. Darrow (1929) represents one of the first researchers to measure heart rate in psychophysiological research. He suggested that a stimulus exposure could make a change in heart rate and blood pressure due to associative processes linked with emotion. Even Sokolov's (1963) orienting reflex is related to this mechanism. The Orienting Response is considered a primary response to a stimulus, which causes specific behavioural reactions. These changes occur in the central and autonomic nervous system, due to alterations in arousal levels associated with an environmental stimulus. Although Sokolov's research did not focus on heart rate, his research created the basic assumption for considering the heart rate as an acceleratory response in the orienting reflex (Reynolds & Richards, 2008). In fact, Lacey and Lacey (1958) assumed that increased levels of heart rate were associated with inhibition of cortical activity. They stated that in situations where the stimulation is unpleasant or painful, the physiological response increasing is more likely to occur. Thus, heart rate is likely to decrease with a reduction in sensory thresholds.

An infinitive number of factors influence cardiovascular activity, such as increases in mental and physical activity or the occurrence of any threats. Generally, any form of specific or general arousal can have an impact on cardiovascular processes, blood pressure and heart rate (Synnott et al., 2015). These factors can be also radically different amongst individuals when considering mental illnesses, anxiety disorders, physiological and biological conditions (Berntson et al., 1997).

Cardiovascular applied in the CIT. The complexity of the cardiovascular activity represents an important implication for polygraph examinations, especially if some techniques try to assume the association between certain cardiovascular patterns and deceptive behaviours. In the CIT setting, the use of cardiovascular levels for detecting deception is considered controversial. In fact, research regarding the validity of the use of cardiovascular measures have shown that cardio signals are not valid in the differentiation between guilty and innocent subjects (OTA, 1983; Podlesny & Raskin, 1977; Saxe, 1994). The central nervous system, for example, allows individuals to voluntary control their respiratory activity, which can have an effect in electrodermal response but also in heart rate (Honts, 1991). In addition, the pressurised cuff reduces blood flow, depending on the specific part of the arm in which it is applied, therefore, this can result in a less efficient and reliable measurement (Podlesny & Raskin, 1977).

In support of this, there is an actual explanation for the increased levels of the heart rate during an examination. In fact, where it is requested to verbally answer to a CIT question, a biphasic response can be seen on the chart, before heart rate returns to its baseline again. This response represents an initial acceleration followed by consecutive deceleration, which usually is more noticeable for relevant than non-relevant items (Bashore & Rapp, 1993). In cases where the examinee is silent, this acceleration is missing or less marked (Verschuere, Crombez, De Clercq, & Koster, 2004). Therefore, it has been suggested that this heart rate increase is related to the preparation of the answer, while the decrease in the heartbeat could be related to enhanced direction of attention towards the environment or to one's own bodily responses (Krapohl et al., 2005).

Typically, during the CIT heart rate decreases within around 3s after item onset and then increases momentarily (Matsuda et al., 2009; Verschuere, Crombez, De Clercq, & Koster, 2004; Verschuereet al., 2009). It then begins to decrease again around 5s after item onset, and this state lasts for over 10s. The first heart rate decrease reflects the sensory intake and information-gathering function of the OR to novel or significant stimuli (Bradley, Keil, & Lang, 2012; Lacey, 1967). While the second and longer-lasting heart-rate decrease would compensate for the increase in blood pressure caused by peripheral vasoconstriction, maintaining the homeostasis of blood pressure (Hirota et al., 2009).

Another cardiovascular measure used in CIT examination is the photoplethysmography technique. This sensor is able to register pulse rate and finger pulse amplitudes, which reflects the degree of constriction of peripheral blood vessels (Lykken, 1998). In order to evaluate the validity of this measure in a CIT, Podlesny and Raskin (1977) conducted a study in which it was found that relevant items could significantly arouse peripheral vasoconstriction. Conversely, no-relevant items were followed by a moderate reduction of finger pulse amplitudes. Although, it is important to differentiate that mechanisms underpinning the heart rate changes and peripheral vasoconstriction are slightly distinct. In fact, the sympathetic and the parasympathetic branches control the heart rate. The reduction on the heart rate mentioned above is, thus, resulting from a reduction of sympathetic activity or an increase of vagal innervation. By contrast, the vascular muscle is primarily controlled by the sympathetic nervous system. In CIT examinations, the relevant information results in a reduction of the heart rate and increase of peripheral vasoconstriction. This demonstrates that both branches of the autonomic system are activated in the CIT mechanism (Association, 2011).

Typically, during the CIT examination, pulse volume starts to decrease around 3/4s after item onset, and it reaches its minimum at around 7/10 s (Hirota et al., 2009; Matsuda et al., 2009).

The minimum is lower for the relevant item than for the irrelevant items. The decrease in pulse volume reflects peripheral vasoconstriction; due to the sympathetic-nerve activity activation (Sawada, Tanaka, & Yamakoshi, 2001). The decrease in cutaneous blood flow also reflects this activity. When individuals are exposed to a significant stimulus, their organisms gather blood in the trunk of the body by constricting peripheral blood vessels so that they can provide blood to the skeletal musculature of the limbs when action is actually required (Turpin, 1986). Although, it has been suggested that cardiovascular measures do not directly correlate with the mechanisms of the OR, reducing the validity of these measures during CIT examinations (Gamer, et al., 2008). In addition, cardiovascular responses behave differently when encountering habituation effects: while SCRs tend to decrease during the test, heart rate remain stable (Gamer, et al., 2008).

3.2.2 Respiratory measures

The respiratory apparatus is a biological system made by structures and organs used for the process of respiration. The respiratory system is responsible for the intake and exchange of oxygen and carbon dioxide between the environment and the body. The respiration rate consists in the number of breaths the individual takes in a specific time (Krapohl & Rosales, 2014).

Respiration represents one of those physiological processes that are easily influenced by emotions. For this reason, different studies have investigated the relationship between emotions and respiration, finding a strong correlation between emotional breathing and the amygdala (Elaad, 2011). During an arousal state, emotions increase human blood pressure and heart rate, while rapid breathing occurs (Iacono et al., 1992). Studies that investigated breath rates in a stressful state found that the sympathetic response causes increased blood flow, leading to an increased respiratory rate. This mechanism exists for enabling the body to maintain the O_2 and CO_2 stability inside the body (Iacono & Patrick, 1997). Although, changes of respiratory rate are also related to individuality.

Factors such as personality differences or levels of individual anxiety can affect the patterns of breathing in an individual (Elaad, 1998). Anxiety is considered as a basic defence mechanisms of the human body (Elaad, 2011); anticipatory anxiety, considered as the time between the warning presentation and stimulation, increases the respiratory rate. These changes, rather than being related to O2 consumption, are related to changes in metabolic demand (Elaad, 1998).

RLL applied in the CIT. In a polygraph examination, respiration rate is measured through pneumographs, which are applied around the examinee's chest and abdomen (Podlesny & Raskin, 1977). Because the stretch of these transducers in not calibrated, the signal produced by the sensors has an arbitrary unit. This makes it impossible a direct comparison between examinees (Bashore & Rapp, 1993). Although the correlation between the respiration rate and deception has been extensively researched, studies report that respiration rate is not considered as a valid measure for discriminating guilty from innocent participants in CIT (Bradley et al., 2011; Podlesny & Raskin, 1977). However, a different method of quantifying respiratory responses was proposed by Timm (1982), called respiration line length (RLL). Timm suggested that instead of scoring respiration rate and amplitude separately, it could be more useful to measure the total length of the respiration tracing in a specific period of time, following the stimulus presentation. In this way, it is possible to obtain a more integrative result of the respiratory activity (Bashore & Rapp, 1993). Studies, which tried to validate this method in polygraph examination, found that RLL is smaller for relevant items, compared to neutral items (Grubin & Madsen, 2005; Krapohl et al., 2005).

In the CIT, respiration is suppressed, resulting in a slower or smaller response from around 1s after the relevant item's onset, lasting for 20s (Kobayashi, 2011; Matsuda, et al., 2009; Matsuda, et al., 2011). During the CIT, instructions to inhibit physiological responses to affective pictures, such as slowing and shallowing breathing, lower the respiratory amplitude and rate (Dan-Glauser & Gross, 2011).

However, the main problem associated with the use of respiration rate in CIT settings is related to the capacity of controlling the respiratory rhythm (Iacono & Patrick, 1997). To bring respiration under voluntary control, examinees could try to manipulate their breathing pattern which could falsify the final results (Bashore & Rapp, 1993). In addition, the action of vocally responding to questions during the test is a behaviour that can break the pattern of breathing during the period in which that pattern would be recorded and then analysed. Therefore, some fragment of the RLL can be associated with the interruption of the respiration process (Iacono, 1991).

There are several doubts on whether respiratory measures can be considered an index of the OR (Barry, 1996); however, some elements have been found related to the OR framework, such as a decreasing of breathing after a novel stimulation (Stekelenburg & van Boxtel, 2002) and that respiration responses tend to habituate with stimulus repetition (Barry, 1983).

3.2.3 Electrodermal measures

The Electro-Dermal Activity (EDA) indicates the variation of the electrical properties of the skin in response to sweat secretion produced by glands, usually on the palms (Fowles et al., 1981). The process of measurement of the skin response is complex, especially because it is influenced by individual properties of the skin. In order to understand this process, a determinant distinction has to be made between *tonic* and *phasic* aspects of electrodermal measurement. The stable, almost unchanging, component is called “tonic”, while “phasic” is related to more rapid changes in few seconds (Schmidt & Segalowitz, 2008).

The second difference concerns “endosomatic” and “exosomatic” measures of electrodermal activity. The first one indicates those measurements of an endogenous potential across the skin. These measurements can occur by applying one electrode on the skin surface and the other at a site that has been abraded. The abrasion eliminates the barrier layer in the skin, therefore the electrode is placed in contact with the conductive interstitial fluid beneath the skin (Fowles et al., 1981). Because of the complexity of this measurement, endosomatic technique is rarely used in the field. Conversely, exosomatic measures try to evaluate electric properties of the skin by placing an external voltage. This causes current to flow through the skin and between the electrodes (Schmidt & Segalowitz, 2008). Thus, the tonic measure is called Skin Conductance Level (SCL), which slowly varies and changes on a time scale of tens of seconds to minutes. The rising and declining of the SCL depends on the individual respondent, on their hydration, skin dryness, or autonomic regulation. The tonic level can also differ markedly across individuals. This has led some researchers to conclude that the actual tonic level on its own is not that informative and for this reason it has not been used for this research (iMotions, 2017). Similarly, the Skin Resistance Level (SRL) represents the phasic measure (Fowles et al., 1981), which shows significantly faster changes compared to the SCL. This measure is sensitive to specific emotionally arousing stimulus events. These bursts occur between 1 to 5 seconds after the onset of emotional stimuli (iMotions, 2017). Due to its characteristics, this type of response has been used as the electrodermal measure in this study.

Between 1888 and 1890, Fere and Tarchanoff conducted the first investigations regarding the correlation between changes in electrical activity of the skin and emotional stimuli. Although in origin “Galvanic Skin Response” (GSR) was the term used to refer to all the electrodermal measures developed, recently the Electro Dermal Activity (EDA) replaced it. In the same way, Electro Dermal Response (EDR) is considered the most convenient way to indicate the phasic changes (Fere, 1888; Tarchanoff, 1890).

One of the main functions of the skin is to protect the organism from bacteria, parasites and chemicals. Although, another important purpose consists in thermoregulation. This mechanism occurs because of the dilation of blood vessels in the skin, which increase sweating, causing a decrease in the body temperature (Andreassi, 2007)

In regard to psychophysiology, the point of interest is the secretion of sweat from glands, which are differentiated in two categories, the apocrine glands, which are larger, usually placed in genital areas and the armpits. The other one is the eccrine, which represents the most important and most interesting type of sweat glands to psychophysiologists. They are distributed all over the body surface, with some exceptions. The areas in which they are most numerous are the hands and the soles of the feet (Larson, 1932).

Since different areas are responsible for the control and the production of EDA activity, EDA appears to be a complex mechanism which has been researched and investigated deeply. Boucsein (2012) dedicated different studies focused on the investigation of EDA and its complex processes. He suggested that EDA is mainly controlled by three systems related to arousal, emotion and locomotion. The hypothalamus, cingulate gyrus and hippocampus are responsible for EDA activity in relation to emotional responses and thermoregulation. The motor cortex is responsible for locomotion, while the reticular formation controls EDA in relation to arousal states.

The thermoregulation of sweating is also involved in changes of skin conductance in different situations, especially those that are emotionally and psychologically arousing. One of the most common and natural phenomena is the sweating in the palms and fingers in response to a strong psychological stimulation, such as anxiety or scare. Different studies have been conducted in order to investigate the correlation between the body sweating and emotional arousal. Some of these studies (Darrow, 1929; Matte, 1996; Reid & Inbau, 1977; Trovillo, 1939) considered the sweat gland activity adaptive, because it occurs in situations of emergency. For example, when the individual incurs in a stressful situation, the glands in the palms start sweating reducing their resistance, thus provoking a spike in the conductance reading (Backster, 1963a). In addition, the amount of sweat produced depends on the intensity and emotional significance of the stressor. This mechanism is activated when the sympathetic branch of the autonomic nervous system releases acetylcholine into those ducts that increases the amount of sweat in the glands (Backster, 1963b).

Therefore, the thermoregulation is also strongly related to human behaviour and the external stimulation. In fact, EDA is sensitive to a different type of stimulations; therefore, it is very important to understand the situation in which the response occurs. Being able to control the situation and the measurement of the response is determinant for the final interpretation of EDA.

SCR applied in the CIT. The SCR was applied for the first time as physiological measure for the CIT setting by Lykken (1998), demonstrating that phasic skin conductance changes can be used as a detector with a good validity. His early research demonstrated for the first time the impressive potential of the SCR in detecting concealed information.

In his research, Lykken (1959) asked some participants to commit two mock crimes, providing several crime-related information about the crime. Just through the recordings of the amplitude of galvanic skin responses to all the items, Lykken correctly identified 100% of the innocent participants and 88% of those who committed the mock crimes. Therefore, his results have shown that crime-related questions provoked a larger skin conductance response compared to neutral alternatives. As previously described, the skin conductance strongly relates with the sympathetic nerve activity; therefore, in a CIT setting, electrodermal responses are the product of a differential sympathetic activation after relevant and a non-relevant stimulus. This, of course, can occur just in situation where the examinee is able to discriminate the item categories (Bashore & Rapp, 1993).

After the development of the CIT, a number of studies tried to investigate the validity of the physiological measures applied to the CIT. Although there is no unique pattern of physiological response associate to deception (Daniels, 2002), different researchers found that electrodermal measures offer the greatest validity when detecting deceptive participants in a concealed information setting (Bradley et al., 2011; Council, 2003b; Holden, 2000; Saxe, 1994; Society, 1986). The research regarding the validity of electrodermal measures was not limited on mock crimes; in fact, other studies tried to evaluate the validity of this measure in different contexts, such as trying to detect autobiographical information (Barland, 1999), or memorized numbers and words (Beardsley, 1999) or in field studies to evaluate the ground truth of some confessions (Holden, 2000). According to a large meta-analysis conducted by Ben-Shakhar and Elaad (2004) in relation to the validity of the SCRs in the CIT, it was found that electrodermal measures had the highest validity with deceptive participants.

Typically, during a CIT examination, SCRs begin to increase between 1 to 3 s after item onset, and it reaches its maximum between 5 to 7s (Selle, Kindt, Vershuere Meijer & Ben-Shakhar, 2016; Hirota, Ogawa, Matsuda, & Takasawa, 2009; IMotions, 2017). The amplitude is greater for a relevant item than for irrelevant items. In other words, when an individual faces salient stimulation, hands and feet start to sweat, due to the surprise, and thus, prepare for action to approach or avoid the salient stimulus source (Bradley, 2009; Codispoti & De Cesarei, 2007). This process presents the same characteristics of the orienting response. This link between SCR and OR has led to the assumption that the CIT replies primarily on the OR, making electrodermal responses the most sensitive measure for this type of test.

3.3 Questionnaire Design

The polygraph examination consists in a series of questions aimed to cause physiological changes in an individual who is being deceptive. This examination is recorded by the DAS, in order to produce an output of the examination.

The CIT consists of a series of multiple-choice questions, which contains one relevant item among several non-relevant items (Iacono, 1991). For the purposes of this study, the use of relevant/non-relevant questions will be adopted. The relevant questions need to address the central details of the tasks: in order to conduct an effective CIT, the test should contain information reflecting the central essence of the event (Krapohl, McCloughan, Senter, 2006).

Each series was structured as shown below:

- One *Irrelevant question*: this is the only not crime-related question. It is always placed in the beginning of the series. It serves as initiating question to the test. This question can be general and it does not have to contain relevant information.

- One *Buffer question*: this question is always an incorrect alternative and it is always placed at the beginning of the series, after the irrelevant question. As the first item presented with the intent to provoke a response, the buffer question's function consists in absorbing the first orienting response reaction to the new set of questions (Krapohl, et al., 2006). Therefore, the relevant item should never be presented first. Consequently, the buffer question is always excluded from the analysis. In addition, the buffer question must be as plausible for innocent examinee as is the relevant item (Meijer, Verschuere & Ben-Shakhar, 2011).

- One *Relevant question*: this question contains the critical information and it is randomly located among the other plausible alternatives. As this type of question should evoke the orienting response of guilty participants, if it appears too salient compared to the others control questions, this will provoke a natural physiological response even in innocent participants. Therefore, it is important that relevant questions should not be more salient or more plausible than the other alternatives (Krapohl, et al., 2006).

- 4/5 *Control questions*: these questions are not critical but they contain plausible alternatives compared to the relevant one. In order to increase the efficacy of the CIT, it is suggested to select items that can be differentiated between each other (Krapohl, et al., 2006). Research about the CIT have been found, in fact, that the orienting response generalizes to items that share common characteristics with the significant item (Ford, 1996). The number of these questions can vary, however it is advised to insert a minimum of 5 questions (Meijer, et al., 2011; Verschuere et al., 2004). Similarly, Ben-Shakhar and Elaad (2002) found that significant detection can be found even with a single question, although there is an high probability of false positive effect, compared to multiple questions. Therefore, the ideal test should contain a minimum of three questions that need to be repeated (Meijer, et al., 2011).

Once the test has been created, the question set needs to be inserted in the *LXSoftware.* An example of a polygraph test structure on the LXSoftware can be seen in Figure 4. The example reported refers to the questionnaire used for the Study II. Irrelevant questions always open the series of questions. The green question represents the Buffer question (called control/comparison by the software); the red questions represent the Control options (which include also the relevant question). For setting reasons, the software considers the control questions as relevant; however, this difference did not affect the results of the study.

This software allows to manage the question set, creating, editing or eliminating the questions. In addition, it is possible to specify the question type in the question editor, specifying if they are relevant, non-relevant or control questions. This procedure facilitates the scoring process, which occurs at the end of the examination.

H&S experiment.lxq 16/02/2018 Page 1 of 2

#	ID	Type	Question Text	Respo...
Question List: 1 - H&S experiment			Other	
1	1	Irrelevant / Neutral	Do you attend the University of Huddersfield?	
2	2	Control / Comparison	Regarding to the balloon you burst, did you burst it with a earring?	
3	3	Relevant	Did you burst it with a fork?	
4	4	Relevant	Did you burst it with a pen?	
5	5	Relevant	Did you burst it with a scissor?	
6	6	Relevant	Did you burst it with a paperclip?	
7	7	Relevant	Did you burst it with a pin?	
8	8	Relevant	Did you burst it with a screwdriver?	
9	1	Irrelevant / Neutral	Is it sunny today?	
10	2	Control / Comparison	Regarding to the egg you smashed, did you smash it with a a shoe?	
11	4	Relevant	Did you smash it with a hammer?	
12	5	Relevant	Did you smash it with a book?	
13	5	Relevant	Did you smash it with a wooden rool?	
14	6	Relevant	Did you smash it with a hand?	
15	7	Relevant	Did you smash it with a plate?	
16	8	Relevant	Did you smash it with a bottle of water?	
17	1	Irrelevant / Neutral	Did you go to University today?	
18	2	Control / Comparison	Regarding to the balloon you burst, did you burst it with a key?	
19	8	Relevant	Did you burst it with a screwdriver?	
20	6	Relevant	Did you burst it with a paperclip?	
21	4	Relevant	Did you burst it with a pen?	
22	7	Relevant	Did you burst it with a pin?	
23	3	Relevant	Did you burst it with a fork?	
24	5	Relevant	Did you burst it with a scissor?	

Figure 4. An example of a polygraph test structure on the LXSoftware.

When the examiner starts to read the question, an "Onset" button needs to be pressed and held down until the question is asked. A grey bar on the chart appears, marking the time it took to read the question. Once the examinee has answered, it is possible to annotate the subject's response by clicking the "yes" or the "No" buttons. This also makes easier the reading of the chart when it is necessary to evaluate the length of the physiological reaction after the question.

Once the examinee has answered, a countdown of 15 seconds, depending on the settings in place, appears, during which it is not possible to ask any other question. This mechanism has been put in place to allow examinee's physiological levels to come back to a base line level after the last question response.

3.4 Tasks Development

Prior to the CIT a priming stage was administrated to participant, consisting of the completion of tasks. Although every experiment of this thesis includes different tasks, they retain commonality in relation to creating the psychological context in which participants need to lie.

The theoretical base behind this stage is focused on the concept of "priming": it enables the exposure to an external stimulation, which might affect a subsequent response. The result of this association is created by the unconscious activation of mental constructs. Priming can be perceptual, semantic or conceptual (Weingarten et al., 2016). In this study, the tasks are developed based on a conceptual priming: manipulations through the completion of a task are used to activate the internal mental representation of elements acquired during the task process (in this case, relevant information about a specific context). During the priming stage, the participant does not realize the association between the activation context (the task used as primer) and the later influence or use of that representation for the examination; in fact, the importance of this stage is the activation of the representation essential for the examination, rather than the task itself (Bargh & Chartrand, 2000). Although this priming process is necessary for creating a context in which the representation can be activated, the mechanism can be successful only if intentional. In fact, the interest of this study is focused on how the intention to lie can actually affect the physiological reactions of participants when the representation created during the priming stage is activated. Therefore, the effectiveness of this representation, which is stored in the participant memory, plays a determinant role. In addition, this method provides standardisation and validity to the study, since every participant is subjected to the same priming stage. As a consequence, they will try to lie about the same elements without being affected by personal experiences.

For all these reasons, it was necessary that participants were aware of the actual stimulation. This type of "conscious" priming is called *supraliminal*: during this process, the participant is stimulated but fully aware of the specificity of the priming stimulation itself. However, he does not need to be aware of the underlying pattern which serves to prime the construct (Bargh & Chartrand, 2000), such as the characteristics of the relevant object, or the basic mechanism of the CIT. The awareness of the potential effect of the priming makes a difference in the process. It has been found that if the primes are very explicit in representing a specific category (e.g. Gandhi and Mandela as primes for "good"), they are clearly too memorable and likely to be used as a conscious standard of comparison (Herr, Sherman, & Fazio, 1984). In other words, the awareness of the participants can negatively affect the results. However, it has been demonstrated that when the individual is aware of the priming material, it produces stronger priming effects than the subliminal priming. This can be explained because the activation of a concept by a conscious and intentional process is stronger; therefore, its accessibility and likelihood of subsequent use will increase (Higgins, King, & Mavin, 1982).

During the priming stage, the research assistant challenged the participant to lie about a specific aspect of the task. This represents another important element of the priming stage, focused on the motivation of the participant (considering the evident limitation of the absence of real stressor-emotional factors in a lab context). Studies researched the role of motivations, and it has been found that stimuli related to valued actions might elicit greater goal activation than those associated with lesser value (Förster, Liberman, & Friedman, 2007). Therefore, by presenting the task as important in the process of beating the polygraph, emphasizing that at the end of the examination the participant will be able to know if the polygraph machine detects him as liar, this methodology will try to increase the effect of the stimulation. In conclusion, the development of the tasks for each represented a critical and important stage.

The first two experiments of this thesis consisted in the application of two action tasks. In contrast, the third and the forth studies consisted of planning a mock crime; the third study invited participants to plan the mock crime individually, while the fourth took place in a group setting.

3.4.1 Action Task

Subject-Performed Tasks (SPTs) are those tasks in which participants are given verbal instruction that invite them to symbolically perform a series of mini-tasks, such as cutting the bread, painting a table or smashing an object (Engelkamp & Zimmer, 1999). These tasks can be performed with real or imaginary objects, with the purpose to memorize the actions. The efficacy of these tasks was analysed through the application of techniques such as free recall or recognition of the verbal instruction. The consistent result obtained from these studies found that memory in SPTs was better than that under other experimental conditions that involved just verbal tasks, in which participants only had to memorize action phrases (Engelkamp, 1998). As a result, a self-performed task effect was investigated, which describes that verb phrases are memorized better when these are physically performed (therefore learned) than those that involve the gathering of verbal information (Cohen, 1981). According to some authors, the use of physical movements and gesture improves the quantity of phrases that can be recalled, improving the length of time in which they can be stored, making also easier the access to this information (Engelkamp & Krumnacker, 1980).

Different studies (Engelkamp & Zimmer, 1985; Engelkamp & Zimmer, 1997; Engelkamp, Zimmer, Mohr, & Sellen, 1994) highlighted how enacted actions are well recollected because of a combination of the verbal-semantic, perceptual, and motor output systems, which are activated during and enactment process. More specifically, motoric output information appeared to be the responsible factor for the enactment advantage. In other words, the motor encoding enhances recall over the visual feedbacks typically available during the encoding process. Therefore, the SPTs effect is the product of different forms of information: visual and motoric, respectively. On the other hand, it needs to be considered that human memory is very flexible and subjective factors need to be included in the analysis. However, more recent studies have shown that the type of SPT can make visual information quite salient, producing greater attention during the process of encoding. In fact, it was suggested that the use of real objects during enactment might improve the process of encoding information, supported by both visual and motoric information (Mulligan & Hornstein, 2003).

For these reasons, two action tasks were developed for each of the first two experiments: bursting a balloon and attacking a mannequin (Study I); bursting a balloon and smashing an egg (Study II).

For the last two studies (Study III and IV), participants were asked to plan a terrorist attack. One of the objectives of this project is, indeed, to investigate if a crime context might have a stronger impact on participants when trying to lie, due to a possible higher psychological and emotional involvement. In addition, Study III and IV have been designed to evaluate potential and limitations of the use of the CIT in identifying criminal intentions.

3.4.2 Visual presentation

Although different authors (Allen, et al., 1992; Farwell, 1991; Seymour, et al., 2000; Ben-Shakhar, 2003) have reported successful results with the CIT, they have largely ignored non-verbal stimuli, despite the prevalence of photographic evidence in forensic cases. Visual stimuli has been used only with a combination of CIT and pupil dilation, using names and photographs as stimuli (Lubow & Fein, 1996). Findings suggested that visual and verbal stimuli contributed equally to detection. Although, in two similar studies using the CIT, results reported differences in electrodermal response. In Ben-Shakhar's study (1987), participants needed to study either schematic faces or verbal description that varied along several dimensions. Later, participants were tested using the previously studied stimulus. It has been found that verbal descriptions lead to a greater mean electrodermal response than schematic faces. Similar results were reported in Gati and Ben-Shakhar (1990), where verbal stimulation produced higher reactions than visual stimulation. Conversely, Seymour and Kerlin (2007) reported no significant differences in participant's responses, when comparing the differences between verbal and visual stimuli effect. The conflicting findings between these studies confirm the lack of theoretical support around this area, which is further investigated in the current study.

In CIT examinations, it is possible to use visual stimulation through slide projectors or other means to present the images (Krapohl et al., 2006). Although, the visual setting that is proposed in this study has not been previously used. Therefore, this represents one novel aspect of this thesis. However, additional care should be used with this additional modality. In fact, Krapohl et. al (2006) recommended to dedicate specific attention to the selection of visual stimuli, as pictures shown can contain more information or distractions than words. Examiners should standardize the illumination level of the images, their size and coloration. Every image should have the same emotional weight and be visually plausible as the relevant item.

Three of the studies within this thesis were conducted with the support of visual stimulation. While the questions were verbally asked, a Microsoft PowerPoint presentation displayed the images related to each item in the question through a projector on a board size screen (Figure 5). Every image has been shown starting with the formulation of the question until the participant had expressed its answer. After the answer, the picture was followed by a blank slide, which lasted for 15 seconds; this corresponded to cool-off period time, in which the physiological levels needed to come back to their baseline after the stimulation. This will also allow to keep recording conditions clean and structured (Krapohl et al., 2006; iMotions, 2017). The researcher ensured that the illumination levels of the images were appropriate and that the participant was looking at the pictures during the examination and not elsewhere (Krapohl et al., 2006).

Figure 5. Example of a visual stimulation during a polygraph test.

3.5 Procedure

3.5.1 Priming stage

The experimental stage included all the participants, excluding the control groups. Although in the four studies the priming stage differs in the modality, the procedure was always the same. According to Engelkamp (1994), who suggested that memories produced after performing a task are stronger than verbal tasks, different "action tasks" were developed and standardized for everyone. In the first two studies, participants were physically involved in solving simple action-tasks. While in the last two, they were asked to accurately plan the details of a mock crime.

- Study I: participants were asked to burst a balloon and to attack a mannequin. The item used for the first task was a fork, while for the second task it was a knife. These two represented the two relevant items for the Study I.
- Study II: participants were asked to burst a balloon and to smash an egg. The item used for the first task was a paper clip, while for the second task it was a hammer. These two represented the two relevant items for the Study II.
- Study III: participants were asked to plan an attack. The items chose for the discussion were Paris as city, the Eiffel Tower as location and Guns as weapon. These three represented the three relevant items for the Study III.
- Study IV: participants were asked to plan an attack in groups of four. The items chosen for the discussion were Paris as city, the Eiffel Tower as location and Guns as weapon. These three represented the three relevant items for the Study IV.

Two researchers ran the experiments: the assistant researcher, which had the role to prime participants, and the polygraph examiner, who was responsible for the use and administration of the CIT. The assistant researcher contributed to the organisation of the study's stages, booked laboratory rooms and aid in the recruitment of participants.

The priming stage occurred in a normal classroom of the university. Once participant arrived, the assistant researcher gave them the time to read and sign the consent forms. They were also asked to point out any heart-pressure problems or any other type of condition that might compromise or impact on the results. Moreover, participants were informed of the experiment's aim, clarifying any possible doubts or concerns about the polygraph examination.

After this brief introduction, participants were asked to complete the tasks corresponding the study introduced above. During the priming stage, the polygraph examiner was not present in the room: a blind researcher setting was necessary in order to ensure validity during the scoring of the results. In fact, Rosenthal and Fode (1996a) stated that expectations and biases of an examiner could be unintentionally transferred to examinees. Therefore, these cues can significantly affect the results, especially in CIT examinations. In addition, in order to prevent these effects, Meijer et al. (2009) suggested that ideally, the examiner should be unaware of the relevant items chose for the tasks. In the present study, all these suggestions were applied.

3.5.2 Examination

The priming stage was always followed up by the administration of the polygraph. As previously mentioned, the process of information storage is complex and essential for an effective CIT. In fact, Meijer et al. (2009), suggested that this process can be hindered by the interference of other information, which usually increases with time. Thus, it is suggested to administrate the CIT as soon as possible after the event.

The examination stage involved experimental and control participants. Although, participants from the control group did not received any information regarding the tasks. The psychology lab was a quiet and discrete room with a desk and the polygraph chair, which was part of the standard polygraph equipment. The neutral and silent characteristics of the psychology lab created the perfect experimental environment, to avoid any distractions or additional stimuli.

Before the beginning of the examination, the examiner briefly introduced themselves, the study and the procedure of the CIT. This pre-test stage is essential in order to acclimate the examinee to the instrumentation, the examiner's voice and to be sure that the examinee understands all the instructions (Iacono, 1991). In addition, it was shown that building a standard level of rapport prior to the examination could decrease stress levels of the interview (Lynch, 2009; Osugi, 2011). During polygraph examination stress could affect the physiological responses of the examinee and impact on the results (Carson, 2009).

The examiner attached the participants to the polygraph equipment, providing information about the type of physiological information that would have been recorded. EDA, breath rate, heart beat and finger pulse were attached and recorded. In accordance to Ben-Shakhar and Elaad (Grubin et al., 2004), participants were instructed also to respond "NO" to each question. In fact, they suggested that verbally answering with a "NO" to all the questions increases the detection efficiency of laboratory examinations. In addition, the examiner was instructed to empathize the fact that he was not aware of the task details. Because the tasks were simple and not psychologically demanding, it was attempted to increase the participants' motivation, soliciting them to lie in order to beat the polygraph machine.

The examiner was always placed on the front left of the participant. This position was considered optimal because it does not limit the visibility of the participant (Iacono, 1991). In addition, considering the complex threat of countermeasures that are present today, it is advised to place the examiner in a position in which it is possible to get all the dynamics possible during the examination (Iacono, 1991). Once the equipment is set up and participants felt comfortable with the testing procedure, the CIT was administrated. The test lasted around 20 minutes. It was important that while reading the questions, the examiner used the same voice inflection, in order to avoid that participants might have accidentally identified the relevant item (Iacono, 1991).

Finally, at the end of the experiment, a debrief sheet (e.g. Appendix D and E) was given to participants, in which all the information and aim of the study were explained. In addition, if participants were interested, the examiner could provide them with more information regarding the polygraph mechanism or possible feedback of their performance during the test.

3.6 Data acquisition

Two types of analysis were conducted for the purpose of this study. Initially, preliminary analysis was conducted only on the SCR using the standard method of acquisition described in the LXSoftware. While the second round of analysis was conducted on all three physiological measures, SCR, RLL and Cardio, by using intervals of time for extracting the data.

3.6.1 Preliminary Analysis

In order to proceed with the scoring and the analysis of the data collected, it was necessary to extract EDA responses from participants. EDA was recorded using a constant voltage system (4 μA constant current, range of 10 kΩ to 2.3 MΩ) using two finger clips (0.1 second time constant, 635 nm.). The LXsoftware provides an on-screen caliper which is able to measure the magnitude and duration of participants' physiological reactions during examinations. At the beginning of each participant's chart, the calipers were set up: by selecting the area between the horizontal and vertical calipsers, it was possible to set a fix caliper that could be moved as a single unit (Instrument, 2006). With this procedure, it was possible to drag the caliper along the chart, without the need to resize at each question. Therefore, it was possible to ensure standardisation for the single measurement and for the whole sample. Once a specific area is selected by the calipers, a “caliper statistics” section is displayed on the screen. This indicates the dimensions of the chart area selected. These bars measure minimum, maximum and average of EDA, heart rate, respiration rate, cuff pressure and cardio baseline characteristics. For the purpose of this preliminary analysis, it was measured only the maximal EDA (see section 3.6). As mentioned before, the Irrelevant question was not considered in the measurement, as it serves only as introduction to each series and it does not contain any relevant information. The second question, which consists of the Buffer question, was also excluded from the measurement, because examinees tend to physiologically react to the first question that are relevant to the event (Krapohl, et al., 2006). When the calipers were applied to the area of interest, a number was produced in the “caliper statistics” bar dedicated to EDA responses. These numbers represented the length of the EDA of the questions highlighted. Each EDA number-related question was reported in an excel sheet.

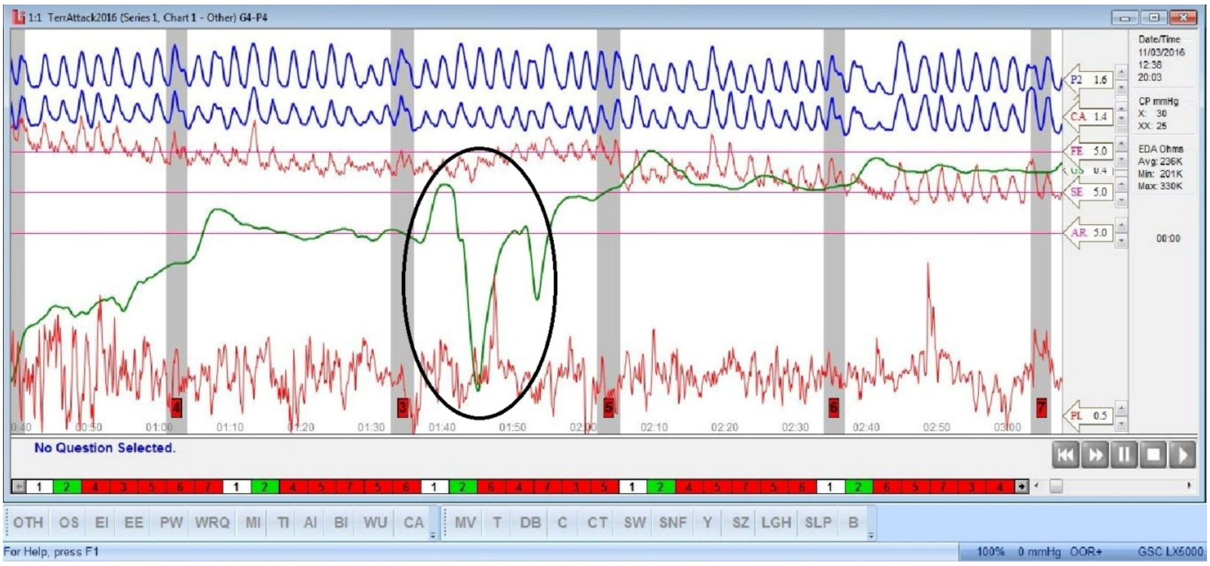

Figure 6. *Example of a screenshot from a polygraph examination record.*

The blue lines at the top indicate the breath rate of the examinee; the following red line indicates the heart; the green line indicates the electro-dermal activity while the bottom red line indicates rate the changes in the blood pressure. The vertical grey lines indicate the precise moment in which the questions were asked. The screenshot shows a complete series of questions (red numbers included in the grey lines); the second question refers to the relevant one. In fact, it is possible to see a typical raise of the electro-dermal activity (highlighted by the black circle) just after the question was formulated. This change is also supported by changes in the blood pressure.

3.6.2 Follow up analysis on SCR, RLL and Cardio measurements

A follow up analysis was conducted after the preliminary results, aiming to use standardised scores of intervals of time for each of the measures available during the interviews (similar to e.g. Selle, Kindt, Vershuere Meijer & Ben-Shakhar, 2016). Five types of data were recorded during the interviews: EDA, thoracic pneumatic measure, abdominal pneumatic measure, heart beat and finger pulse (see section 3.2). However, due to technical reasons, finger pulse recording was compromised during the interviews, caused by a malfunction of the equipment. Therefore, only heart rate could provide information regarding the cardiovascular responses of participants.

Each interview was extracted as text files from the Lafayette machine, with the option "*download polygraph file*", and imported into the statistical computing language environment *R* (R Core Team, 2017). The text file was then opened in R in order to prepare the data for the scoring analysis. In order to reduce the signal noise, data filtering has been applied, which is considered one of the most common strategies adopted to analyse physiological polygraph data. In fact, all virtual data requires signal processing in order to optimise the data available (Nelson, 2018). This process aims to remove artefacts from the physiological data due to external factors, e.g. movement of the participant (Kircher & Raskin, 1988; National Research Council, 2003; Nelson, 2018; Braithwaite, Watson, Jones, Rowe, 2015).

However, the algorithms adopted to filter and smooth the data are not standard and can vary based on the particular signal processed, the software used and the scoring strategy adopted for the final analysis (Nelson, 2018).

For the above reasons, in the current study, the filtering and smoothing techniques have been applied only to limit the values to physiological boundaries.

The strategy adopted for all the 3 channels can be summarized in 5 steps:

1- High-pass and low-pass filtering to remove the values exceeding the physiological limits chosen;
2- Signal smoothing to remove the minor peaks in the data;
3- Identification of the peaks in the signal;
4- Evaluation of the mean rate at each peak;
5- Calculation of the average value in the interval.

The filtering and smoothing operations on the signals have been performed in the frequency domain.

EDA was recorded using a constant voltage system (4 µA constant current, range of 10 kΩ to 2.3 MΩ) using two finger clips (0.1 second time constant, 635 nm.). As explained in section 3.2.3, EDA consists of two main components: SCL and SCR. The interpretation of polygraph data involves only the use of phasic EDA responses, as SCL is considered not to be related to arousal of high-amplitude "spikes" (Nelson, 2018; iMotions, 2017). In addition, some examinees have unstable tonic EDA, caused by a variety of physiological and environmental factors. This instability can substantially increase the difficulty of extracting and interpreting phasic EDA responses. For this reason, field and laboratory polygraph systems have been used EDA filtering options for several decades (Nelson, 2018). In order to remove the tonic component of the EDA signal from the data of this study, a median filter was applied using R. The median EDA score based on a +/-4 s time interval centred on each current sample using the function *medianFilter*. Then, the average was subtracted from the sample, leaving only the phasic data (SCR). Subsequently, a high-pass Butterworth filter (Butterworth, 1930) has been applied, using the *butter* function from the *signal* package in R, with a cut-off of frequency 0.1Hz. Then a low-pass filter in the form of a moving average period of 0.5 seconds has been used, corresponding to a corner frequency of 0.883Hz, along with a smoothing filter at 0.443Hz (Raskin, Honts and Kircher, 2014).
Subsequently, the data from SCR was processed by calculating the difference between the maximum and the minimum value in the interval between 1s after the onset of the trial and 5s after the onset (Selle, Kindt, Vershuere Meijer & Ben-Shakhar, 2016; iMotions, 2017). This allowed to calculate the amplitude of the EDA, used for scoring the analysis.

Respiration was recorded using two pneumatic tubes placed in the thoracic and abdominal area (See section 3.2). Respiration responses were defined on the basis of the total Respiration Line Length (RLL). RLL is a composite measure of respiratory amplitude (depth of breathing) and respiratory cycle (rate of breathing). In particular, for the two channels associated with breathing, namely *UPneumo* (Thoracic) and *LPneumo* (Abdominal) the signal has been limited between 12 and 40 breaths per minute. These values conform to the limits of the natural respiratory rate of 12 to 20-25 breaths per minute (Mead, 1960), with more provision left for higher breathing rates.
The segment of data has been first filtered with a high-pass Butterworth filter with a cut-off frequency of 0.2Hz (12 breaths per minute). The first 2s of data from the resulting segment were discarded and the data was filtered with a low-pass filter with a cut-off frequency of 0.667 Hz, corresponding to 40 breaths per minute.

The length of the RLL was measured starting from stimulus onset up to 13s later (Ben-Shakhar & Dolev, 1996). Timm (1982) has highlighted how the length of the RLL might be disproportionally affected by the start of measurement. In order to solve this issue, each RLL should be calculated as the mean of 10 RLL's: from 0.1s stimulus onset to 13.1s later, from 0.2 s after stimulus onset to 13.2 s after stimulus onset, etc. (Vandenbosch, Verschuere, Crombez & De Clercq, 2009; Selle, Verschuere, Kindt, Meijer & Ben-Shakhar, 2016).
For this study, UPneumo (thoracic) and LPneumo (abdominal) data were initially taken in the interval from 4s before stimulus onset to 13s after stimulus onset. A filtering procedure similar to the one used for the EDA channel has been applied, with different cut-off frequencies, as reported in table 1. After the filtering, the first 4s of data has been discarded, to avoid filtering artefacts at the beginning of the interval. Therefore, 10 windows were created in the 13s interval, each beginning 0.1s after the previous one.
The data was subjected to a second filtering step resulting in smoothing with a moving window of 1.03s, using the *filter* function from the *stats* package with the *method* parameter set to "convolution". The process of smoothing of the signal aims to eliminate the minor peaks.
Finally, the average breathing rate at each peak starting from the second one was calculated by dividing the number of samples in one minute by the number of samples between the current and the preceding peak. Any values outside 8 and 40 breaths per minute has been discarded as too extreme (Selle, Verschuere, Kindt, Meijer & Ben-Shakhar, 2016). The mean of the remaining values, recorded as UPneumorate and LPneumorate in an Excel file, has produced the RLL used for the scoring analysis.

The heart rate (HR) data was processed in a similar way to the RLL, using the parameters reported in table 1. The length of the interval post stimulus was 15s (Selle, Verschuere, Kindt, Meijer & Ben-Shakhar, 2016). The number of ignored samples at the boundaries of the interval was 4 on each side. Only rates between 50 bpm and 180 bpm were used in the calculation of the average heart rate. The post-stimulus difference scores were computed by subtracting the HR in the last second prior to item onset (pre-stimulus baseline) from the HR-score of each post-stimulus second. Finally, the largest deceleration and the mean HR within the post-stimulus period were calculated (Gamer, Vershuere, et all., 2008; Selle, Verschuere, Kindt, Meijer & Ben-Shakhar, 2016).

Table 2. Parameters used for the filtering and smoothing processes of this study

	Filtering				Smoothing	
	High Pass filter		Low Pass filter		Smoothing	
	Type	Cut-Off/frequency	Type	Cut-Off/frequency	Type	Window size
EDA	Butterworth	0.1Hz	Moving average	0.883 Hz	/	/
RLL	Butterworth	0.2Hz	Generic	0.667 Hz	convolution	1.03 sec
Cardio	Butterworth	0.2Hz	Generic	0.667 Hz	convolution	0.23 sec

3.7 Scoring

Preliminary analysis

For this procedure, the same blind examiner run the scoring with the assistance of a blind assistant researcher. As every interview process requires a standardised scoring system (Patterson & Hennessy, 2013), the Lykken Scoring method has been applied in the current study (Lykken, 1998). This procedure is one of the scoring systems used for scoring CIT data to facilitate the analysis of the data. Specifically, this method assigns a score of 2 if the largest response was provoked by the relevant item. A score of 1 is awarded if the second largest response was provoked by the same relevant item and a score of 0 if the highest reaction occurred on a non-relevant item. "With *n* questions, the Lykken-score varies between 0 and 2*n*, and the threshold for a guilty outcome is typically set at *n*, meaning that a score of at least *n* indicates recognition" (Meijer at al., 2011, p.299). In other worlds, these values are then summed up across the set of questions to produce a single detection score for each participant on the physiological measure used (SCR for this first part of the analysis) (Iacono, 1991). This method is considered extremely practical, which can be used without quantification and parameter estimations. In addition, as responses are ranked within each block, corrections are not required even if physiological levels change between blocks as a result of habituation (Matsuda, Nittono & Allen, 2012). For the preliminary analysis, this approach was directly applied on the raw SCRs extracted from the data.

Follow up analysis

For the follow up analysis, the method of Z-score averaging was applied, which consists of a widely used procedure to capture quantitative differences between items (Ben-Shakhar, 1985; Braithwaite, Watson, Jones & Rowe, 2015). In this method, a response to each item is standardised using the mean and the standard deviation of each measure. The aim of this approach is to cancel the differences in physiological levels and to treat multiple measures that have different units in the same dimension (Matsuda, Nittono & Allen, 2012; Braithwaite, Watson, Jones, Rowe, 2015). The idea of using standardised scores is based on the principle that standardization considers all the raw data of each subject, rather than just the maximal response. In fact, mean and standard deviation seem to be more stable estimates than the maximum (Ben-Shakhar, 1985). As for the Lykken method, this approach needs no parameter estimation a priori, therefore it is easy to use with the CIT (Matsuda, Nittono & Allen, 2012).

Standardization. In order to eliminate individual differences in responsivity and to allow meaningful index of the response differences between relevant and irrelevant items for the whole test, standard differences scores for each subject and measure (SCR, RLL, Cardio) were calculated accordingly to the following procedure.

Firstly, the data available for each study have been imported in MATLAB to evaluate the z-scores values. A MATLAB script has been created for each parameter used in the study (SCR, RLL, Cardio). The Excel spreadsheet, which contained all the data processed in R, as discussed in the previous section, has been imported using the function *xlsread*. Each column of the Excel file has been stored as a separate variable, ordered per participant and ordered questions.

The indices corresponding to a specific question, e.g. the relevant question, have been firstly identified using the *function find*, then furtherly separated between Control and Experimental group. A vector with the corresponding values of the parameter to study, e.g. Cardiorate, has been extracted for each question, obtaining a vector with the values of the aimed parameter for all the participants to the study. For each vector, the corresponding z-scores have been calculated using the *function zscore*, based on the mean and standard deviation.

Afterwards, the Lykken method (1959) has been applied on the relevant questions: a value of 2 was assigned to the maximum z-score (smallest z-score in the case of RLL and Cardio). The second maximum value has been replaced with 1 and the remaining with zeros for each physiological measure used (SCR, RLL, Cardio) (Lykken, 1959). A *for* loop has been used to automate the identification and the replacement of the values for each line of the matrix. After repeating the same process for all the group of questions, the final matrix has been exported as a text file. Once the scoring stage was completed, all the data was transferred to SPSS for statistical analyses.

Finally, difference scores between the responses to the relevant items and the mean of the four irrelevant items within each of the six multiple-choice questions were calculated using the method described by Gamer, Rill, Vossel and Godert, (2006). Subsequently, for each group of questions (relevant and irrelevant questions) a matrix containing all the z-score vectors has been created. By doing so, each line of the matrix contained the z-scores of the specific group of questions, with the first value corresponding to the relevant question, while the others corresponding to the irrelevant ones. These values have been used for calculating the index of the differential responsivity in each physiological measure.

3.8 Analysis

The study consists of a group of four experiments independently conducted. This study aimed to examine the accuracy rate of the CIT, which is expressed by the percentage of correctly classified cases: true positives (truth teller examinees), true negatives (deceptive examinees), false positives (truth tellers detected as deceptive examinees) and false negatives (deceptive examinees detected as truth tellers).

The present study also run factorial ANOVAs between groups, to measure the variance within and between experimental and control groups. A factorial ANOVA compares means across two or more independent variables. In addition, the factorial ANOVA assumes a cause- effect relationship (Field, 201). This infers that one or more variables (like the priming process or visual stimulation) *cause* the significant difference of one or more characteristics. If this cause-effect occurs, it causes the difference in the mean value of the groups. Preliminary analysis run ANOVA on the data extracted from SCR. Diversely, in the follow-stage, the same type of analysis was run on all the three measures. However, as Rosenfeld (2018) has suggested, when elicited by the relevant stimulus EDA increases, pulse volume produce smaller responses, and respiration is supressed; therefore, each measure has a different direction. For this reason, the ANOVAs were run separately.

In order to get an index of the differential responsivity in each physiological measure, the mean of all the measures used was computed. It was assumed that this aggregation of values should be around zero for innocent subjects, as they are supposed to show a non-systematic response pattern when irrelevant and relevant items are presented. Conversely, a negative value (for RLL and Cardio) or a positive value (for SCR) should indicate knowledge of relevant information (Gamer, Rill, Vossel & Godert, 2006).

3.9 Ethical Clearance

This study received full ethical approval by the University of Huddersfield's School Research Ethics Committee (SREP) and adhered to guidelines in regards to the safe management and storage of personal data. The study needed to adhere to a number of different guidelines, which could ensure the privacy of the participants. All the personal details about the participants that took part in this study have been removed and complete anonymity was guaranteed. For the purposes of this study, information regarding their physical state were asked prior the polygraph examination. This information was used only in case of possibility of exclusion from the study, only the physiological reactions from participants were considered as important information from participants.

CHAPTER 4: STUDY I

4.1 Background

Over the last 80 years, many studies have been conducted in order to determine the validity of the polygraph as a reliable machine able to detect deception. Although the polygraph was applied in many forensic settings, as outlined in Chapter 2, there are controversial issues around the application of the polygraph machine in investigative and legal contexts. Different experiments have tried to estimate which one of the many polygraph tests is the most valid and the outcome of these studies showed that the CIT polygraph test is considered to be the most effective, even though it is not the most utilised test internationally.

Many studies tried to investigate the validity of the polygraph through the application of different stimulation procedures, in order to assess which one was the most effective. According to previous studies (Bradley et al., 1996; Gamer et al., 2008), the support of visual stimulation was found to have a strong impact on physiological reactions. However, almost no research concerning visual stimulation has been conducted on the CIT. In order to investigate the accuracy of the CIT using alternative methods, it is first necessary to conduct an explorative study which will evaluate standard procedure in the CIT. This will enable the current study to gain familiarity with the test and to acquire new insight into it. The exploratory aspect of this study will also take into consideration how relevant information acquired can affect participants' physiological reactions in different ways.

As discussed in Chapter 5, the priming stage involves the stimulation of participants' memory by presenting relevant information, which will be asked about during the examination. For Study I, two action tasks were created: in the first one, participants needed to burst a balloon with a fork, while in the second participants needed to attack a mannequin with a knife. In line with Cohen's (1981) research, individuals are more likely to remember self-performed tasks than just verbal or visual tasks. These results can be explained because performed actions involve richer and more complex representations than simple verbal phrases or passive tasks. In addition, those actions that also involve the motor system are likely to have a greater impact on memory (Engelkamp & Zimmer, 1997).

The rationale behind the development of the Balloon Task is based on previous research which highlighted how the echoic memory can be one of the strongest components of the sensory memory (Clark, 1987). It was found that auditory sensation can be very evocative and is able to elicit highly emotional reactions (Royet et al., 2000). Research suggested how sensory stimuli, in particular auditory stimuli, could be related to emotional information. For example, a specific sound can be strongly related to emotional information, being stored in the auditory cortex, such as the memory of fear (see section 2.3). This shows that the sound can acquire an emotional meaning, which can have an impact on memory (Grosso et al., 2015). In addition, the reaction time for auditory stimuli has been found to be faster than visual stimuli (Jain et al., 2015). This statement surely supports how important the verbal administration of polygraph tests is and the recollection of any echoic memories associated with the information stored. In fact, it was hypothesized that during the administration of the polygraph, the result of the process of revocation of this echoic memory (the fork that burst a balloon) could release a greater physiological reaction. The analysis of the aspects of the relevant information and how they impact physiological reactions represents one of the novel elements of this study. In addition, to the best of the researcher's knowledge, this study is the only one that has examined the CIT with the use of action tasks.

In order to compare the efficacy of this effect, it was necessary to create a different task that could explore other aspects of the relevant information. With this in mind the Mannequin Task was designed, which required participants to attack a mannequin with a knife. In this case, the task focused on the physical involvement of the participant by performing an action with the use of a weapon. The decision to use a weapon instead of a normal object was related to the attempt to determine whether the presence of a weapon could have a greater effect on the physiological response of participants during the polygraph examination. In fact, researchers suggest that when individuals are exposed to weapons, their reaction time is faster than when exposed to neutral objects (such as a simple balloon).

In a similar study, it was found that reaction times toward aggressive words were faster than to neutral words when primed with pictures showing weapons than when primed with neutral photos. Therefore, the presence of weapons can affect thoughts, appraisal and behaviour. This can also be seen in research that shows how people can become more aggressive even when weapons are concealed instead of visible. It is clear that the presence of a weapon can have an impact on memory and on physiological responses. For these reasons, the following experiment used the same concept supported by the other research, although applied to the performance of a task that involved the use of a weapon. As a result, the comparison between the two tasks might enable this exploratory study to determine which performance had a stronger impact on physiological reactions when trying to conceal information.

4.2 Aims of the experiment

In the first stage of this study, the aim is to evaluate the validity of polygraph through verbal stimulation, which represents the standard procedure for all polygraph examinations. This will enable an analysis of the basic mechanisms of the CIT and the impact of relevant information.

This study will examine the deception levels of participants from the experimental group and the possible level of false positive effect in participants from the control group. In addition, the habituation effect will be analysed and its effectiveness evaluated.

Two action tasks were designed to act as a priming function for the experimental stage: a Balloon Task and a Mannequin Task. The role of the tasks is not limited solely to prime participants; they will also be studied to investigate how specific elements involved in the tasks (echoic memory or presence of a weapon) could have an impact on the memory of participants, thus resulting in a greater physiological reaction.

4.2.1 Hypotheses:

- Accuracy rates in the CIT will be above the level of chance
- A higher deception level in the experimental group will be found.
- A low false positive effect will be found.
- Habituation will reduce the accuracy rate of the test during the examination.
- Greater physiological responses during the relevant questions will be found
- A various in the accuracy rate between the questions series will be found

4.3 Participants and Recruitment Process

Sixty-three students (41 women and 22 men) took part in this study. Their mean age was 27.2 years (*SD= 3.79* range 19 to 35). Forty-three of the participants were randomly assigned to an experimental group and 20 to the control group. Participants were randomly approached on the university campus of Huddersfield. Undergraduate and postgraduate students and members of the university staff constituted the majority of the sample. Recruitment was completed by approaching potential participants directly on campus in areas such as in the library, lecture halls, cafés, and several social areas. In addition, a number of posters were placed around the campus, which generated interest in the study.

4.4 Equipment and materials

The physiological data for this experiment was recorded with the Lafayette LX5000 SW Polygraph System. Heart rate (HR), breathing rate and EDA were recorded for this study. The first task required the use of a fork and a balloon. The second task required the use of a mannequin and a knife. The objects used for the tasks (balloons, fork and knife) were purchased at the Poundland shop in Huddersfield city centre. The mannequin was booked in advanced and borrowed from the Department of Fashion and Design at the University of Huddersfield.

4.5 Questionnaire Structure

The questionnaire had 48 questions in total. It contained three series of questions for each task (six series in total). Every series of questions contained eight questions: one irrelevant question, one buffer question, five control questions and one relevant question. The Balloon Series was presented first, while the Mannequin Series was presented last. The presentation of the questions followed the performance order of the tasks that the participants engaged with. Each series of questions was alternated in order to ensure validity (see Appendix E) and in line with the standard polygraph methodological approach.

4.6 Procedure

The sample was distributed into an experimental and a control group. The experimental condition involved the completion of two tasks followed by the administration of the CIT.

The control group did not perform any tasks. Participants were provided with full details of the study, including the implications of participating. Participants were encouraged to sign an informed decision, without any pressure or coercion, and they were notified of their right to withdraw from the study at any point during the examination. Furthermore, they were informed of their right to have their information and data immediately removed and destroyed from the study. In addition, every participant needed to indicate any possible physical conditions that could have affected the results of the test.

Experimental Group:

Once participants signed the consent forms and were provided with all the information regarding the experiment, they were invited to complete two different action tasks. For more details, see Section 3.4.

Every participant from the experimental group had to complete the tasks individually: first the Balloon Task (Task 1), then the Mannequin Task (Task 2). In order to analyse the results at group level, the objects used during the task were chosen by the research team who then instructed every participant from the experimental group.

Task 1: The research assistant had to inflate the balloon and give it ready to the participant. The participant was then instructed to burst the balloon with a fork provided by the research assistant. The participant was free to attempt to burst the balloon as many times as necessary to successfully complete the task.

Task 2: The research assistant invited the participant to attack a mannequin. The assistant researcher carefully provided the knife. For practical reasons, the mannequin was placed in the middle of the room. The participant was free to hit the mannequin in any part of the body he/she preferred. However, the participant was instructed to hit the mannequin only once.

The priming stage took approximately 10 minutes. Once the priming stage was concluded, every participant was instructed by the research assistant not to reveal the object used during the tasks. Rather, they were instructed to deny the use of the real object during the polygraph examination by answering “No” to all the questions. After this information, every participant joined the polygraph examiner in the Psychology Lab for the administration of the polygraph. In order to avoid any judgment bias during the examination and during the analysis of the data, a blind examiner conducted the polygraph examination.

Control Group:

After having signed the consent forms, participants from the control group were instructed to enter the Psychology Lab for the administration of the polygraph. This group of participants was not primed with any tasks.

Second stage

In the second stage of the experiment, participants from both experimental and control conditions were given a polygraph examination. The questions were asked verbally and every series of question was repeated three times.

The test took approximately 20 minutes: five extra minutes were used for presentation and five minutes for debriefing (for more details regarding the general examination process, see section 3.5). The data was stored on the laptop through the Data Recording System. The coding and scoring of the data followed the criteria and guidelines described in Chapter 3 (section 3.7).

4.7 Results

Participants (N=63) from Study II were divided into an experimental group (N=43 *M=1.04, 95% CI [.94,1.16], SD=.888*)) and a control group (N=20 *M=.48, 95% CI [.35, .62], SD=.756*).
Effect sizes (Cohen's d) comparing the experimental and control groups averaged were computed. Cohen's d = (-2.35, -0.07) ⁄ 0.79 = 3.025668, showing a large effect size.

4.7.1 Preliminary analysis conducted on the EDA with the Lykken's approach

4.7.1.1 Descriptive data

Table 3.Overall psychological responses using Lykken's Method

Conditions	***Overall number of responses***	***Highest EDA***	***Second highest EDA***	***Zero and low EDA***

Experimental group **Task Balloon**	189	87	60	42
Experimental group **Task Mannequin**	189	68	70	51
Control group **Task Balloon**	60	11	10	49
Control group **Task Mannequin**	60	10	18	32

Table 4. *Frequencies for the Balloon and Mannequin Task*

Task	***M***	***SD***	***Median***	***Mode***	***Variance***	***Sum***
Ball1	1.33	.808	2	2	.653	57
Ball2	1.05	.899	1	2	.807	45
Ball3	.95	.872	1	0	.760	41
Man1	1.07	.910	1	2	.828	46
Man2	.98	.859	1	0	.738	42
Man3	.93	.910	1	0	.828	41

Table 5. Percentages of detection rate among the series for the Balloon Task and Mannequin Task

Task	*No Detection*	*Detection*
Ball1	21%	79%
Ball2	37%	63%
Ball3	40%	61%
Man1	37%	63%
Man2	37%	63%
Man3	44%	56%

In regards to the Balloon Task, the analysis of frequencies made possible a comparison between the levels of detection in the three series presented to the experimental group. In Series 1, the CIT was able to detect deception in 34 participants (*M=1.33, SD=.808*), with a rate of accuracy of 79%; similarly, in Series 2, 27 participants (*M= 1.05, SD= .899*) were detected with an accuracy rate of 63%. Finally, in the last Series, 26 participants (*M=.95, SD=.872*) with 61% of accuracy rate.

The same analysis of frequencies was conducted on the Mannequin Task, In Series 1, the CIT was able to detect 27 participants as deceptive (*M=1.07, SD=.910*), with an accuracy rate of 63%; in addition, in Series 2, 27 participants (*M= .98, SD= .859*) were detected with the same rate of the Series 1. Finally, in the Series 3, 24 participants were detected as deceptive (*M=.93, SD=.910*) with 56% of accuracy rate.

As for the experimental group, an analysis of the frequencies was conducted on the control group, in order to calculate any false positive levels in the CIT. Table 5 shows the frequencies related to the Balloon Task and the Mannequin Task, while Table 6 shows levels of accuracy in both tasks, which are presented and expressed in percentages.

Table 6. Frequencies *for the Balloon and Mannequin Task for the Control Group*

Task	***M***	***SD***	***Median***	***Mode***	***Variance***	***Sum***
Ball1	.35	.745	0	0	.555	7
Ball2	.30	.470	0	0	.221	6
Ball3	.65	.671	1	0	.450	13
Man1	.45	.826	0	0	.682	9
Man2	.75	.967	0	0	.934	15
Man3	.40	.754	0	0	.568	8

Table 7. *Percentages of detection rate among the series for the Balloon Task and Mannequin Task in the Control Group*

Task	***No Detection***	***Detection***
Ball1	80%	20%
Ball2	70%	30%
Ball3	45%	55%
Man1	75%	25%
Man2	60%	40%
Man3	75%	25%

Regarding the Balloon Task, the analysis of frequencies made possible a comparison between the levels of accuracy rates in the three series presented to the control group. In Series 1, the

CIT was able to detect 16 participants as truth tellers (*M=.35, SD=.745*), which accuracy rate accounted for 80%; in Series 2, 14 participants (*M= .65, SD= .671*) were detected by the CIT as truth tellers with 70% of accuracy rate. Finally, in the last Series, nine participants were detected as truth tellers (*M=.95, SD=.872*).

The same analysis of frequencies was conducted on the Mannequin Task. In Series 1, the CIT was able to detect 15 participants as truth tellers (*M=.45, SD=.826*), which accuracy rate accounted for 75%; similarly, in Series2, 12 participants (*M= .75, SD= .967*) were detected by the CIT as truth tellers with 60% of accuracy rate. Finally, in Series 3, 15 participants (*M=.40, SD=.754*) were correctly detected (25%).

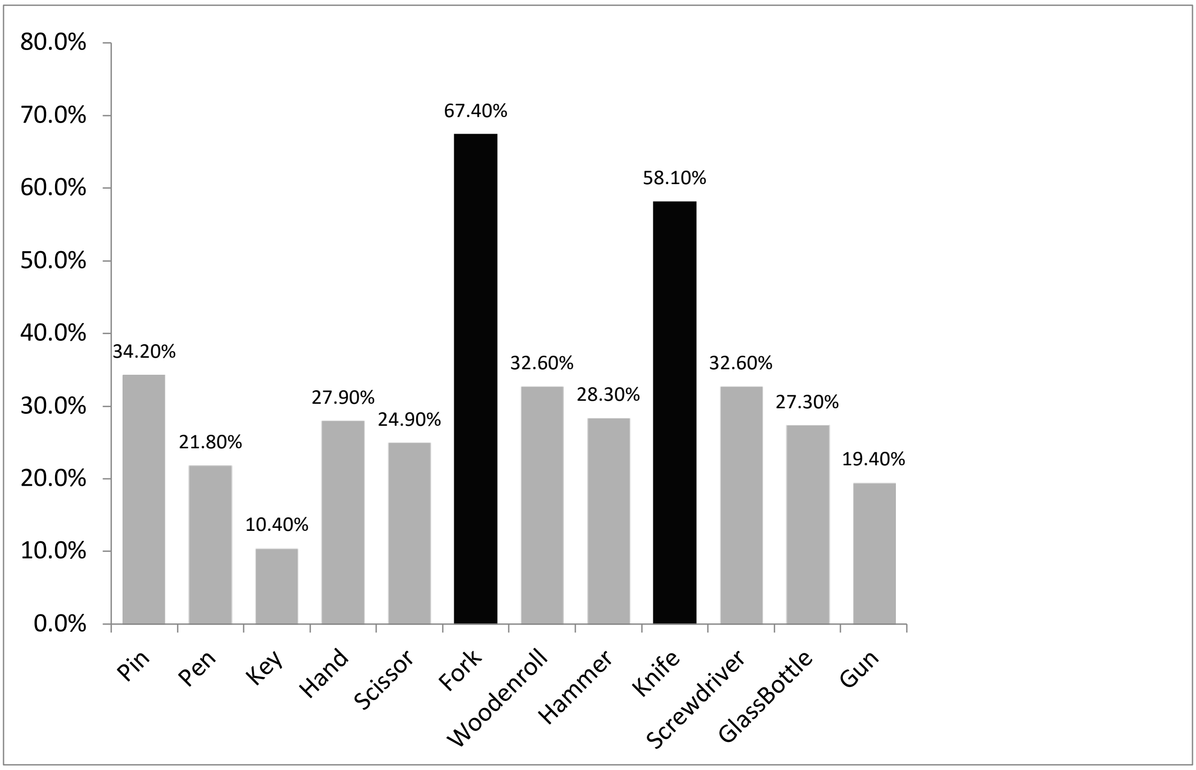

Figure 7. Physiological reactions' levels to all stimuli presented in both tasks from the experimental group

It was also conducted an analysis of responses levels from the experimental group towards every relevant and irrelevant item presented during the CIT was conducted. Figure 7 shows that the highest physiological responses occurred on the relevant items (Fork for the Balloon Task and Knife for the Mannequin Task).

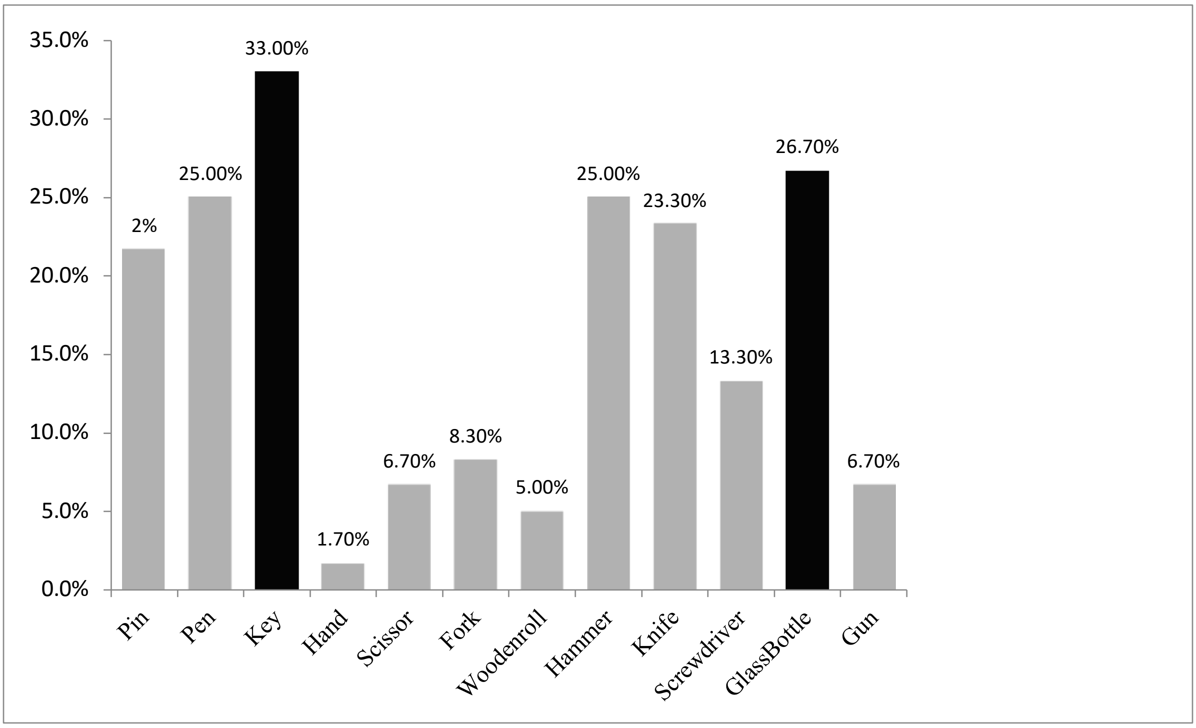

Figure 8. Physiological reactions' levels to all stimuli presented in both tasks from the control group

Similarly, the analysis was conducted on the control group too by measuring the physical response levels of participants towards every relevant and irrelevant item presented during the CIT. Figure 8 shows the highest physiological responses occurred to some irrelevant items (Key for the Balloon Task and Glass bottle for the Mannequin Task).

4.7.1.2 Inferential analysis

In order to conduct appropriate inferential analysis, it was important to determine if the data from this study was normally distributed. In order to obtain this analysis, a single variable was created which included both experimental and control conditions. The analysis contained also a grouping variable, which specified the specific condition of each participant.

As the sample from Study I was a relatively small size, a Shapiro-Wilk test was found to be the most appropriate and powerful compared to the Kolmogorov-Smirnov test for normality (Field, 2013). The analysis suggests that the overall level of deception from experimental and control group, $D{=}545{=}.382, p{=}.000$, was significantly non-normal. When trying to determine the normality of the data graphically from a normal Q-Q Plot, however, it is possible to see that the data distribution is almost normal distributed. However, even if the data is not normally distributed (possibly due to the difference between the control and the experimental group) an ANOVAs is the most appropriate test to employ for this sample as it is considered to be a robust test with deviations from normality (Maxwell & Delaney, 2004).

The assumption of homogeneity of variances was violated, as assessed by Levene's test for equality of variances, *F(2.966) =.001 p<.05*. For this reason, Welch's t-test was preferred for general conclusions of this analysis, as it is considered more robust with equal variances not assumed and with a sample size that differs significantly (Kohr & Games, 1974). In fact, the two conditions indicated, experimental and control group, differed significantly. This was determined by the Welch's t-test, which indicated a statistically significant difference between groups *F(1.376) = 37.184, p<.001.* Mauchly's test was non-significant for the main effect of the tasks (p=.579); no results were shown for the main effect of the Series as there are only two levels.

A 2 (Group conditions: control and experimental group) x 2 (Task: Balloon and Mannequin Task) x 3 (Series: three series for each task) Mixed Factorial ANOVA was conducted.

Factorial ANOVA revealed a significant main effect of group conditions, *F(1.61) = 146.034, p<.001* η^2*=.24*. In fact, Post-hoc tests with Bonferroni corrections applied for multiple comparison (a=.017) revealed a significant difference between control group (p<.001) and experimental group (p<.001). However, the number of the series were found no significant, *F(1.473) = .669, p=.839.* In addition, this analysis found no significant effect regarding the type of tasks *F(1,36) = .009, p=.925*.

4.7.2 Follow up analysis conducted on SCR, RLL and Cardio Using Z-Scores

4.7.2.1 Descriptive data

The first analysis consisted in the classification obtained through the Lykken's method (see table 7), which shows that SCR was the measure that accurately detected more participants, compared to the RLL and Cardio. Also, from this method, it was possible to obtain deception rates for each measure, where the highest was obtained by the SCR, the second highest by the RLL and the lowest by the Cardio.

Table 8. Detection rates computed for the three physiological measures and across measures in Study I for the experimental group using the Lykken's method

Measures	n. of detection (experimental)	n. of detection (control)	Percentage detected (experimental)
RLL	46	15	57%
SCR	52	18	61%
Cardio	34	12	53%

In a first step, the mean z-standardised response differences between relevant and irrelevant items were compared between the two experimental groups. Afterwards the mean of these measures was computed as an overall index of the differential responsivity in each physiological measure, expecting negative values for RLL and Cardio and a positive value for SCR as indication of knowledge of relevant information. As expected, the typical pattern of differential responses to relevant and irrelevant items was found for all the measures in the group of guilty participants, whereas these differences tended to zero for innocents (see table 8).

Table 9. Means and standard deviation of the standardised physiological response differences between relevant and irrelevant items within each experimental condition

Measure	Guilty (43)		Innocent (20)	
	M	SD	M	SD
RLL	-1.41	0.85	0.02	0.47
SCR	2.85	0.77	0.03	0.63
Cardio	- 0.5	0.03	- 0.07	0.18

The reaction levels of participants from the experimental group towards every relevant and irrelevant item presented during the CIT was examined. Figure 9 shows that the highest physiological responses from SCR, RLL and Cardio occurred on the relevant items (Fork for the Balloon Task and Knife for the Mannequin Task).

The same analysis was applied to the control group. Figure 10 shows the reaction levels of participants towards every relevant and irrelevant item presented during the CIT was measured. As expected, the highest physiological responses occurred on some irrelevant items (for the RLL it occurred on Screwdriver; for SCR it occurred on Hand; for Cardio it occurred on Scissor)

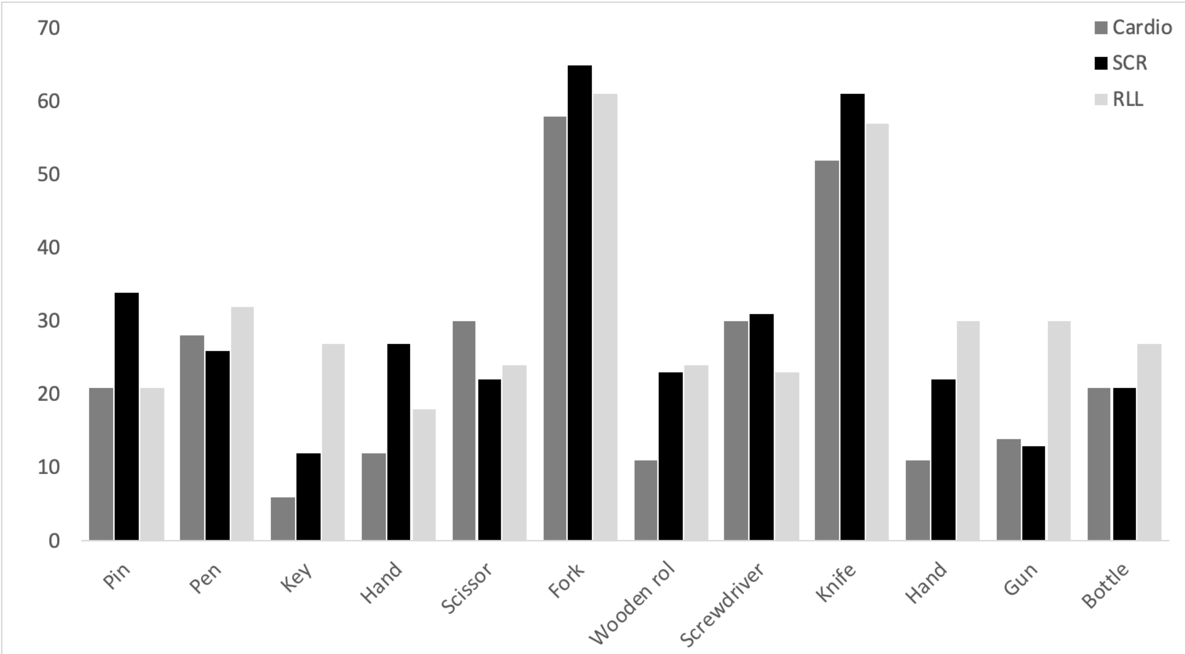

Figure 9. Physiological reactions' levels to all stimuli presented in both tasks from the experimental group

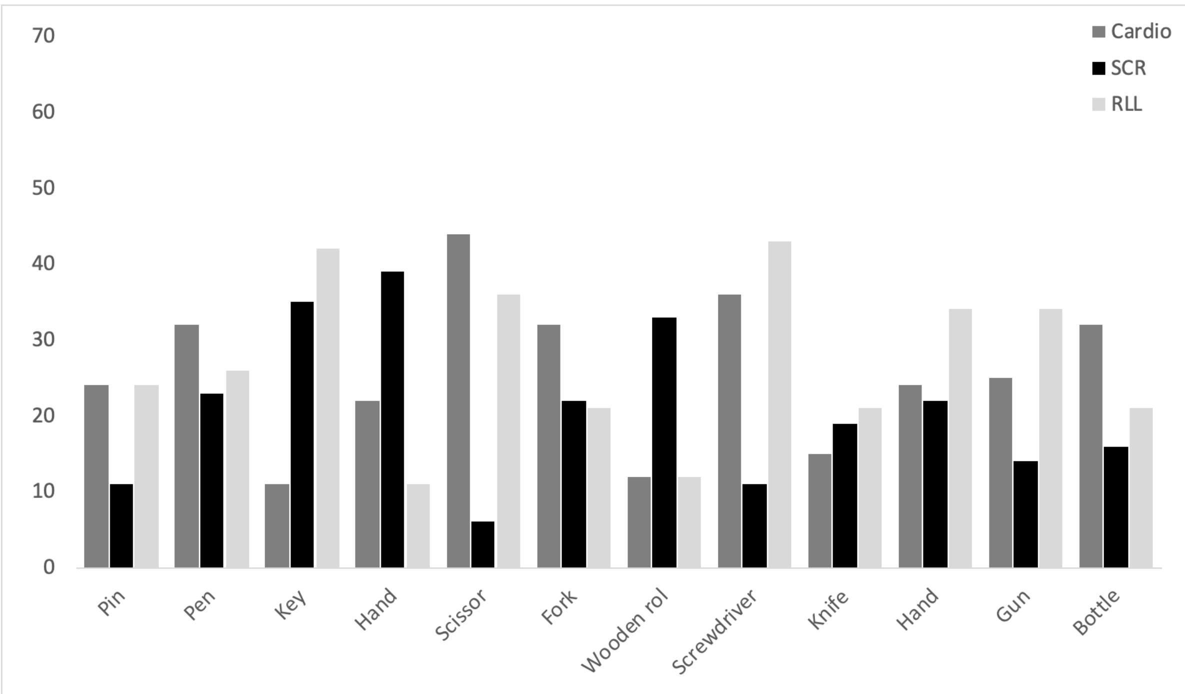

Figure 10. Physiological reactions' levels to all stimuli presented in both tasks from the control group

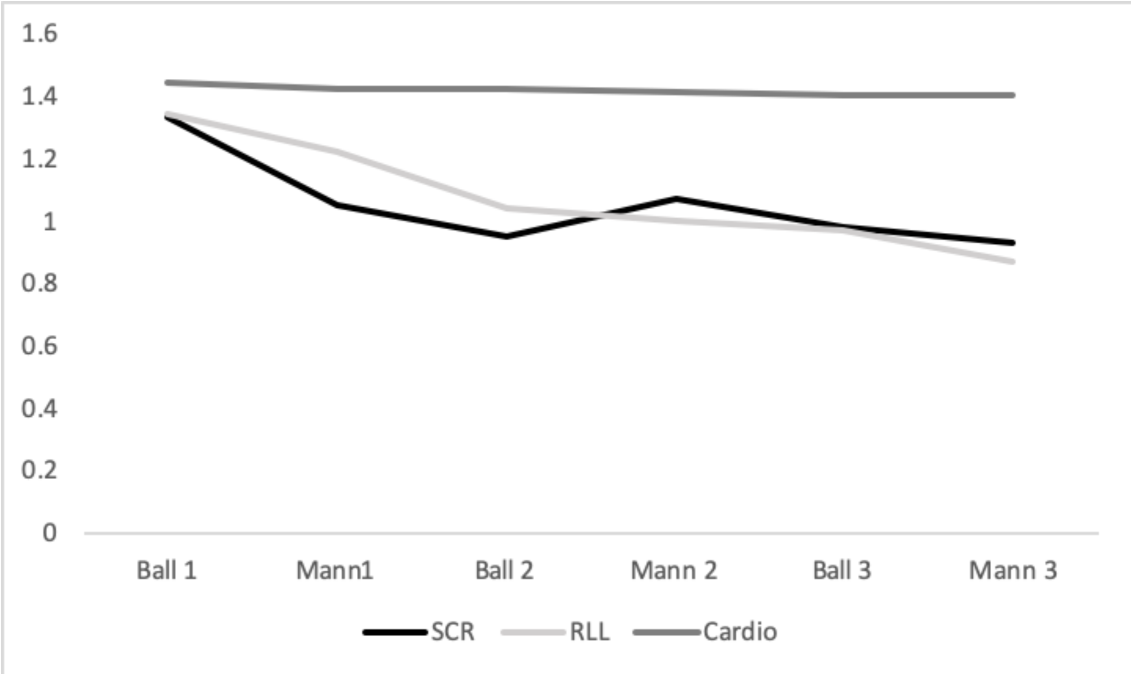

Figure 11. Means of the relevant items presented during the test from SCR, Cardio and RLL

From figure 11 it is possible to see how the means of each measure develop along the duration of the polygraph test. The figure suggests a decrease of the SCR and RLL. The most dramatic decrease is visible on the SCR, which decrease along the duration of the test. A small increase starts from Series 2 of the Balloon task. This result can be confirmed by the preliminary analysis (see table 6), which showed that the highest SCR detection rate can be found in the Balloon task. This suggests that the habituation effect affected SCR. However, as these are only preliminary results, the significance of this effect will need to be confirmed by inferential analysis. Conversely, RLL levels decrease smoothly along the duration of the test, with a very small increase during Series 2 of the Balloon task. Finally, Cardio levels appear stable during the whole duration of the test.

4.7.2.2 Inferential analysis

It is important to determine if the data from this study is normally distributed by running a test for normality. In order to obtain this analysis, a single variable was created, which included both experimental and control conditions from all the physiological measures. The analysis contained a grouping variable, which specified the condition of each participant.
As the sample from Study I was a relatively small size, a Shapiro-Wilk test was found to be the most appropriate and powerful compared to the Kolmogorov-Smirnov test for normality

(Field, 2013). The analysis suggests that the overall level of deception from experimental and control group, D=*323=.382, p=.000*, was significantly non-normal. When trying to determine the normality of the data graphically from a normal Q-Q Plot, however, it is possible to see that the data distribution is almost normal distributed. However, even if the data is not normally distributed (likely due to the difference between the control and the experimental group) an ANOVAs is the most appropriate test to employ for this sample as it is considered to be a robust test with deviations from normality (Maxwell & Delaney, 2004).

In line with the previous analysis, the assumption of homogeneity of variances was violated, as assessed by Levene's test for equality of variances [SCR $F(1,65) =.543\ p<.05\ (.031)$; Cardio $F(1,65) =86.2\ p=.031$; RLL $F(1,65) =.776\ p=.025$. For this reason, Welch's t-test was preferred, as it is considered more robust with equal variances not assumed and with a sample size that differs significantly (Kohr & Games, 1974). In fact, the two conditions indicated (experimental and control group) differed significantly. This was determined by the Welch's t-test, which indicated a statistically significant difference between groups for each measure ($p= .000$).
Mauchly's test was non-significant for the main effect of the Series [SCR $p=.566$; Cardio $p=.482$; RLL $p=.767$]; no results were shown for the main effect of the Task as there are only two levels.

A Mixed 2 (experimental-control) x 2 (Balloon – Mannequin) x 3 (Series 1-Series 2-Series 3) Factorial ANOVA was conducted for each measure (SCR, RLL and Cardio).

There was no a statistically significant main effect for the Tasks on none of the measures [SCR $F(2,123)=1.25,\ p=001,\ \eta^2=.053$; RLL $F(2,43) = 1.39,\ p=0.3,\ \eta^2=.023$; Cardio $F(1,765) = 2.28,\ p=0.15,\ \eta^2=.064$ for Cardio, meaning that the deception levels were not significantly different over the two tasks.
There was no statistically significant main effect for Time any of the measures [SCR $F(1,423) = 1.567,\ p=.988,\ \eta^2=.012$; RLL $F(2,78) = .982,\ p=.867,\ \eta^2=.031$; Cardio $F(2,543) = .223,\ p=.421,\ \eta^2=.04$], suggesting that the habituation effect had not significantly affected the results.

Finally, ANOVA revealed a significant main effect of groups conditions for each measure [SCR $F(1.45) = .687, p=0.01;$ RLL $F(1.32) = 2.344, p=0.01;$ Cardio $F(2.65) = 66.7, p=0.01$. In fact, post-hoc tests with Bonferroni corrections applied for multiple comparison (a=.017) revealed significant differences between control group ($p<.001$) and experimental group ($p<.001$) for each measure.

4.8 Discussion

4.8.1 Accuracy Rate

The aim of this first study consisted of investigating the validity of the CIT at a very basic stage, by analysing the accuracy rate of the test in relation to verbal stimulations. For this study, it was expected that the detection rate of the test would be above the level of chance, according to the previous literature (Vrij, Mann, & Leal, 2013). This first research question has been confirmed, as all three measures used for this study detected levels of accuracy above the level of chance. Specifically, EDA and SCR from both analyses presented the highest levels of accuracy, while Cardio the lowest. These results can be confirmed by the number of detections for each measure. In fact, SCR represents the measure with more detection (52 out of 63) while Cardio the one with less (34 out of 63). From the preliminary analysis conducted on EDA, the results reported that the highest accuracy rate occurred in the first series of the Balloon Task, which represented the first series of questions in the test. The accuracy rate for this series of questions accounted for 79%, where 34 participants out of 43 were detected as deceitful. Similarly, the first series of questions for the Mannequin Task was the one that presented the highest accuracy rate (63%); however, the level of deception was lower compared to the Balloon Task. In addition, the number of participants detected in this series was 27. These results might be explained by the fact that the Balloon Series was presented as the first series of the test, therefore the levels of stress were possibly higher at the beginning of the test.

Regarding the control group, the research questions put forward that no false positive effects or results under the level of chance would be found. Preliminary analysis found that 14% of false positives were detected from this study. The polygraph was able to detect 16 participants out of 20 as truth tellers, with 80% accuracy. Similarly, in the other series of questions, almost two thirds of the sample was detected as truth tellers, with a rate that ranged from 60 to 75%.

Surprisingly, in the third series of the Balloon Task, more than half of the sample (11) were detected as deceitful with a 55% rate of deception (false positive effect). The second analysis using z-scores confirmed these results. In fact, SCR was able to detect 18 participants (two more compared to the preliminary analysis), RLL was able to detect 15 participants, while Cardio was able to detect 12. Therefore, SCR has the strongest detection rate for the control group, while Cardio has the smallest. However, the overall result is positive, as the false positive rate of this study is lower compared to other studies.

From the overall index produced with the z-scores, it was possible to see the differential responsivity of each measure. RLL and Cardio showed negative values, as both measures resulted in suppression and decrease of the physiological levels when stimulated by relevant items. Conversely, SCR's index produced a positive value, as this measurement increased when stimulated by relevant items. As expected, the typical pattern of differential responses to relevant and irrelevant items was found for all the measures in the group of guilty participants, whereas these differences tended to zero for innocents.

The accuracy of the CIT was confirmed by the analysis of participants' physiological reactions in relation to all the stimuli presented during the test. Results from all the measures found that in the experimental group participants reacted more to the two relevant items (Fork=Balloon; Knife=Mannequin), which confirms the previous results. Similarly, the physiological responses of the control group were analysed; results found that participants reacted more to different control items, such as the Key, the Glass Bottle, the Hammer and the Pen. These results are in line with previous studies, which supported the idea that the CIT's mechanisms are effective, and guilty participants are able to recognize and discriminate relevant and familiar items among irrelevant and unfamiliar items (Bashore & Rapp, 1993; Lewis & Cuppari, 2009).

Furthermore, results from univariate analysis confirmed that the control group and the experimental group significantly differed for every physiological measurement analysed, showing higher levels of detection in the experimental group. This was also supported by the analysis of the Cohen's *d*, which reported a large effect size. Therefore, part of the aim of Study I, which consisted of finding higher levels of possible detection in the experimental group than in the control group, was achieved.

4.8.2 Tasks Differences

The second part of this analysis focused on the nature and the type of the tasks. This examination was conducted in order to determine if some aspects of the tasks (sound, weapon) could have an effect on participants' accuracy rate. Although from descriptive statistics it was possible to observe that the Balloon Task had a higher impact compared to the Mannequin Task (through the accuracy rates), inferential results reported no significant outcomes. Therefore, the nature of the tasks did not play a significant role in this study. However, it would not be wise to discard completely the fact that the Balloon Task was more effective than the Mannequin Task. Preliminary analysis of EDA had been confirmed by the analysis on the other physiological measurements using z-scores. In fact, when looking at the responses from the single physiological measurements, it was possible to see some peaks from SCR and RLL on the Balloon questions; conversely, SCR and RLL tended to decrease towards the questions on the Mannequin Task. As previous studies proposed (Grosso et al., 2015; Vaidya et al., 2002; Clark, 1987), sound has one of the highest recall values when spoken or heard, compared to other sensory information. Perhaps, the loud noise of bursting a balloon might have had a greater impact on participants, compared to attacking a mannequin. In fact, the fake attack on a fake body shell like a mannequin might not have produced the psychological effect expected.

4.8.3 Habituation Effect

Finally, the habituation effect was analysed for each measure, by statistically determining if time had an effect on participants' physiological reactions during the administration of the test. As each question series was presented three times in alternate order, it was anticipated that a habituation effect would be observed. This could have decreased the accuracy rates of participants along each series of questions. From the descriptive analysis conducted on each measure, it was possible to notice a slight decrease in the levels of response on SCR and RLL, although inferential analysis found no significant results. Interestingly, Cardio showed a stable level of response across the duration of the test. This tendency can be supported by previous researchers (e.g. Gamer, et al., 2008; Elaad & Ben-Shakhar, 2006), who suggested that cardiovascular measurements are less sensitive to habituation than SCR. This has important practical implications for CIT examinations, as this measure might be able to serve for longer periods of interrogation under less than optimal conditions. To conclude, the habituation did not significantly affect participants' performance. However, a subtle decrease is clear from the descriptive statistics, which demonstrates that a natural habituation effect could affect some of the physiological measures within each participant.

4.8.4 Limitations

This final section will discuss the limitations encountered in Study I, although a more detailed section regarding the limitations of the studies will be discussed in Chapter 8 (section 8.2).

This study presented a few methodological limitations: the first one refers to one of the tasks developed. In the Mannequin Task, the object used for the task was a knife, which differs greatly from the other control items chosen for the test. Therefore, it is possible that a knife among other objects such as a screwdriver or a wooden rolling pin, is considered too salient, compared to the other control questions. It was suggested that if relevant items are more salient or more plausible than the control questions this might provoke a natural physiological response even in innocent participants (Krapohl et al., 2006). In fact, the control group had a high physiological reaction to the knife when it was presented, although they were not primed. This factor might have possibly increased the overall false positive rate of Study I. Therefore, in Study II, it would be advisable to use relevant questions that are similar to and as plausible as the other control questions.

In addition, although there were no statistical differences between the two tasks, the descriptive results demonstrate observable differences in the accuracy rates, which might be related to the low effectiveness of the Mannequin Task or to the order of the questions. Therefore, in order to clarify these ambiguous results, it could be useful to create two tasks that are very similar in nature. In this way, if an obvious difference in the accuracy rate still occurs, it could be possible to relate the main cause to the order in which the questions are delivered.

4.9 Conclusions

This exploratory study aimed to evaluate the accuracy of the CIT in its standard procedure by using different physiological measurements, in order to gain familiarity with the test and to acquire new insight into it. The exploratory aspect of this study also took into consideration how relevant information acquired can affect participants' physiological reactions in different ways.

This study was able to evaluate these elements by analysing and comparing participants' physiological response rates between an experimental group and a control group. The current study was able to demonstrate the validity of the CIT in different ways: the overall detection rate of the test was above the level of chance, which is in line with previous studies. By comparing the experimental and the control group, there was an observable difference in physiological reactions towards the relevant items. Finally, the study showed that EDA and SCR were the most accurate measurements compared to the others used for recording the results. In fact, it was able to detect correctly a greater number of participants with a higher detection rate. Conversely, the Cardio measurement was able to detect correctly the smallest number of participants with the lowest detection rate.

In conclusion, this exploratory study was able to examine the efficacy of the standard procedure of the CIT. Based on these results, it will be possible to build the next study which will try to apply methodological changes that might help to increase the validity of the CIT.

CHAPTER 5: STUDY II

5.1 Background

Study I was developed with the aim of investigating the validity of the polygraph through the application of the standard CIT procedure, which consisted of stimulating participants verbally. From Study I, it was possible to obtain some important results, which were taken into consideration while developing Study II. The results showed a sufficient rate of accuracy and a low false positive level. The next step of the research focuses on replicating Study I, by investigating the validity of the CIT with the introduction of an additional stimulation method: visual stimulation.

Although different authors (Ekman, 1992; Grubin et al., 2004; Hancock et al., 2004; Vrij, 2008) have reported successful results with the CIT, they have largely ignored the investigation of different methods of stimulation during an examination, and more specifically, whether an alternative stimulation can affect CIT examinations by increasing the validity of the accuracy rate.

Traditionally, polygraph tests are based upon the recording of physiological reactions to verbal questions that are also answered audibly by a simple “yes” or “no” (Horvath & Reid, 1972). However, despite the prevalence of photographic evidence in forensic cases, visual stimulation has not been applied as a standard method of examination with the polygraph. It was found that visual and verbal stimuli contributed equally to detection (Ekman, 1988). Although, in two similar studies using the CIT results showed differences in electrodermal response: in fact, in Ben-Shakhar’s study (2003), participants needed to study either schematic faces or verbal descriptions that varied along several dimensions. Later, participants were tested using the previous studied stimulus. Also, it has been found that verbal descriptions lead to a greater mean electrodermal response than schematic faces (DePaulo et al., 2003; Serota, Levine, & Boster, 2010). It is not clear why the Ben-Shakhar studies show an enhanced response to verbal stimuli compared to visual, while Lubow and Fein (1988) reported no differences. Seymour and Kerlin (1988) conducted a study in which verbal and visual stimuli effect was compared. They found no significant differences in participants’ responses. For this reason, it was hypothesized that lack of modality differences reported by

Lubow and Fein (Ekman, 1988) may have been due to a similar level of encoding for verbal and visual stimuli. Therefore, it is not clear whether verbal stimulation is actually more effective or more impactful on physiological reactions compared to visual stimulation.

A few other studies have been conducted in order to examine the significance of visual stimulation during polygraph examinations, applying different visual presentation methods. One of the first attempts was developed by Bradley and Ainsworth (1984), with the use of a tape recorder for delivering the questions. A few years later, Forman and McCouley (1986) administrated the polygraph by showing pictures of the relevant items on a piece of paper. A similar experiment was conducted by Bradley et al. (1989), where the stimuli were typed on cards. With the use of a similar method, Bradley et al. (1996) continued to search for the best presentation technique applying a different element to the previous methodology: participants were asked to read themselves the questions aloud. They reported that this method was very effective, as the detection rate improved. Other researchers, who supported the efficiency of the technique with different successful studies, then tested this procedure (Krapohl et al., 2006; Meijer, Bente, Ben-Shakhar, & Schumacher, 2013). In an attempt to find the best stimulation, Gamer et al. (2008) tried to use pre-recorded audio samples supported by corresponding pictures projected onto a small screen and they found that physiological reactions increased. Finally, in 2014 Elaad conducted a study in which the questions of the test were presented on a small monitor, while the examiner was verbally asking them. The study produced similar results.

In CIT examinations it is possible to use visual stimulation through slide projectors or other means to present the images (Iacono, 1991). However, a thorough search of the relevant literature yielded only a little research focused on the application of visual stimulation with the use of the CIT. One of the few applications of visual stimulation within the CIT was based on the P300 method (Onitsuka, Shiotsuka, & Iramina, 2015). The results from this study reported that the accuracy rate was higher when participants were visually stimulated, compared to a simple verbal stimulation.

Visual materials, such as photographic documentation of evidence, play an important role in the investigative process. Photographs, for example, are employed for documenting evidence and details from the crime scenes, reconstructing events, identifying suspects and enabling forensic analysis (Porter, 2007). Therefore, in forensic settings, there are different visual materials that can help the investigation in representing and interpreting facts, reconstructing visual narratives or identifying suspects; most of these elements can be used and adopted as critical elements for a CIT investigation. Due to the lack of research on the application of visual stimulation during the CIT, it is important to attempt to create an alternative method of stimulation that can be used to measure the physiological reactions of the examinees during a polygraph examination.

In line with these results, this study will examine the validity of the CIT with the support of visual stimulation. The setting proposed in this study represents one of the unique aspects of this study, as only a few studies have examined the potential of visual stimulation within the CIT. Nonetheless, Krapohl et al. (2006) recommended dedicating specific attention to the selection of visual stimuli, as pictures could contain more information or distractions than words. Examiners should standardize the illumination level of the images, their size and coloration. Every image should have the same emotional weight and be as visually plausible as the relevant item.

5.2 Aims of the experiment

The aim of this study is to examine the validity of the CIT through the support of visual stimulation. The results of this current study will then be compared to the results obtained from Study I. Results from this comparison will establish if the accuracy rate has been improved by the additional visual stimulation. Due to this methodological change, the results will reveal which stimulation method affected more participants' physiological arousal. This study will examine false positive rates on participants from the control group. In addition, the habituation effect will be analysed and its effectiveness evaluated.

Study II will also try to address some limitations highlighted in Study I: although the difference between the two tasks was found not to be statistically significant, descriptive data suggested a lower accuracy rate for the Mannequin Task. This was interpreted as a lack of strength in its efficacy as a task, due to the unbalanced relevance of the items presented and to the nature of the task itself. Therefore, an "Egg Task" will replace the second task for Study II. The Egg Task consists of smashing an egg with a hammer. In this way, the nature of the task will be very close in nature to the Balloon Task, focusing on the echoic and physical effect of the performance. In addition, all the control items presented during the examination will be no more or less relevant than the relevant one. Due to the inclusion of these small methodological amendments and the inclusion of the visual element, it was expected that more valid and homogeneous results would be obtained from the current study.

Finally, the role of the tasks will not be limited to prime participants; it will also be possible to investigate how specific elements involved in the task performance might have an impact on the participants' memory, potentially enhancing a greater physiological reaction. Therefore, detection rates between the two tasks will be analysed and compared.

5.2.1 Hypotheses:

- Accuracy rates in the CIT will be above the level of chance
- A higher deception level in the experimental group will be found
- A low false positive effect will be found
- Low levels of habituation effect will be found
- Greater physiological responses during the relevant questions will be found
- A various in the accuracy rate between the questions series will be found
- Habituation will reduce the accuracy rate of the test during the examination
- A difference between the Mannequin Task and the Egg Task will be found
- A higher overall accuracy rate will be found in Study II compared to Study I, due to the application of the visual stimulation

5.3 Participants and Recruitment Process

The current study had a sample of N=141 participants (42 males and 80 females). The sample was divided in two groups: an experimental group (*N = 111)* and a control group (*N= 30*). Participants were randomly approached in the Huddersfield University campus. Undergraduate and postgraduate students and members of the university staff constituted the great majority of the sample. The recruitment method was similar to the Study I, with the only difference that in this study SONA Systems was also used (See section 3.1); this additional method increased the participation of students.

5.4 Equipment and materials

The physiological data for this experiment was recorded with the Lafayette LX5000 SW Polygraph System. Heart rate (HR), breath rate and EDA were recorded for this study. The first task required the use of a paperclip and a balloon. The second task required the use of a hammer and an egg. The objects used for the tasks (a fork, balloons, a hammer and eggs) where purchased at the Poundland shop in the Huddersfield city centre.

Due to the nature of this study's setting, a projector and a white screen were used, for visually stimulating participants through a Power Point presentation. The Power Point presentation was created in advanced by the researcher, which contained a series of pictures carefully chosen from the browser.

5.5 Questionnaire Structure

The questionnaire had 48 questions in total. It contained three series of questions for each task (6 series in total). Every series of question contained eight questions: one irrelevant question, one buffer question, five control questions and one relevant question. The Balloon series was the first one presented, while the second was the Egg series. Each series of question was alternated in order to ensure validity (See Appendix F).

5.6 Procedure

The sample was distributed into an experimental and a control group. The experimental condition involved the completion of two tasks followed by the administration of CIT. The control group did not perform any tasks. Participants were provided with full details of the study, including the implications of participating. Participants were encouraged to sign an informed decision, without any pressure or coercion; they were notified of their right to withdraw from the study at any point during the examination. Furthermore, they were informed of their right to have their information and data immediately removed and destroyed from the study. In addition, every participant needed to indicate any possible physical conditions that could have affected the results of the test.

5.6.1 STAGE ONE

Experimental Group:

Once participants signed the consent forms and were provided with all the information, they were invited to complete two different action tasks.

In order to create a psychological context in which participants could lie about something, two tasks were created: the first one consisted in bursting a balloon with a fork. The rationale behind the development of this task it was already explained in Study I. The reason why it was decided to keep the same task relates to the intention to compare the results from the Study I to the results obtained from Study II, with the additional support of visual stimulation. On the other hand, results obtained from the Study I regarding the second task (Mannequin one) revealed a low effectiveness. Therefore, it was decided to change the typology of the task, creating an action task more similar to the first task: smashing an egg with a hammer. In this way, the echoic effect and the physical action was included in the performance of the second task (Clark, 1987; Grosso et al., 2015; Royet et al., 2000).

Every participant from the experimental group had to complete the two tasks individually, completing first the Balloon Task (Task 1), followed by the Egg Task (Task 2). In order to the analysis of the results at group level, the objects to use during the task were chosen by the research team and then instructed to every participant from the experimental group.

Task 1: the research assistant had to blow the balloon and give it ready to the participant. The participant then was instructed to burst the balloon with a fork provided by the research assistant. The participant was free to attempt to burst the balloon as many times as necessary for successfully complete the task.

Task 2: the research assistant asked every participant to smash an egg with a hammer. The research assistant had to place an egg in the middle of a desk over a small towel. The participant was instructed to smash the egg with the hammer provided by the research assistant. The participant was invited to smash the egg only once.

The priming stage took approximately 10 minutes. Once the priming stage was concluded, every participant was instructed by the research assistant not to reveal the object used during the tasks. Rather, they were instructed to deny the use of the real object during the polygraph examination by answering “No” to all the questions. After this information, every participant joined the polygraph examiner in the Psychology Lab for the administration of the polygraph. In order to avoid any judgment bias during the examination and during the analysis of the data, a blind examiner conducted the polygraph examination.

Control Group:

After having signed the consent forms, participants from the control group were instructed to enter the Psychology Lab for the administration of the polygraph. This group of participants was not primed with the tasks.

5.6.2 STAGE TWO

In the second stage of the experiment, participants from both experimental and control conditions were subjected to the administration of the polygraph. The questions were asked verbally and every series of question was repeated three times.

While the questions were verbally asked, a PowerPoint presentation displayed the images related to each item in the question through a projector on a white board size screen. Every question lasted for the duration of the formulation of the question, until the participant's answer has been expressed. After the answer, the picture was followed by a blank slide, which lasted for 25 seconds; this amount of time corresponding to the period time in which the physiological levels need to come back to normality after the stimulation (Instrument, 2003). It was ensured that the illumination level of the images was appropriate and that the participant was looking at the pictures during the examination (Iacono, 1991)

The test took approximately 20 minutes, plus 5 minutes for presentation and 5 minutes for debriefing (for more details regarding the general examination process, see Chapter 3 (section 3.5).

The data was stored on the laptop through the Data Recording System.

5.7 Results

Participants (N=143) from Study II were divided into an experimental group (N=111 *M=1.01, 95% CI [.87,1.18], SD=.902*) and a control group (N=30 *M=.45, 95% CI [.26,.74], SD=.479*). Effect sizes (Cohen's d) comparing the experimental and control groups averaged were computed. Cohen's d = (-2.41, -4.77)⁄0.982494 = 2.407516, showing a large effect size.

5.7.1 Preliminary analysis conducted on EDA with the Lykken's approach

5.7.1.1 Descriptive data

Table 10. *Overall psychological responses using Lykken's Method*

Conditions	*Overall number of responses*	*Highest EDA*	*Second highest EDA*	*Zero and low EDA*
Experimental group Task Balloon	333	173	65	95
Experimental group Task Egg	333	132	77	124
Control group Task Balloon	90	10	10	70
Control group Task Egg	90	12	23	55

Table 11. Frequencies for the Balloon and Egg Task

Task	*M*	*SD*	*Median*	*Mode*	*Variance*	*Sum*
Ball1	1.30	.859	2	2	.738	144
Ball2	1.05	.897	1	2	.804	116
Ball3	.93	.860	1	0	.740	103
Egg1	1.10	.904	1	2	.817	122

Egg2	1.03	.847	1	2	.717	114
Egg3	.84	.910	0	0	.828	93

Table 12. Percentages of detection rate among the series for the Balloon Task and Egg Task

Task	***No Detection***	***Detection***
Ball1	15%	85%
Ball2	27%	73%
Ball3	31%	69%
Egg1	26%	74%
Egg2	34%	66%
Egg3	51%	49%

Regarding the Balloon Task, the analysis of frequencies enabled a comparison between the different levels of detection in the three series presented to the experimental group. In Series 1, the CIT was able to detect 92 participants as deceptive (*M=1.30, SD=.859*), which accuracy rate accounted for 80%; similarly, in Series 2, 80 participants (*M= 1.05, SD= .897*) were detected with 73% of accuracy rate. Finally, in the last Series, 66 participants (*M=.93, SD=.860*) were detected which rate accounted for a 70%.

The same analysis of frequencies was conducted on the Egg Task. In Series1, the CIT was able to detect correctly 81 participants (*M=1.10, SD=.904*), with an accuracy rate accounted of 74%; equivalently, in Series2, 73 participants (*M= 1.03, SD= .847*) were detected with an accuracy rate of 66%. Finally, in the Series3, 55 participants were detected (*M=.84, SD=.910*) with 49% of accuracy rate.

As for the experimental group, an analysis of the frequencies was conducted in order to calculate and compare any possible accuracy rate detected from the control group. Table 12 shows the frequencies related to the Balloon Task and the Egg Task, while Table 11 shows the percentage of accuracy in both tasks in the control group.

Table 13. *Frequencies for the Balloon and Egg Task for the Control Group*

Task	***M***	***SD***	***Median***	***Mode***	***Variance***	***Sum***
Ball1	.67	.661	1	1	.437	20
Ball2	.37	.556	0	0	.309	11
Ball3	.37	.556	0	0	.309	11
Egg1	.23	.504	0	0	.254	7
Egg2	.57	.504	1	1	.254	17
Egg3	.50	.682	0	0	.682	15

Table 14. *Percentages of detection rate among the series for the Balloon Task and Egg Task in the Control Group*

Task	***No Detection***	***Detection***
Ball1	63%	37%
Ball2	67%	33%
Ball3	67%	33%
Egg1	80%	20%
Egg2	43%	57%

Egg3	60%	40%

The analysis of frequencies of the Balloon Task made possible a comparison between the levels of accuracy rates detected in the three series presented to the control group. In Series1, the CIT technique in this study was able to detect 20 participants as truth tellers (*M=.67, SD=.667*), with an accuracy rate of 63%. In Series 2, 20 participants were detected correctly (*M= .37, SD= .557*) with an accuracy rate of 67%. Finally, in the last Series, 20 participants were detected as truth tellers (*M=.37, SD=.556*) with an identical accuracy rate of Series 2.

The same analysis of frequencies was conducted on the Egg Task. In Series 1, the CIT was able to detect as truth tellers 24 participants (*M=.235, SD=.504*), with an accuracy rate of 80%; however, in Series 2, only 13 participants were detected correctly (*M= .57, SD= .504*) with an accuracy rate of 44%. Finally, in the Series 3, 18 participants were detected as truth tellers (*M=.50, SD=.682*) with 60% of accuracy rate.

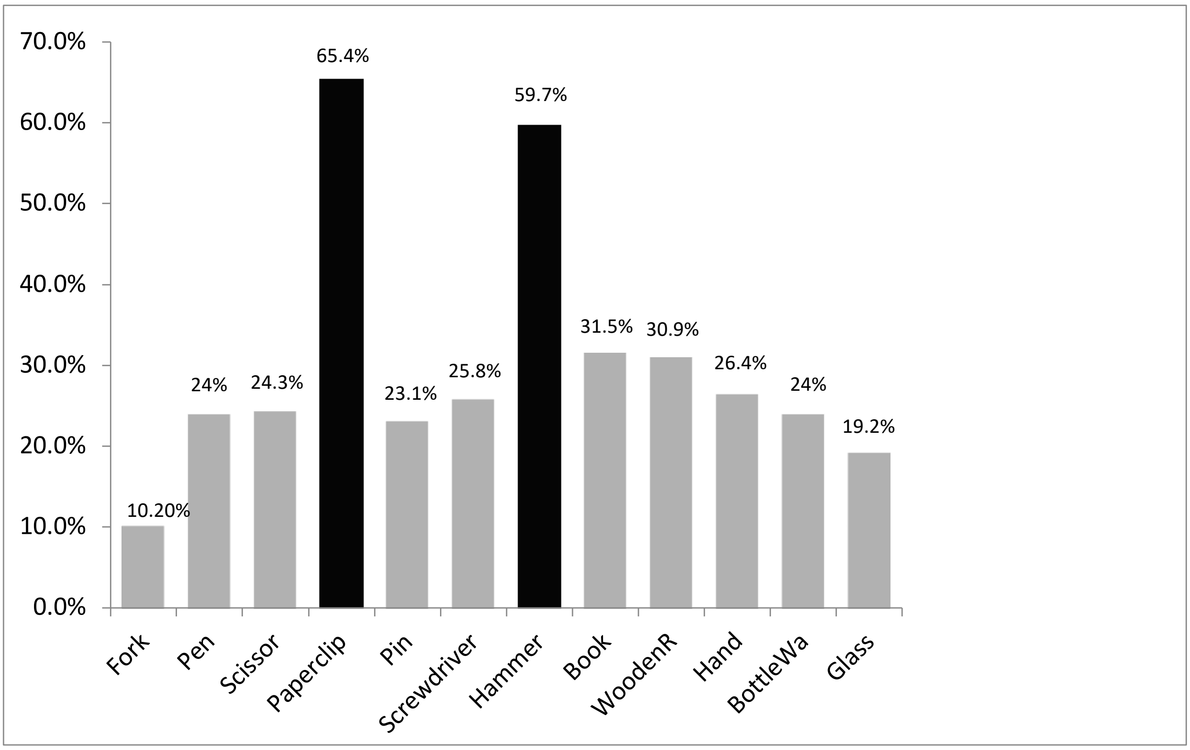

Figure 12. Physiological reactions' levels to all stimuli presented in both tasks from the experimental group

The reaction levels of participants from the experimental group towards every relevant and irrelevant item presented during the CIT was examined. Figure 12 shows that the highest physiological responses occurred with the relevant items (Fork for the Balloon Task and Hammer for the Egg Task).

The same analysis was applied to the control group. Were the reaction levels of participants towards every relevant and irrelevant item presented during the CIT was measured. Figure 13 shows that the highest physiological responses occurred to some irrelevant items (Scissor for the Balloon Task and Screwdriver for the Egg Task).

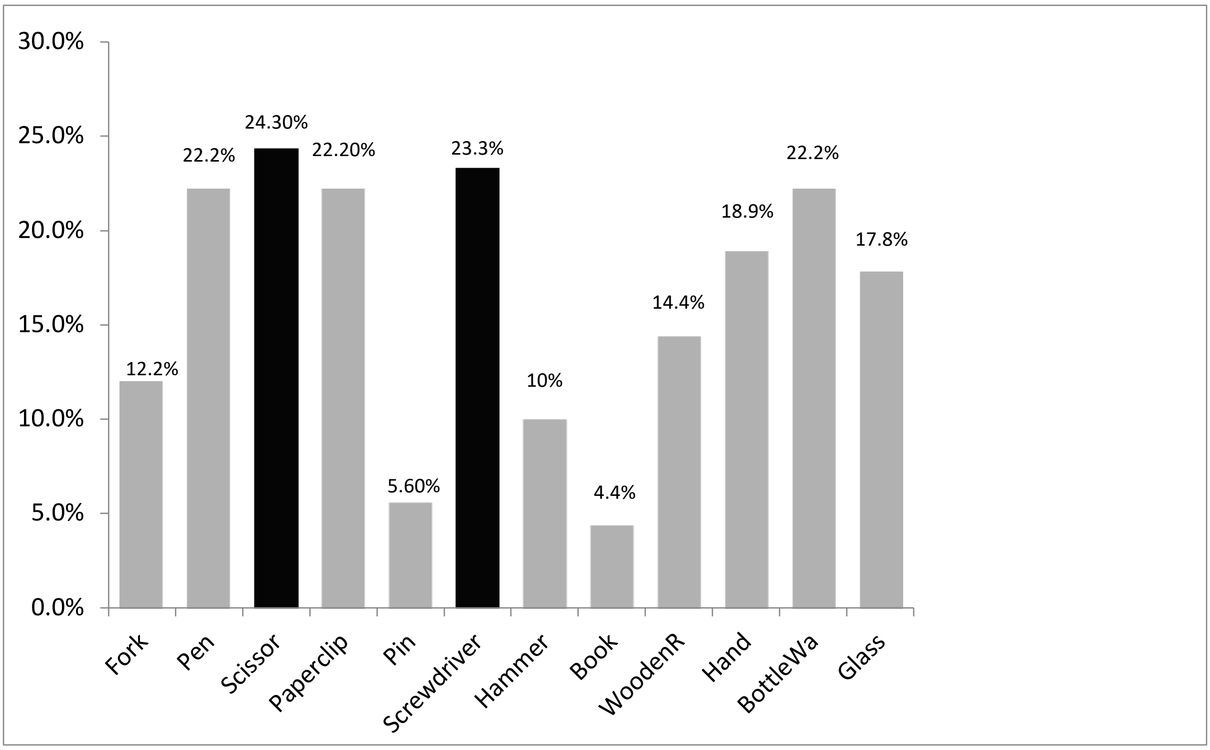

Figure 13. Physiological reactions' levels to all stimuli presented in both tasks from the control group

5.7.1.2 Inferential Analysis

It is important, as in the first analysis, to determine if the data form this study is normally distributed, by running a test for normality. In order to obtain this analysis, single variable was created, which included both experimental and control conditions. The also analysis contained a grouping variable, which specified the condition of each participant.

As the sample from Study II is quite large, the Kolmogorov-Smirnov test was the most appropriate test (Field, 2013). The analysis suggests that the overall level of deception from experimental and control group, $D=845=.286$, $p<.001$, was significantly non-normal. However, even if the data is not normally distributed (likely due to the difference between the control and the experimental group), it was suggested (Maxwell & Delaney, 2004) that it is still possible to run ANOVAs, which is robust enough to tests deviations from normality.

In line with the previous analysis, the assumption of homogeneity of variances was violated, as assessed by Levene's test for equality of variances, $F(11) =12.220\ p<.001$. For this reason, Welch's t-test was preferred, as it is considered more robust with equal variances not assumed and with a sample size that differs significantly (Kohr & Games, 1974). In fact, the two conditions indicated (experimental and control group) differed significantly. This was determined by the Welch's t-test, which indicated a statistically significant difference between groups $F(81.620) =120.953$, $p= .000$.

Mauchly's test was non-significant for the main effect of the Series (p=.208); no results were shown for the main effect of the Task as there are only two levels.

A 2 (Group conditions: control and experimental group) x 2 (Task: Balloon and Egg) x 3 (Series: three series for each task) Mixed Factorial ANOVA was conducted.

Factorial ANOVA revealed a significant main effect of group conditions, $F(1.38)=75.91$, $p=.031$, $\eta^2=.35$. In fact, Post-hoc tests with Bonferroni corrections applied for multiple comparison (a=.017) revealed a significant difference between control group ($p<.001$) and experimental group ($p<.001$).

The analysis revealed a non-significant main effect of tasks, $F(1,87.77) =1.450$, $p=.231$, $\eta^2 =.010$, which means that the deception levels were not significantly differed over the two tasks. Finally, significant results were shown from the main effect of Series $F(2,269) =.468$, $p=0.26$, $\eta^2=.019$. These results suggest that a habituation effect had significantly affected participants' physiological reactions during the test.

The analysis reported a significant effect for time $F(1,19) = .987$, $p=.021$. Post-hoc tests with Bonferroni corrections applied for multiple comparison (a=.017) revealed a significant difference between Series 1 ($p<.001$) and Series 2 ($p<.001$).

These results suggest that habituation effect had a significant effect on participants' physiological reactions during the test. These results can be confirmed also from the general means displayed on the Table 11, in which it is visible a clear decrease in the physiological responses among the duration of the test.

5.7.2 Follow up analysis conducted on SCR, RLL and Cardio Using Z-Scores

5.7.2.1 Descriptive data

The first analysis consisted in the classification obtained through the Lykken's method (see Table 14), which shows that SCR was the measure that accurately detected more participants, compared to the RLL and Cardio. Also, from this method, it was possible to obtain deception rates for each measure, where the highest was obtained by the SCR, the second highest by the RLL and the lowest by the Cardio.

Table 15. *Detection rates computed for the three physiological measures and across measures in Study 2 for the experimental group using the Lykken's method*

Measures	n. of detection (experimental)	n. of detection (control)	Percentage detected
RLL	67	22	60%
SCR	88	25	65%
Cardio	48	19	52%

In a first step, the mean z-standardised response differences between relevant and irrelevant items were compared between the two experimental groups. Afterwards the mean of these measures was computed as an overall index of the differential responsivity in each physiological measure, expecting negative values for RLL and Cardio and a positive value for SCR as indication of knowledge of relevant information. As expected, the typical pattern of differential responses to relevant and irrelevant items was found for all the measures in the group of guilty participants, whereas these differences tended to zero for innocents (see Table 15).

Table 16. *Means and standard deviation of the standardised physiological response differences between relevant and irrelevant items within each experimental condition*

Measure	Guilty (111)	Innocent (30)

	M	SD	M	SD
RLL	-0.72	0.32	0.04	0.63
SCR	0.95	0.82	0.02	0.81
Cardio	-0.54	0.40	- 0.05	0.31

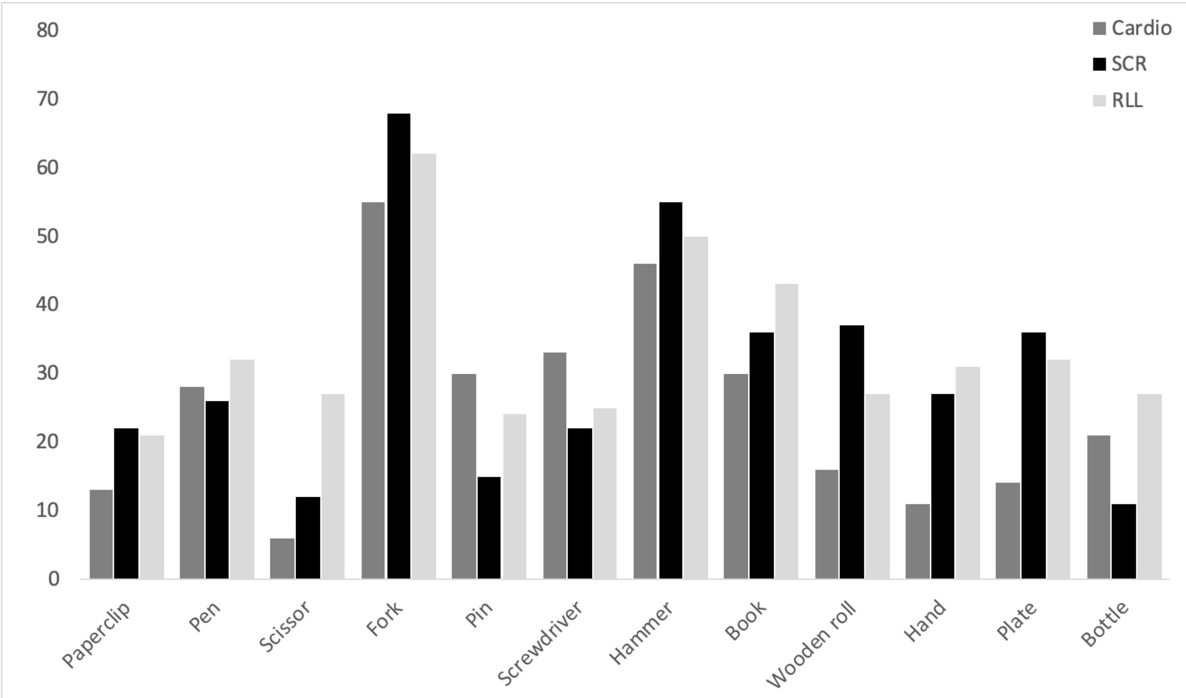

Figure 14. Physiological reactions' levels to all stimuli presented in both tasks from the experimental group

The reaction levels of participants from the experimental group towards every relevant and irrelevant item presented during the CIT was examined. Figure 14 shows that the highest physiological responses from SCR, RLL and Cardio occurred on the relevant items (Paperclip for the Balloon Task and Hammer for the Egg Task).

The same analysis was applied to the control group. Figure 15 shows the reaction levels of participants towards every relevant and irrelevant item presented during the CIT was measured. As expected, the highest physiological responses occurred on some irrelevant items (for the RLL it occurred on Book; for SCR it occurred on the screwdriver; for Cardio it occurred on the Pin)

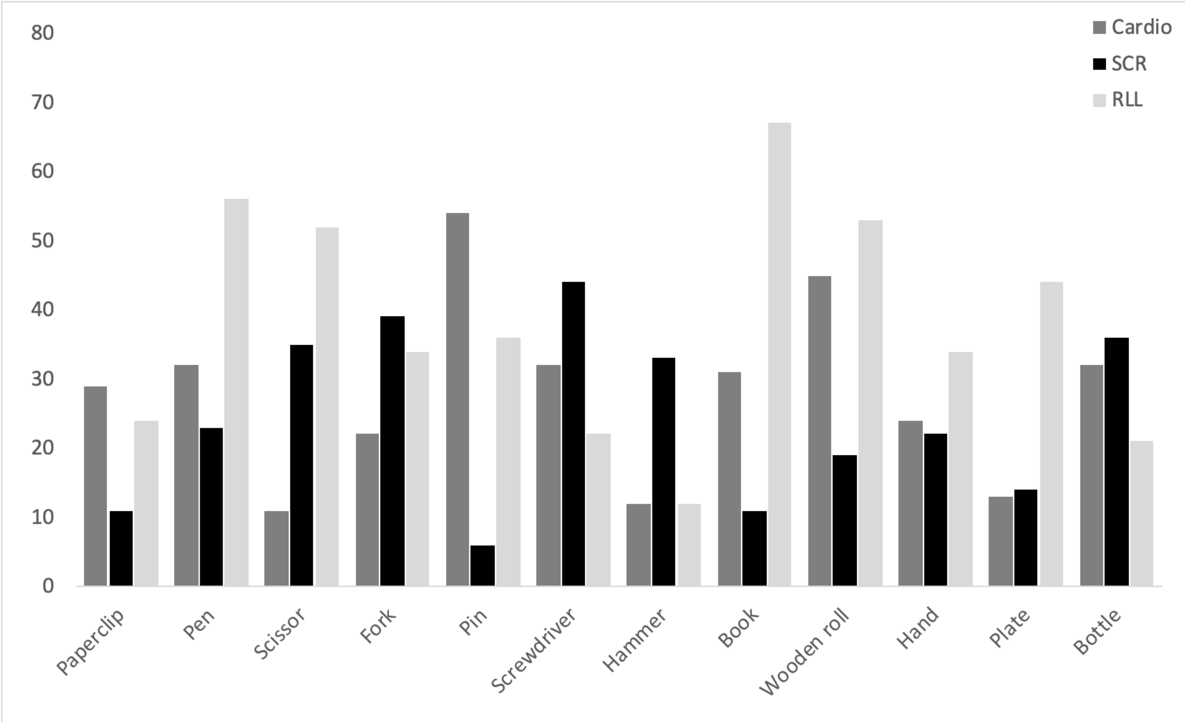

Figure15. Physiological reactions' levels to all stimuli presented in both tasks from the control group

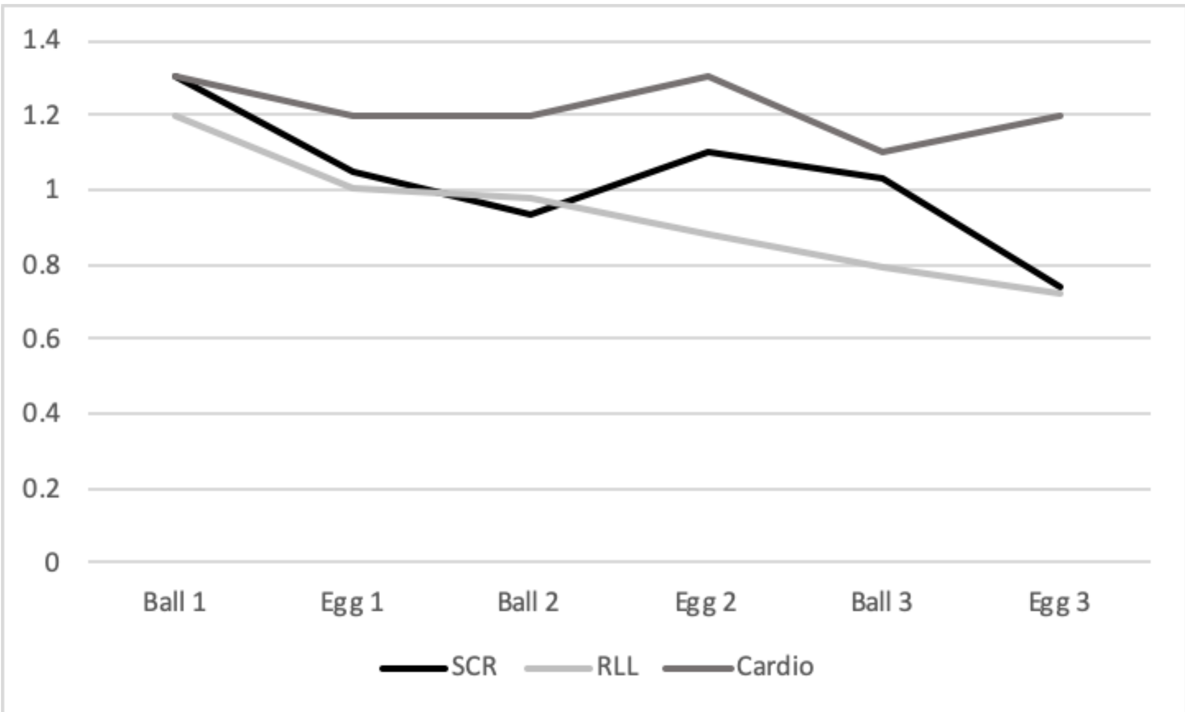

Figure16. *Means of the relevant items presented during the test from SCR, Cardio and RLL*

From figure 16 it is possible to see how the means of each measure develop along the duration of the polygraph test. The figure suggests a clear decrease of the SCR and RLL. The most dramatic decline is visible on the SCR until Series 2 of the Balloon task, then it increases for the Series 2 of the Egg task, following another clear decrease along Series 3. This suggests that the habituation effect affected SCR. However, as these are only preliminary results, the significance of this effect will need to be confirmed by inferential analysis. Similarly, RLL levels presented a slight but continuous decrease along the duration of the test. Finally, Cardio levels appear relatively stable during the whole duration of the test, with a small increase for Series 2 of the Egg task.

5.7.2.2 Inferential analysis

It is important to determine if the data from this study is normally distributed by running a test for normality. In order to obtain this analysis, a single variable was created, which included both experimental and control conditions from all the physiological measures. The analysis contained a grouping variable, which specified the condition of each participant.
As the sample from Study II is quite large, the Kolmogorov-Smirnov test was the most appropriate test (Field, 2013). The analysis suggests that the overall level of deception from experimental and control group, $D=150 = .958$, $p<.001$, was significantly non-normal. However, even if the data is not normally distributed (likely due to the difference between the control and the experimental group), it was suggested (Maxwell & Delaney, 2004) that it is still possible to run ANOVAs, which is robust enough to tests deviations from normality.

In line with the previous analysis, the assumption of homogeneity of variances was violated, as assessed by Levene's test for equality of variances [*SCR* $F(1,57) = 5.43$ $p<.05$ $(.023)$; Cardio $F(1,57) = 31.5$ $p=.002$; RLL $F(1,57) = .433$ $p=.003$. For this reason, Welch's t-test was preferred, as it is considered more robust with equal variances not assumed and with a sample size that differs significantly (Kohr & Games, 1974). In fact, the two conditions indicated (experimental and control group) differed significantly. This was determined by the Welch's t-test, which indicated a statistically significant difference between groups for each measure ($p= .000$).

Mauchly's test was non-significant for the main effect of the Series (SCR *p=.317*; Cardio *p=.645*; RLL *p=.321*); no results were shown for the main effect of the Task as there are only two levels.

A Mixed 2 (experimental-control) x 2 (Balloon–Egg) x 3 (Series 1-Series 2-Series 3) Factorial ANOVA was conducted for each measure (SCR, RLL and Cardio).

There was no a statistically significant main effect for the Task on none of the measures [SCR $F(1,102) = 4.30, p=0.72, \eta^2=.004$; RLL $F(2,84) = 0.37, p=1.3, \eta^2=.016$; Cardio $F(1,105) = 3.2, p=1.43, \eta^2=.023$], meaning that the deception levels were not significantly different over the two tasks.

There was no statistically significant main effect for Time for RLL and Cardio [RLL $F(2,13) = 0.72, p=0.73$; Cardio $F(1,23) = 2.14, p=.863$]. However, there was a statistically significant main effect for the series for SCR [$F(2, 68) = 0.53, p=0.12$ $\eta^2=.019$]. This was confirmed by post-hoc tests with Bonferroni corrections applied for multiple comparison (a=.017), showing a significant difference between Series 1 ($p<.001$) and Series 3 ($p<.001$). This suggests that a habituation effect had a significant effect on the SCR results.

Finally, ANOVA revealed a significant main effect of groups conditions for each measure [SCR $F(1.42) = 76.73, p=.01;$RLL $F(2.23) = 53.4, p=0.21;$ Cardio $F(2.42) = 46.1, p=0.10$]. In fact, post-hoc tests revealed significant differences between control group (p<.001) and experimental group (p<.001) for each measure.

5.8 Discussion

5.8.1 Accuracy Rate

The aim of Study II was to examine the validity of the CIT through the support of visual stimulation. The additional support was applied with the intention to determine whether a different stimulation could increase participants' physiological arousal, thus improving the CIT accuracy. In line with previous studies, the test was able to detect deception above the level of chance (Vrij et al., 2013). This first research question has been confirmed, as all three measures used for this study detected levels of accuracy above the level of chance. Similarly to Study I, EDA and SCR presented the highest levels of accuracy while Cardio the lowest. These results can be confirmed by the number of detections for each measure. In fact, SCR represents the measure with more detection (88 out of 111), while Cardio the one with least (48 out of 111). From the preliminary analysis conducted on EDA, the highest accuracy rate was found in the first series of questions, with an 85% accuracy rate, where 92 participants out of 111 were detected as deceitful. The second highest occurred in the first series of the Egg Task, with a 74% accuracy rate. In the second series of the Balloon Task, the polygraph was able to detect 80 participants correctly, with a 73% accuracy rate. As in Study I, the Balloon Task produced higher accuracy rates. This could be explained by the fact that this task was presented before the Egg Task, so the level of stress was higher at the beginning of the questionnaire. This can be confirmed by the means from descriptive results, where accuracy rate levels were higher at the beginning of the first series and lower during the last one (3rd Egg Series: 49%).

The accuracy of the CIT was confirmed by the analysis of participants' physiological reactions in relation to all the stimuli presented during the test. Results from all the measures showed that in the experimental group participants reacted more to the two relevant items (Fork=Balloon Task; Hammer=Egg Task). In the same way, the physiological responses of the control group were analysed; results showed that participants reacted more to irrelevant items, such as Scissors and a Screwdriver. This is in line with previous studies which supported the idea that the CIT's mechanism works, and guilty participants will be able to recognize and discriminate relevant and familiar items among irrelevant and unfamiliar items (Bashore & Rapp, 1993; Lewis & Cuppari, 2009).

From the overall index which indicates the differential responsivity of each measure, RLL and Cardio showed negative values, as both measures resulted in suppression and decrease of physiological levels when stimulated by relevant items. Conversely, SCR's index produced a positive value, as this measurement increased when stimulated by relevant items. As expected, the typical pattern of differential responses to relevant and irrelevant items was found for all the measures in the group of guilty participants, whereas these differences tended to zero for innocents.

As for the previous study, preliminary analysis found that 13% of false positives were detected from this study. Therefore, the false positive effect was slightly lower than in the previous study. As for the accuracy rate, the CIT was more accurate in detecting truth tellers in the first series of questions; in fact, in the first series of the Balloon Task, the polygraph was able to detect correctly 20 participants out of 30, with a 63% accuracy rate. Similar results were found for the rest of the series of the same task. In the first series of the Egg Task, the polygraph detected 24 participants as truth tellers with 80% accuracy, which represents a good level of accuracy. However, this rate dramatically decreased in the second series of the Egg Task, with only 13 participants detected as truth tellers and just a 43.3% accuracy rate. The second analysis using z-scores confirmed these results. In fact, SCR was able to correctly detect 25 participants, RLL was able to detect 22 participants, while Cardio was able to detect 18. Therefore, SCR had the strongest detection rate for the control group, while Cardio had the smallest. Nonetheless, the overall level of false positive effect rate for this study is lower compared to the existing literature (Podlesny & Raskin, 1977; D. Raskin et al., 1976). In addition, through the initial analysis conducted, it was possible to see a clear and significant difference between the control group and the experimental group in relation to the accuracy rate detected during the test. This was supported by analysis of the Cohen's *d*, which reported a large effect size. Therefore, part of the aim of Study II was achieved.

5.8.2 Visual Stimulation

The second main aim of this study, the visual stimulation, was applied during the administration of the test. It was expected that we would find a higher accuracy rate compared to the previous study, in which only verbal stimulation was applied. The comparison between Study I and Study II accuracy rates was necessary in order to determine if the application of the visual stimulation helped to increase the physiological reactions of participants when trying to lie. As in Study II, the second task was changed from a Mannequin Task to an Egg Task; it was possible to make a comparison just with the Balloon Tasks. Results revealed an increase in the detection rates of all the measures applied in Study II. Specifically, EDA, SCR and RLL presented higher detection rates compared to Study I. Therefore, in line with previous studies (Onitsuka et al., 2015), it can be put forward that visual stimulation during the administration of the test played a significant role, increasing the accuracy of the CIT.

This conclusion is in line with previous studies (Intraub & Nicklos, 1985), which stated that compared to other sensory information, visual input creates the strongest recall value and allows the widest spectrum of levels of processing modifiers; this is because pictures and visual stimuli have a great potential for both physical and semantic richness. Although not extensive, literature exists regarding the application of this particular methodology, and these results are in line with the few studies who found beneficial effects in the application of visual stimulation (Elaad, 2014; Gamer et al., 2008). In addition, this study clarifies some doubts around the idea that verbal and visual stimuli contribute equally to detecting deception during the administration of a polygraph examination (Seymour & Kerlin, 2007). Due to these positive results, visual stimulation will be applied in the subsequent studies of this thesis, as it can be considered a useful supplement during polygraph administration.

5.8.3 Tasks Differences

The second part of this analysis focused on the nature and the typology of the tasks, in order to determine if the typology of the task could have affected participants' accuracy rates differently. Although from descriptive statistics it was possible to notice a slightly higher impact from the Balloon Task compared to the Egg Task, the outcome from univariate analysis reported no significant results. Due to results from this study presenting similar consequences, clearly the discrepancy between the accuracy rates of the task should be associated with the order in which the questions are presented. In fact, the Balloon Task was the first series presented as in Study I.

5.8.4 Habituation Effect

Results from both preliminary and follow-up inferential analysis showed a significant main effect of time during the examination of EDA and SCR. Due to the structure and the length of the test, a habituation effect was expected, which could have negatively impacted the accuracy rates of participants along the series of questions. From an initial inspection, a slight decrease can be noticed from descriptive data (specifically from the percentages and the mean differences of the series of questions). Preliminary analysis of EDA showed that the highest level of accuracy rates was found in the first series of the Balloon Task (85%), while the lowest was found in the last series, which accounted for the Egg Task (49%). From the differences in the means and from the accuracy rates, it is clearly the impact of the habituation effect. In fact, participants' physiological reactions gradually decreased with the repetitions of the questions, habituating their physiological responses to the relevant item presented (Sokolov, 1963). Therefore, these results reveal that the first series of questions for both tasks

(1^{st} series and 2^{nd}, as the task series were alternated) had the strongest impact on physiological levels of participants. From the descriptive analysis conducted on each measure, it was possible to notice a slight decrease of the levels of response on SCR and RLL. Although, the results from inferential analysis found significant results only for SCR, the means of the RLL measure presented a smooth decrease along the series of questions. Similarly, the Cardio measure presented a slight but stable level of response across the duration of the test. This tendency can be supported by previous researchers (e.g. Gamer et al., 2008; Elaad & Ben-Shakhar, 2006), who suggested that cardiovascular measurements are less sensitive to habituation than SCR. This has important practical implications for CIT examinations, as this measure might be able to serve for longer periods of interrogation under less than optimal conditions. As the habituation effect has played an important role in this study, it is worth considering that multiple series in the test can affect negatively the accuracy of the test, by decreasing the intensity of the SCRs during the examination. In support of this, results showed that the Balloon Task always presented a higher accuracy rate than the second task, perhaps because it was always presented as the first series of questions in the test. Therefore, it is put forward that the habituation effect significantly determined the performance of the participants over time for this study.

5.8.5 Limitations

This final section will discuss the limitations encountered in Study II, although a more detailed section regarding the limitations of the studies will be discussed in Chapter 8 (section 8.2).

Conversely from Study I, the present study showed a significant habituation effect, which affected the final results. In fact, physiological reactions were higher during the first series of questions and lower in the final series of questions. Results from both descriptive and inferential analysis show a clear decrease of participants' responses during the test. These conclusions are in line with the existing literature (Ben-Shakhar & Elaad, 2002; Podlesny & Raskin, 1977; Thompson & Amy, 2009) in which it is indicated that the habituation effect is a limitation of the CIT, especially in a laboratory setting. Hence, this limitation needs to be considered for the following studies. Although a minimum of two repetitions is necessary in order to ensure standardization to the test's results, the reduction of repetitions of questions could help to decrease the habituation effect. This methodological issue will be taken into consideration for future studies.

5.9 Conclusions

The purpose of this study was to evaluate the accuracy of the CIT through an alternative method of stimulation: visual stimulation. The current study was able to demonstrate the validity of the CIT in different ways: the overall detection rate of the test was above the level of chance, which results are in line with previous studies.

This study also proposed the application of an additional method of stimulation, consisting of visual stimulation, which represents the innovative aspect of this study. As expected, physiological reactions were higher when participants were stimulated not just verbally but also visually during the CIT examination. Finally, the study showed that EDA and SCR were the most accurate measurements compared to the other used for recording the results. In fact, it was able to detect correctly a greater number of participants with a higher detection rate. Conversely, Cardio was able to detect correctly the smallest number of participants with the lowest detection rate.

These results provide a great contribution to the existing literature, supporting the application of visual stimulation as a positive additional supplement during polygraph administration. In order to increase the validity of the present results, the next study will need to investigate how the accuracy rate of the CIT could change if applied in a more realistic forensic context. This will help to increase the external validity of this study and make the results relevant for real future applications.

CHAPTER 6: STUDY III

6.1 Background

Study II was developed with the aim of investigating the validity of the polygraph through the application of visual stimulation. The results have shown that the additional application of visual stimulation increased the accuracy of the CIT, compared to Study I where only verbal stimulation was applied. The previous two studies tried to investigate the mechanisms of the CIT and its validity through the performance of simple action tasks; the aims around the previous studies focused on investigating cognitive processes that could increase physiological reactions associated with deception, such as the effect of echoic, visual and physical factors. However, as the CIT has been designed to detect relevant information regarding criminal cases, Study III aims to evaluate the validity of the test in a criminal context. For this reason, Study III will involve a mock crime setting. Much research has been conducted in the area of psychophysiological detection of deception with the support of mock crimes (Ben-Shakhar & Furedy, 1990; Kleiner, 2002; Lykken, 1998; MacLaren, 2001). Although the use of mock crimes has many disadvantages compared to field research, such as a limited external validity, many advantages need to be considered, such as internal validity.

In a mock crime setting it is possible to successfully control the validation criterion: in polygraph studies, the validation criterion plays a crucial role when analysing the results, as it represents the real status of participants as being truthful or deceitful (Gödert, Gamer, Rill, & Vossel, 2005). Therefore, it would be important for this study to investigate the practical utility of the CIT in a mock crime situation. However, a lot of research investigates the validity of the CIT through a mock crime scenario, but in this study instead of performing a crime, participants will be asked to plan one.

Many studies in the area of cognitive neuroscience have tried to investigate deception using mock crime scenarios (Kozel et al., 2005; Mohamed & Sloane, 2006). Typically, studies investigating the CIT used the same methodology, based on the detection of mock crime knowledge after having performed a mock crime (Ben-Shakhar & Dolev, 1996; Farwell & Donchin, 1991; Langleben et al., 2002; Meijer et al., 2013; Meixner & Rosenfeld, 2011;

Rosenfeld et al., 2008). Then, during the examination, the CIT will test for crime-related details that were central or mentioned during the mock crime. This procedure is based on the standard theoretical mechanism behind the CIT; specifically, the exposure to crime-related details might lead to differences between deceitful and truthful examinees (Meixner & Rosenfeld, 2011). However, diversely from the studies conducted on the validity of the polygraph, this study will not investigate participants' performances during the crime simulation, but it will focus on their involvement in planning a crime.

This mechanism makes the CIT potentially suitable for different crime scenarios (property crimes, murder and espionage); however, only a small amount of studies have investigated this aspect of the CIT's accuracy. One of the relatively new criminal contexts in which the CIT has been used experimentally is terrorism; in fact, only a few studies have tested the CIT as an antiterrorism tool (Meijer, Smulders, & Merckelbach, 2010; Meixner & Rosenfeld, 2011). Therefore, this area of application has not been extensively researched. Although much research has been conducted around the CIT validity within the context of criminal scenarios, only a few studies have investigated the potential application of the CIT as a tool for preventing crimes, such as suspect prioritization. For this reason, the present study will try to investigate the accuracy of the CIT when trying to detect relevant information associated with a crime.

6.2 Aims of the experiment

The aim of this third part of the study is to examine the validity of the CIT with the use of a mock crime scenario. Among others, one of the most problematic criminal behaviours that has grown more in recent decades is terrorism; therefore, the need to develop new techniques of counter-terrorism has become a priority. For these reasons, this study will combine the investigation of the validity of the CIT with a relatively new application: trying to investigate the intentions or a simple involvement in a crime. The innovation of this method can be found in the application of an alternative way to investigate planning crime-activities. The unique aspect of this study is related to the investigation of personal involvement or intentions by also using visual stimulation. In fact, the proposed approach has not been combined before in any of the published literature to the best of the researcher's knowledge.

The mock crime will have the function of priming the experimental group. However, through the task, it will also be possible to investigate whether some elements of the plan might have a different cognitive impact on the memory of participants, enhancing greater physiological responses.

6.2.1 Hypotheses:

- Accuracy rates in the CIT will be above the level of chance
- A higher deception level in the experimental group will be found.
- A low false positive effect will be found.
- Low levels of habituation effect will be found.
- Greater physiological responses during the relevant questions will be found
- A various in the accuracy rate between the questions series will be found
- A difference between the City, the Location and the Weapon will be found.

6.3 Participants and Recruitment Process

This study included 53 participants (37 males and 16 females). The sample was divided into two groups: an experimental group (*N = 34)* and a control group (*N= 19*). Participants were recruited as described in Study II (see Chapter 5, section 5.3. Also, see section 3.1 for more information regarding the recruitment procedure).

6.4 Equipment and materials

The physiological data for this experiment was recorded with the Lafayette LX5000 SW Polygraph System, owned by the department of Psychology at the University of Huddersfield. The equipment was purchased by the university from the official American Lafayette website (http://lafayettepolygraph.com/). Heart rate (HR), breath rate and EDA were recorded for this study. The mock crime was conducted with the use of pictures of some cities, locations related to the cities and weapons. The pictures were carefully selected from the browser and inserted in a Power Point presentation. Due to the nature of this study's setting, a projector and a white screen were used, for visually stimulating participants through the Power Point presentation. The Power Point presentation was created in advanced by the researcher, which contained a series of pictures carefully chosen from the browser. (see Chapter 3, section 3.5).

6.5 Questionnaire Structure

The CIT had 46 questions in total. It contained two series of questions for each crime area (2 series in total). Every series of question contained seven questions: one irrelevant question, one buffer question, four control questions and one relevant question. The City series was the first presented, followed by the Location and finally the Weapon series (see Appendix G).

6.6 Procedure

The sample was distributed into an experimental and a control group. The experimental condition consisted of a discussion with the research assistant followed by the administration of CIT. The control group did not have any discussion. Participants were provided with full details of the study, including the implications of participating. Participants were encouraged to sign an informed decision, without any pressure or coercion. They were notified of their right to withdraw from the study at any point during the examination. Furthermore, they were informed of their right to have their information and data immediately removed and destroyed from the study. In addition, every participant needed to indicate any possible physical conditions that could have affected the results of the test.

6.6.1 STAGE ONE:

Experimental Group:

Once participants signed the consent forms and were provided with all the information, they were invited to participate in a discussion with the assistant researcher. In order to create a psychological context in which participants could lie about something, they were asked to plan a criminal attack. In order to the analysis of the results at group level, the three crime options were chosen by the research team, and then instructed to every participant from the experimental group. Although, the polygraph examination was conducted by a blind examiner.

The relevant options chosen for the experiment were:

City of the attack: **PARIS**

Location of the attack: **THE EIFFEL TOWER**

Method of the attack: **GUNS**

For guaranteeing standardisation and consistency, all the cities included in the questionnaire (relevant and control questions) included places in which a minimum of one crime attack occurred. The locations, instead, involved spots of the cities where no criminal attacks occurred before. In fact, there were chosen only famous touristic spots that could be easily associated with the correspondent city (e.g. New York- Empire State Building). This choice was made to avoid any associations of generic locations with more than one city (e.g. restaurant, tube). Weapons were chosen based on the most used for criminal purposes. The relevant items were chosen by draw. Clearly, the location needed to match with the correspondent city.

For the current study, the priming stage involved a discussion around the relevant information provided by the research team. Therefore, participants were not involved in any decision process or action task. For this reason, the risk for participants in encountering in a weak encoding and memory retrieval was high (Meijer, Verschuere, & Merckelbach, 2010). For this reason, participants were invited to plan the criminal attack by providing details and suggestions on the relevant options. This process was supported by asking a few open questions, in order to push participants to discuss the topics. For this reason, nine open questions were asked, which helped the participants to discuss details of the plan and consolidate the significance around the relevant options (e.g. why do you think Paris would be the perfect city for the attack? Any disadvantages?).

Every participant from the experimental group had to complete the priming discussion providing the following information:

City: Describe how you are going to enter the city; Justify why Paris might be a good city for a terrorist attack; highlight any the disadvantages.

Location: Describe why the Eiffel Tower might be the best location for a terrorist attack; describe in which specific area of the Tower the attack will take place; describe who would be the target of the attack and why.

Weapon: Describe from where you would take the weapons for the attack; highlight the advantages of using guns for this type of attack; highlight any disadvantages of using guns for this type of attack.

While planning the attack, three images were shown to every single participant. The images pictured the crime-related information (Paris, the Eiffel Tower, and Guns). The same pictures were then shown in the PowerPoint presentation during the CIT examination. This procedure was conducted in order to increase the effectiveness of the priming process, making the recognition of the items during the examination stronger.

The priming stage took approximately 10 minutes. Once the priming stage was concluded, every participant was instructed by the research assistant not to reveal the object used during the tasks. Rather, they were instructed to deny the use of the real object during the polygraph examination by answering "No" to all the questions. After this information, every participant joined the polygraph examiner in the Psychology Lab for the administration of the polygraph. In order to avoid any judgment bias during the examination and during the analysis of the data, a blind examiner conducted the polygraph examination.

Control Group:

After having signed the consent forms, individual participants from the control group were instructed to enter the Psychology Lab for the administration of the polygraph. This group of participants were not primed with the discussion.

6.6.2 STAGE TWO

In the second stage of the experiment, participants from both experimental and control conditions were subjected to the administration of the polygraph. The questions were asked verbally, and each series of questions were repeated three times.

While the questions were being asked, a Microsoft PowerPoint presentation displayed the images related to each item through a projector onto a large white board screen. Every question lasted for the duration of the formulation of the question until the participant's answer has been expressed. After the answer, the picture was followed by a blank slide, which lasted for 25 seconds. This amount of time corresponds to the period time in which the physiological levels need to come back to normality after the stimulation. It was ensured that the illumination level of the images was appropriate and that the participant was looking at the pictures during the examination (Iacono, 1991).

The test took approximately 20 minutes, with an additional 10 minutes for presentation and priming discussion and a further 5 minutes for debriefing (for more information regarding the general examination process, see Chapter 3, Section 3.5). The data was stored on password-protected laptop in the Data Recording System of the polygraph software.

6.7 Results

Participants (N=53) from Study II were divided into an experimental group (N=34 *M=1.11, 95% CI [.98,1.26], SD=.889*)) and a control group (N=19 *M=.42, 95% CI [.53,.89], SD=.398*).
Effect sizes (Cohen's d) comparing the experimental and control groups averaged were computed. Cohen's d = (5.1142 - 2.2794) ⁄ 0.385497 = 7.353623, showing a large effect size.

6.7.1 Preliminary analysis conducted on EDA with the Lykken's approach

6.7.1.1 Descriptive data

Table17. *Overall psychological responses using Lykken's Method*

Conditions	***Overall number of responses***	***Highest EDA***	***Second highest EDA***	***Zero and low EDA***

Experimental group City	68	37	13	18
Experimental group Location	68	29	18	21
Experimental group Weapon	68	23	15	30
Control group City	38	6	10	22
Control Group Location	38	2	6	30
Control group Weapon	38	1	7	30

Table 18. Frequencies for the Terrorist Attack Plan from the experimental group

Topic	*M*	*SD*	*Median*	*Mode*	*Variance*	*Sum*
City1	1.38	.779	2	2	.607	143
Location1	1.15	.892	1	2	.796	101
Weapon1	1.03	.904	1	2	.817	110
City2	1.15	.925	1.50	2	.857	127
Location2	1.18	.869	1	2	.756	101
Weapon2	.82	.904	.50	0	.816	89

Table 19. *Percentages of detection rate among the series for the City, Location and Weapon topics in the experimental group*

Topic	***No Deception***	***Deception***
City1	18%	82%
Location1	32%	68%
Weapon1	38%	62%
City2	35%	65%
Location2	29%	71%
Weapon2	50%	50%

In regards to the City Topic, the analysis of frequencies made possible a comparison between the levels of accuracy rate detected in the two series presented to the experimental group. In Series1, the CIT was able to detect 28 participants out of 34 as deceptive (*M=1.38, SD=.779*), where the accuracy rate accounted for 82%; in Series2, 22 participants (*M= 1.15, SD= .925*) were detected with an accuracy rate of 65%.

The same analysis of frequencies was conducted on the Location Topic. In Series 1, the CIT detected 23 participants out of 34 correctly (*M=1.15, SD=.892*), which accuracy rate accounted for 68%; similarly, in Series2, 24 participants (*M= 1.18, SD= .869*) were detected with an accuracy rate of 71%.

Finally, for the Weapon Topic, 21 participants (*M=.84, SD=.910*) were detected, with 62% of accuracy rate. While in Series2, 17 participants (*M=.82, SD=.904*) were detected, with an accuracy rate of just 50%.

Table 20. *Frequencies for the Terrorist Attack Plan in the Control Group*

Topic	*M*	*SD*	*Median*	*Mode*	*Variance*	*Sum*
City1	.89	.937	1	0	.877	17
Location1	.58	.838	0	0	.702	11
Weapon1	.63	.895	0	0	.801	12
City2	.74	.733	1	0	.538	14
Location2	.95	.848	1	0	.719	18
Weapon2	.47	.697	0	0	.485	9

Table 21. *Percentages of detection rate among the series for the City, Location and Weapon topics in the Control Group*

Topic	*No Detection*	*Detection*
City1	63%	37%
Location1	79%	21%
Weapon1	79%	21%
City2	53%	47%
Location2	53%	47%
Weapon2	79%	21%

In regards to the City Topic, the analysis of frequencies made possible a comparison between the different levels of accuracy detected in the two series presented to the control group. In

Series1, the CIT was able to detect 12 participants out of 19 (*M=.89, SD=.837*) as truth tellers, which accuracy rate accounted for 63%; in Series2, 10 participants (*M= .74, SD= .733*) were detected correctly with an accuracy rate of 53%.

The same analysis of frequencies was conducted on the Location Topic. In Series1, the CIT was able to detect 15 participants out of 19 (*M=.58, SD=.838*), with 79% of accuracy rate; in Series2, 15 participants (*M= .95, SD= .848*) were detected as truth tellers with an accuracy rate of 53%.

Similarly, for the Weapon Topic 15 participants (*M=.63, SD=.895*) were detected, with 79% of accuracy rate. While in Series2, 15 participants (*M=.47, SD=.697*) were detected, with an accuracy rate of just 50%.

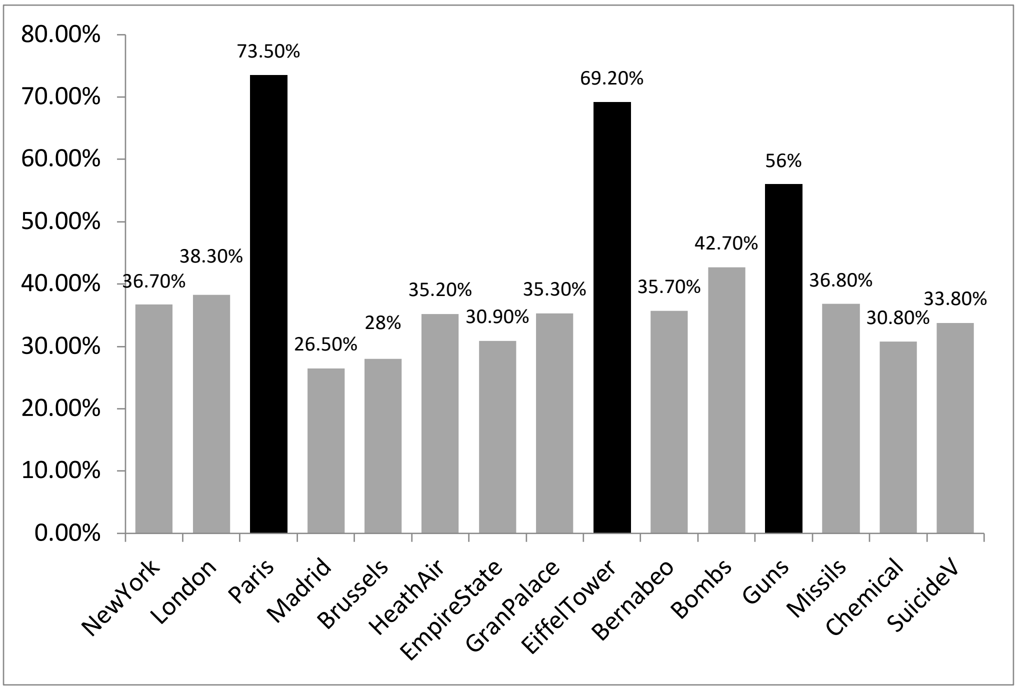

Figure 17. Physiological reactions' levels to all stimuli presented in both tasks from the experimental group

Through descriptive analysis, it was also examined the reaction levels of participants from the experimental group towards every relevant and irrelevant item presented during the CIT. Figure 17 shows that the highest physiological responses occurred to the relevant items (Paris for the City, the Eiffel Tower for the Location and Guns for the Weapon).

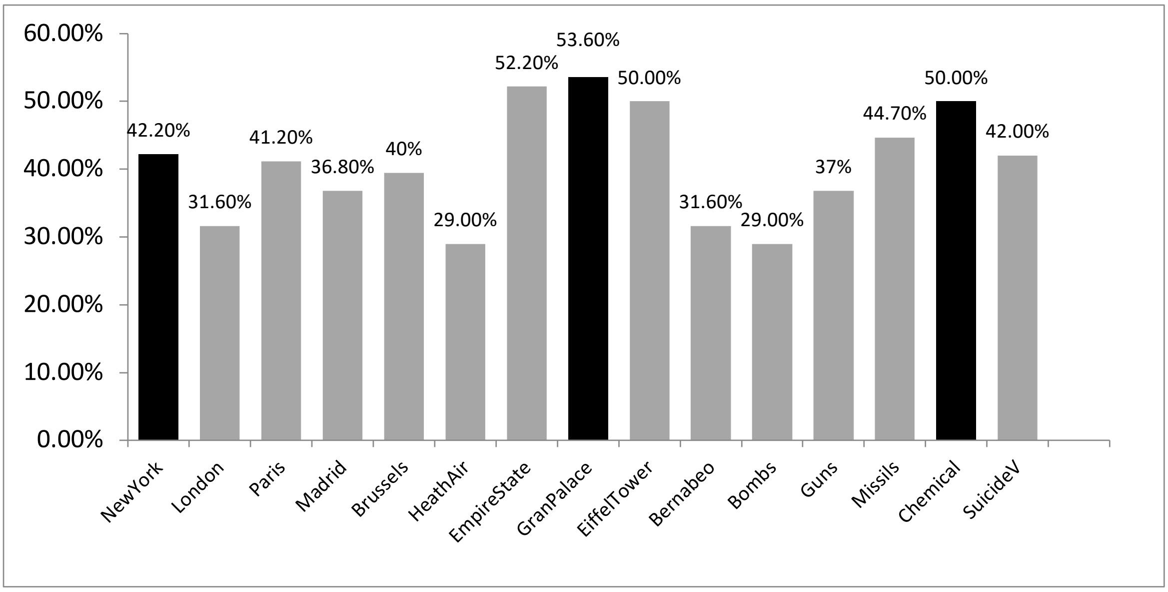

Figure 18. Physiological reactions' levels to all stimuli presented to the control group

Similarly, the same analysis was applied to the control group. The reaction levels of participants towards every relevant and irrelevant item presented during the CIT was measured. Figure 18 shows the highest physiological responses occurred to some irrelevant items (New York for the City, the Grand Palace for the Location and chemical weapons for the Weapon).

6.7.1.2 Inferential analysis

It is important, initially to determine if the data for this study is normally distributed by running a test for normality. In order to obtain this analysis, a single variable was created which included both experimental and control conditions. The analysis contained also a grouping variable, which specified the specific condition of each participant.

As the sample from Study III was relatively small, a Shapiro-Wilk test is the most appropriate and more powerful compared to the Kolmogorov-Smirnov test (Field, 2013). The analysis suggests that the overall level of deception from experimental and control group, D=476=.256, p<.005, was significantly non-normal. However, even if the data is not normally distributed (probably due to the difference between the control and the experimental group), it was suggested that it is possible to run ANOVAs, considered to be robust tests to deviations from normality (Maxwell & Delaney, 2004).

The assumption of homogeneity of variances can be assumed, as assessed by Levene's test for equality of variances, $F(11) = 1.263$ $p > 0.001$ Mauchly's test was non-significant for the main effect of the topics (p=.830); no results were shown for the main effect of the Series as there are only two levels.

A 2 (Group conditions: control and experimental group) by 3 (Topic: City, Location, Weapon) by 2 (Series: two series for each topic) Mixed Factorial ANOVA was conducted.

Factorial ANOVA revealed a significant main effect of group conditions, $F(1.305) = 16.597$, $p=.001$ $\eta^2=.052$. In fact, post-hoc tests with Bonferroni corrections applied for multiple comparison (a=.017) revealed a significant difference between control group (p=.001) and experimental group ($p=.001$). However, the analysis revealed a non-significant main effect of topics, $F(2.102) = 3.440$, $p=.36$, $\eta^2=.063$, which means that the deception levels were not significantly differed over the three topics. Finally, non-significant results were shown from the main effect of Series $F(1,51) = .468$, $p=.49$, $\eta^2=.009$. These results suggest that no habituation effect had significantly affected participants' physiological reactions during the test.

6.7.2 Follow up analysis conducted on SCR, RLL and Cardio Using Z-Scores

6.7.2.1 Descriptive data

The first analysis consisted in the classification obtained through the Lykken' s method (see Table 21), which shows that SCR was the measure that accurately detected more participants, compared to the RLL and Cardio. Also, from this method, it was possible to obtain deception rates for each measure, where the highest was obtained by the SCR, the second highest by the RLL and the lowest by the Cardio.

Table 22. *Detection rates computed for the three physiological measures and across measures in Study 2 for the experimental group using the Lykken's method*

Measures	n. of detection	n. of detection (control)	Percentage detected
RLL	25	15	58
SCR	29	17	67
Cardio	18	10	54

In a first step, the mean z-standardised response differences between relevant and irrelevant items were compared between the two experimental groups. Afterwards the mean of these measures was computed as an overall index of the differential responsivity in each physiological measure, expecting negative values for RLL and Cardio and a positive value for SCR as indication of knowledge of relevant information. As expected, the typical pattern of differential responses to relevant and irrelevant items was found for all the measures in the group of guilty participants, whereas these differences tended to zero for innocents (see Table 22).

Table 23. *Means and standard deviation of the standardised physiological response differences between relevant and irrelevant items within each experimental condition*

Measure	Guilty (34)		Innocent (19)	
	M	SD	M	SD
RLL	-1.41	0.16	-1.27	0.89
SCR	7.39	0.67	2.22	0.57
Cardio	-2.43	0.39	2.58	0.51

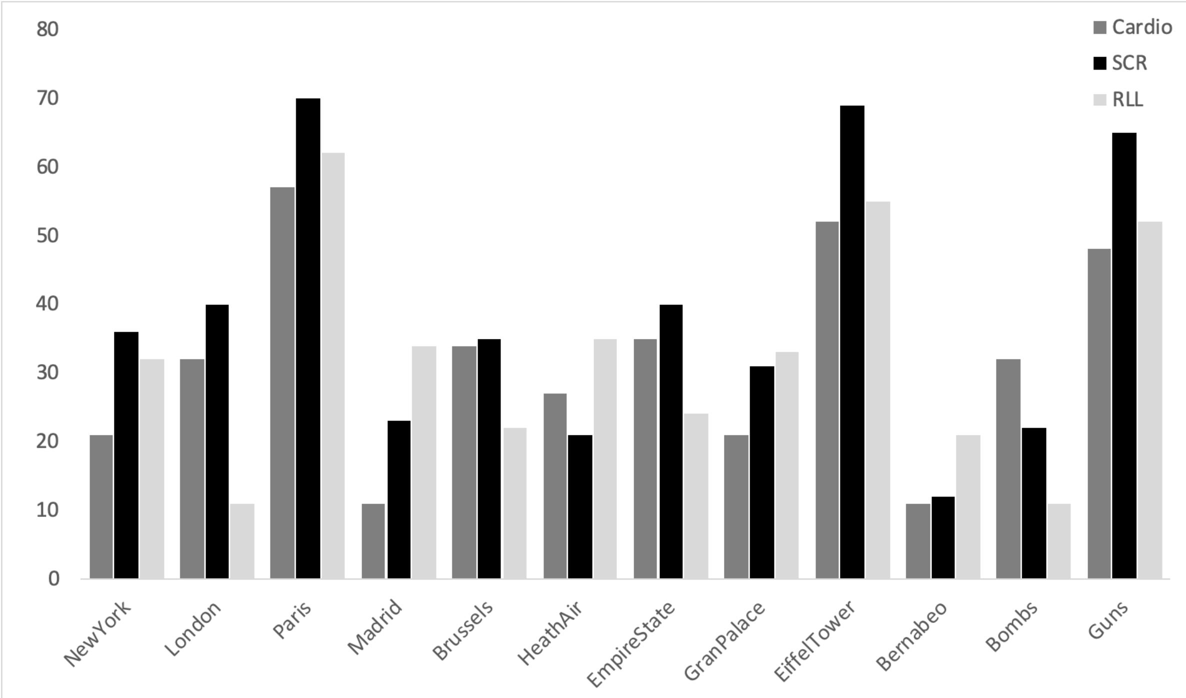

Figure19. Physiological reactions' levels to all stimuli presented to the experimental group

The reaction levels of participants from the experimental group towards every relevant and irrelevant item presented during the CIT was examined. Figure 19 shows that the highest physiological responses from SCR, RLL and Cardio occurred on the relevant items (Paris, Eiffel Tower, Guns).

The same analysis was applied to the control group. Figure 20 shows the reaction levels of participants towards every relevant and irrelevant item presented during the CIT was measured. As expected, the highest physiological responses occurred on some irrelevant items (for the RLL it occurred on the Bernabeo; for SCR it occurred on the Gran Palace; for Cardio it occurred on the Heathrow Airport)

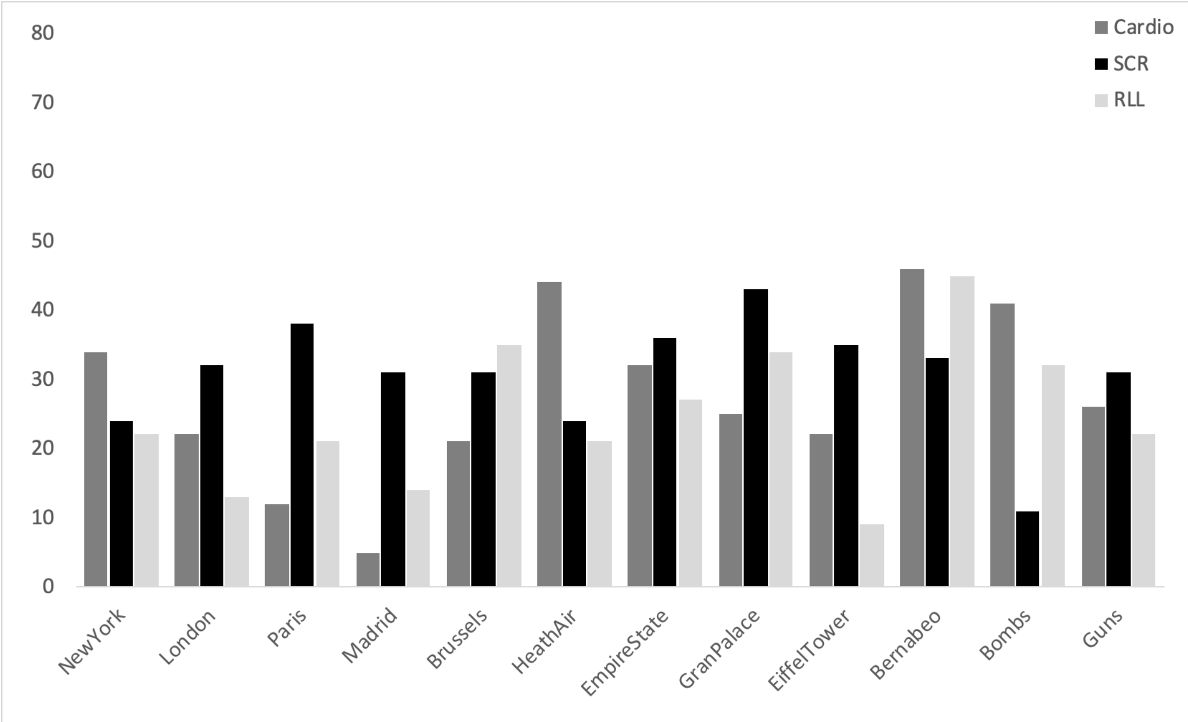

Figure20. Physiological reactions' levels to all stimuli presented to the control group

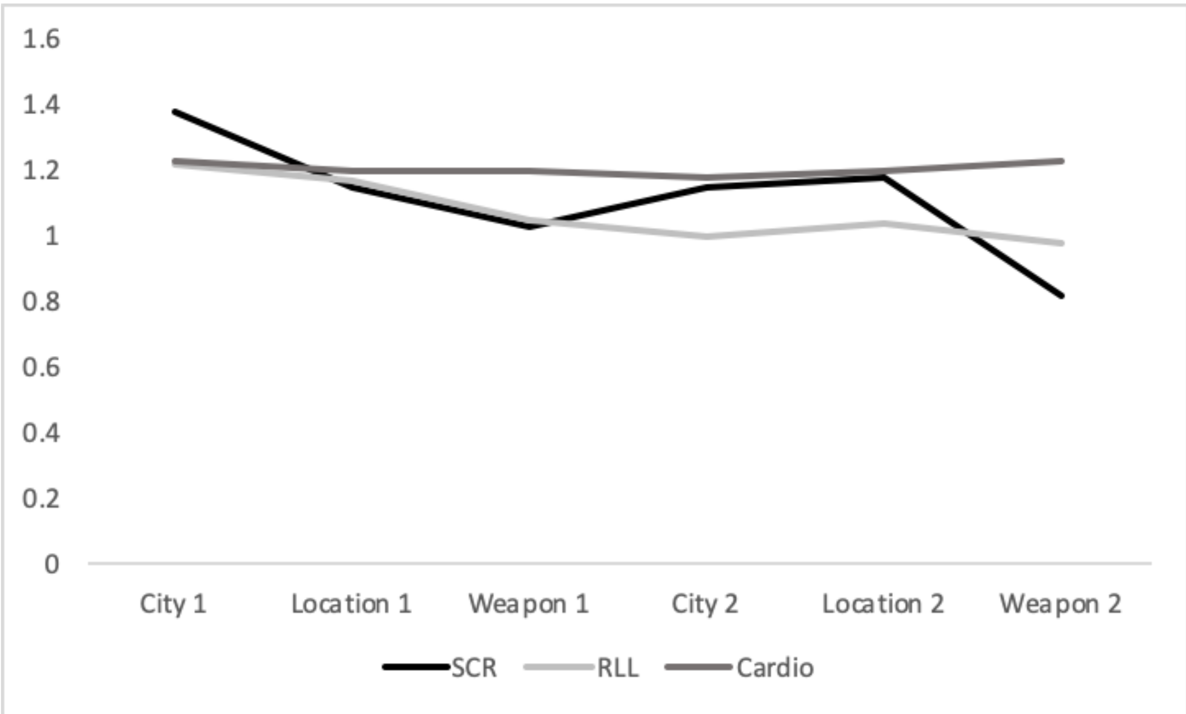

Figure 21. Means of the relevant items presented during the test from SCR, Cardio and RLL

From figure 21 it is possible to see how the means of each measure develop along the duration of the polygraph test. The figure suggests a clear decrease of the SCRS and RLL. The most dramatic decline is visible on the SCRS until Series 2 of the City, then it increases for the Series 2 of the Location, followed by a sharp decrease for the Series 2 of the Weapon. This suggests that the habituation effect affected SCR. However, as these are only preliminary results, the significance of this effect will need to be confirmed by inferential analysis. RLL levels presented a slight but continuous decrease along the duration of the test. Finally, Cardio levels appear relatively stable during the whole duration of the test.

6.7.2.2 Inferential analysis

It is important to determine if the data from this study is normally distributed by running a test for normality. In order to obtain this analysis, a single variable was created, which included both experimental and control conditions from all the physiological measures. The analysis contained a grouping variable, which specified the condition of each participant.
As the sample from Study III was relatively small, a Shapiro-Wilk test is the most appropriate and more powerful compared to the Kolmogorov-Smirnov test (Field, 2013). The analysis suggests that the overall level of deception from experimental and control group, $D=241=.52$, $p<.005$, was significantly non-normal. However, even if the data is not normally distributed (likely due to the difference between the control and the experimental group), it was suggested that it is possible to run ANOVAs, considered to be robust tests to deviations from normality (Maxwell & Delaney, 2004).

The assumption of homogeneity of variance can be assumed, as assessed by Levene's test for equality of variance, [*SCR* $F(1,51,)= 11.91$, $p>.005$; Cardio $F(1,51)=132.1$, $p>.005$; RLL $F(1,51)=.460$, $p>.005$]. Mauchly's test was no significant for the main effect of the topics [*SCR* ($p=.797$); *Cardio* ($p=.672$)); RLL ($p=.952$). No results were shown for the main effect of Time as there were only two levels.

A Mixed 2 (experimental-control) x 3 (City, Location, Weapon) x 2 (Series 1-Series 2) Factorial ANOVA was conducted for each measure (SCR, RLL and Cardio).

There was no a statistically significant main effect for the Topics on both RLL $F(2,102) = 2.161, p=.120, \eta^2=.041$ and Cardio $F(2,64)= .488, p=.616 , \eta^2=.016$; however, there was a main effect of task on SCR results $F(2,102) = 4.853, p=.010, \eta^2=.087$. This significance investigated further with post-hoc tests with Bonferroni corrections applied for multiple comparison (a=.017), showing a significant difference between the City topic and the Weapon topic ($p<0.001$).

There was no statistically significant main effect of Time for RLL $F(1,51) = .774, p=.383$ $\eta^2 =.15$ and Cardio $F(1,32) = 8.692, p=.65$ $\eta^2=.21$. But there was a statistical main effect of time for SCR $F(1,51) = .534, p<0.05(.022)$ $\eta^2=.015$.

Finally, ANOVA revealed a significant main effect of groups conditions for each measure [SCR $F(2.44) = .872, p=0.12;$RLL $F(1.63) = 5.11, p=0.23;$ Cardio $F(1.221) = 23.34, p=0.01$]. In fact, post-hoc tests revealed significant differences between control group (p<.001) and experimental group (p<.001) for each measure.

6.8 Discussion

6.8.1 Accuracy Rate

The aim of this third part of the study was to examine the validity of the CIT by using a mock crime scenario. For this study, as suggested by the previous literature (Vrij et al., 2013), it was expected that the accuracy rate of the test would be above the level of chance. This first research question has been confirmed, as all three measures used for this study detected levels of accuracy above the level of chance. SCR represents the measure with more detection (29 out of 34) while Cardio the one with less (18 out of 34). From the preliminary analysis conducted with the EDA, the highest accuracy rate was found in the first series of questions, with an 82% accuracy rate for the City Topic and 68% for the Location Topic, while Series 1 of the Weapon Topic produced the lowest detection rates at 62%. As previously observed in Study I and Study II, higher levels of deception for Series 1 of each measure might be explained by the fact that participants could present higher pre-existing levels at the beginning of the questionnaire.

Preliminary analysis found that 13% of false positives were detected from this study, which approaches the rates of the previous studies. The second analysis using z-scores confirmed these results. In fact, SCR was able to correctly detect 17 participants, RLL was able to detect 15 participants, while Cardio was able to detect 10. Therefore, SCR has the strongest detection rate for the control group, while Cardio has the smallest. The occurrence of a false positive effect could be caused by many factors, such as the influence of the habituation effect or by the relevance of the questions asked. Although both factors will need to be investigated more, the level of accuracy for the control group in this study represents the highest obtained so far in this study (88%). This rate approaches the 97.4% obtained by Elaad et al. (1992) and it is lower than the false positive levels obtained by other studies (Podlesny & Raskin, 1977; D. Raskin et al., 1976). In addition, through the initial analysis conducted, it was possible to see a clear significant difference between the control group and the experimental group in relation to the accuracy rate detected during the test. This was supported by the analysis of the Cohen's *d*, which reported a large effect size. Therefore, part of the aim of Study III was achieved.

The accuracy of the CIT was also confirmed by the analysis of participants' physiological reactions in relation to all the stimuli presented during the test. Results from all the measures showed that in the experimental group participants reacted more to the three relevant items (City=Paris; Location=The Eiffel Tower; Weapon=Gun), which confirms the principle that guilty participants will be able to recognize and discriminate relevant and familiar items from irrelevant and unfamiliar items (Bashore & Rapp, 1993; Lewis & Cuppari, 2009). In fact, when analysing the physiological responses of the control group results showed that participants reacted more to different control items, such as New York, the Grand Palace and Chemical Weapons.

From the overall index which indicates the differential responsivity of each measure, RLL and Cardio showed negative values, as both measures resulted in suppression and decrease of the physiological levels when stimulated by relevant items. Conversely, SCR's index produced a positive value, as this measurement increased when stimulated by relevant items. As expected, the typical pattern of differential responses to relevant and irrelevant items was found for all the measures in the group of guilty participants, whereas these differences tended to zero for innocents.

6.8.2 Tasks Differences

Another research question of this study focused on determining if the nature of the topics discussed by the participants could have affected their physiological responses in a different way. Preliminary analysis conducted on EDA showed that there was no statistically significant difference between the three topics. Interestingly, analysis conducted with the use of z-scores found a significant effect on the topics only for the SCR measure, showing a significant difference between the questions related to the City and Weapon. This result is not unexpected, as descriptive analysis from both preliminary analysis and z-scores showed higher levels of detection for the questions related to the City. Therefore, this topic must have affected participants' physiological reactions more, possibly due to the presentation order or to the significance of the question.

In addition, both results suggested that relevant questions received a relatively high percentage of reactions from the control group too (compared to Study I and II). This could be the consequence of having used as questions for the test elements that could be related to participants' personal experiences. Therefore, participants from the control group could have possessed personal associations with certain cities and locations enquired about during the test (Krapohl et al., 2006). Differently to this study, Studies I and II included neutral objects such as a hammer or a screwdriver; generally, it is less likely to have personal associations with generic objects like those used for the studies. Therefore, when using items that could trigger personal significance, even more caution needs to be applied, by verifying that participants do not perceive that information as familiar or relevant.

Levels of deception and accuracy rates in this study are the highest obtained so far in this study. The reason why these levels improved could be related to the nature of the priming stage, which is different from the one used in the previous studies. In fact, it is possible that a mock crime scenario was more impactful than completing a simple task; perhaps the psychological involvement of participants appears to be stronger in this study's design than in the previous two studies.

6.8.3 Habituation Effect

The present study tried to determine the occurrence of the habituation effect and its level of influence. Results from preliminary analysis suggested that that habituation effect did not significantly affect SCRs in this study. However, analysis conducted with z-scores found that

SCR was affected by the habituation effect. This suggests that although the number of repetitions was reduced in Study III, results were still affected by the habituation effect. As suggested by Kraphol (2009) the minimum number of repetitions for ensuring standardization to the test's results must be two. However, other critical factors might have promoted the occurrence of the habituation effect, such as the overall length of the test and the number of topics included within the test.

6.8.4 Limitations

This final section will discuss the limitations encountered in Study III, although a more detailed section regarding the limitations of the studies will be discussed in Chapter 8 (section 8.2).

Firstly, Study III presented a difference in the results regarding the significance of the habituation effect during the test and the significance of the topics. In fact, while preliminary results showed that none of these factors affected the outcomes, results with the use of z-scores suggested the opposite. This difference is due to the nature of the methodologies used, which present some differences in validity and accuracy that will be discussed in more detail in the section 8.2. However, results from the z-scores are also confirmed by descriptive data, which showed a change in the SCRs during the test.

Secondly, Study III presents a major limitation related to the nature of the questions. In fact, the physiological reactions from the control group to the relevant questions were higher, compared to Study III, in Studies II and I. This could be the consequence of having used as questions for the test elements that could be related to participants' personal experiences, such as popular cities and locations. In this case, it is more likely that participants could possess personal associations with certain elements enquired about during the test (Krapohl et al., 2006). Perhaps physiological rates were lower in Study I and II due to the use of neutral objects, such as a hammer or a screwdriver. Thus, participants were less likely to have personal associations with generic objects like those used for the first two studies. For this reason, more caution needs to be applied when using items that could trigger personal significance. This limitation, which represents the main cause for false positive effects, could be decreased by verifying whether participants perceive certain elements of the test as familiar or relevant and remove them from the questionnaire. This methodology will not be applied in the next experiment however, as results from the current study will need to be compared with results from Study IV. But, with the limited results obtained from descriptive data from this study, it is not possible to determine if this factor surely affected the results. Therefore, before applying major changes to the content of the questions, it would be useful to evaluate whether personal associations have a real impact on participants' physiological responses. This evaluation will be applied in the next experiment by asking participants to indicate whether they perceived those cities, locations and weapons presented in the test as familiar or relevant to them. This methodological addition will not avoid the influence of personal relevance in the next experiment; however, it will clarify the extent of its impact in the occurrence of false positive effects.

6.9 Conclusions

The purpose of this study was to evaluate the validity of the CIT through the application of a mock crime. The unique aspect of this study consists of the investigation of intention to conceal relevant information with the combination of visual stimulation.

The current study was able to demonstrate the validity of the CIT in different ways: the overall accuracy rate of the test was above the level of chance, with these results in line with previous studies. Similarly to the previous studies, Study III showed that SCR was the most accurate measurement compared to the others used for recording the results. In fact, it was able to detect correctly a greater number of participants with a higher detection rate. Conversely, the Cardio measurement was able to detect correctly the smallest number of participants with the lowest detection rate.

In conclusion, this study contributes to the present literature, which dedicated little attention to the evaluation of the CIT with the use of mock crimes and visual stimulation, and can therefore be considered a novel approach compared to the standard procedure of CIT examinations. Thus, these findings might help to improve preventive interviews with suspects that possess knowledge or involvement in a crime.

CHAPTER 7: STUDY IV

7.1 Background

Techniques such as the CIT were developed with the purpose of designing strategic tests that could help investigators to detect whether an examinee is deceitful or a truth teller (Vrij, 2008). Research around detecting deception through the application of the polygraph has focused on solo crime and individual examinations. Clearly, this is related to the fact that standard polygraph examinations involve interviewing offenders or potential suspects alone; individual interviews in a standard police interview setting are conducted in order to increase anxiety and avoid any attempt to plan responses regarding the crime (Kassin & Gudjonsson, 2004). However, group offending, such as gang offences, are common crimes (McGloin, Sullivan, Piquero, & Bacon, 2008) that are rising (Van Mastrigt & Farrington, 2011). Although, research on in-group offending has received little attention, with the research gap being even wider in the area of the application of the polygraph, and specifically the CIT, within group settings.

The necessity to investigate the validity of the CIT in this alternative setting is based on the theoretical concept that individuals are different within groups: Warr (2002) suggested that social dynamics in groups are extremely different, as this setting involves, for example, sharing of responsibilities and peer support. In line with this premise, research found that co-offenders present different characteristics from individual offenders and that they are usually younger than solo offenders (Van Mastrigt & Farrington, 2011). Research found that groups have a different perception of honesty and deception compared to individuals, which leads them to lie more (Cohen, Gunia, Kim-Jun, & Murnighan, 2009). This is explained by the *Fear and Greed Theory* (Cohen et al., 2009), which states that groups are more likely to strategically use honesty to maximize their own outcomes, as groups are greedier than individuals are. Similarly, groups are more likely to be distrusted, provoking a fear of groups. In addition, groups have minor concerns about the consequences of using deception for achieving their collective and/or their individual self-interests (Cohen et al., 2009).

Consequently, the variations highlighted make a difference when trying to explain criminal behaviour, strategies and motives applied to co-offenders compared to solo offenders. Therefore, this difference can have an impact on the dynamic of the interview process. Groups of criminals can be interviewed individually, by placing every member in different interview rooms, or collectively, usually with a simultaneous interrogation (Vernham, Granhag, & Mac Giolla, 2016). The application of one of these two interview settings allows investigators to gather different type of information; in fact, when interviewed alone, group members might allow investigators to examine within-group consistency. This aspect cannot be obtained in simultaneous interrogation however; instead, in this setting, it is possible to examine the communication and interactions between the members of the criminal group (Granhag, Rangmar, & Strömwall, 2015; Vrij et al., 2012).

The type of interview also determines the type of strategies applied by each group member; in fact, when questioned simultaneously, every member of the group loses some control over the other members' reactions and over what information is provided. Diversely, if interviewed alone, they need to decide whether they can trust the other members regarding the information they might decide to reveal (Vernham et al., 2016). This is one of the reasons why group situations are considered cognitively more demanding: in a typical interview situation, the examinee needs to consider what information to reveal but also what the other members of the group would say (Vernham et al., 2016). In support of this, Vrij et al. (2008) stated that increased cognitive demand generates more cues to deceit, therefore in a group setting the presence of deceitful cues might increase.

Interesting differences and similarities can also be found between groups of truth tellers and liars. According to Granhag and Hartwig (2008), when interviewing groups of suspects regarding any information about the crime that occurred or was planned, truth tellers and liars share the same common aim: trying to convince the interviewer about their innocence. Although researchers found that both groups apply different verbal strategies, they use the same non-verbal strategies, consisting of supressing nervous behaviours (Vrij, Mann, Leal, & Granhag, 2010).

Social theories found that one of the best ways to describe social dynamics in groups is by examining mechanisms of memory: *Collective Memory Theory* (Bartlett, 1932) touches many aspects of social dynamics within groups, such as collaborative learning, collaborative remembering and joint recall. Specifically, it examines how group members can recall information collectively. This concept is crucial for understanding the social nature of memory, as it suggests that previous experiences and past events are considered as memories created and shared with others (Rajaram, 2011).

As mentioned before, the way in which groups encode, store and then retrieve information in-group is extremely different from the processes of storing and recalling information individually. The *Transactive Theory* (Wegner, 1987) is one of the first approaches that tried to describe the mechanism of communication between members of a group. It suggests that members of the same group share common cognitions, which make them able to 'think together'; this process occurs by learning information about each other's domains of expertise. This mechanism results in a "common" memory, greater than the total of all the individual memories, which is property of the group (Wegner, 1987). The transactive memory system covers all stages involved in processing information:

- Encoding: in this stage, members of a group encode information about shared experience; responsibilities are divided and shared to all the members of the group. In this way, members know what information they need to remember individually and what information they need to know about the other members of the group.

- Storing: when information is stored, every member is able to remember the roles and responsibilities assigned to him/her and those assigned to the rest of the group.

- Retrieval: in this stage, the retrieval process is a social mechanism. In fact, the group is able to retrieve information together by interacting and communicating with one another. This helps to enhance their recollection (Wegner, 1987).

In line with this theory, the learning process of invented information occurs with the same mechanism. Deceitful groups invent, share and store fake information, creating a 'common fake memory'. The *Transactive Theory* helps to explain why the three stages are determinant for the deceitful group; diversely, members from the group should count on their individual cognitive abilities in order to create a story that matches with the other group members (Hintz, 1990). Clearly, this does limit the consistency of the group, as it will be more difficult to show coherence and spontaneous interaction between the members of the group. Diversely from truthful groups, deceitful groups when interrogated collectively do not need to interact or communicate for retrieving information, as they recall the invented story and provide answers to expected questions (Vrij et al., 2010). Finally, it found a difference between encoding, storing and recalling information together (collaborative groups) and encoding and storing information together but retrieving information individually (nominal groups) (Vernham et al., 2016). Results from a study show that usually, nominal groups recall significantly more information than collaborative groups, this is because when interviewing group members separately you avoid the occurrence of social inhibitions; therefore, individual interviews are considered better in terms of recalling information (Basden, Basden, Bryner, & Thomas, 1997).

At present, studies regarding deception involved only small groups (dyads and triads) (Granhag et al., 2015); furthermore, only a few studies tried to investigate the application of the CIT with groups (Meijer et al., 2010; Meijer et al., 2011; Meixner & Rosenfeld, 2011), leaving space for investigating alternative applications of the CIT in group settings.

7.2 Aim of the experiment

The aims of this last part of the study is to investigate the validity of the CIT with the use of a mock crime scenario in a group setting. Due to the occurrence of offending groups, such as terrorism, organized crime, drug trafficking and burglary, it is important to find the utility of the CIT in alternative settings that could fit other typology of crimes. As this area has not been extensively researched, this fourth exploratory study will try to investigate the accuracy of the test by proposing another area of practical application of the CIT. The unique aspect of this study is in applying the CIT within a group mock crime scenario with the use of visual stimulation, which has never been used before to the best of the current researcher's knowledge. Due to the positive results from Studies I, II and III, the addition of the visual stimulation will also be kept in a group-setting interview. In addition, cognitive and memory processes will be taken into consideration during the priming stage, as the intention of this study is also to identify if memory processes can have an impact on presenting greater physiological reactions.

The mock crime will have the function of priming the experimental group. However, through the task, it will also be possible to investigate whether some elements of the plan might have a different cognitive impact on the memory of participants, enhancing greater physiological responses. In addition, in order to enhance the impact of group memory, different "roles" related to specific aspects of the plan will be assigned to each member of the group. The impact of the roles will also be examined, trying to reveal which role was the most impactful.

7.2.1 Hypotheses

- Accuracy rates in the CIT will be above the level of chance
- A higher deception level in the experimental group will be found.
- A low false positive effect will be found.
- Low levels of habituation effect will be found.
- Greater physiological responses during the relevant questions will be found
- A various in the accuracy rate between the questions series will be found
- A difference between the City, the Location and the Weapon will be found.
- A difference between the roles assigned to the participants will be found.
- An increase in the accuracy rate will be found, compared to Study III.

7.3 Participants and Recruitment Process

This fourth exploratory study included 80 participants (66 females and 23 males). The sample was divided into two groups: an experimental group (*N = 40)* and a control group (*N= 40*). Ten groups were formed for the experimental group; each group included four members. Participants were randomly approached on the university campus. Undergraduate and postgraduate students and members of the university staff constituted the great majority of the sample. This study applied the same recruitment method used in the previous studies (see Chapter 3, section 3.1 for more information regarding recruitment procedure).

7.4 Equipment and materials

The physiological data for this experiment was recorded with the Lafayette LX5000 SW Polygraph System, owned by the department of Psychology at the University of Huddersfield. The equipment was purchased by the university from the official American Lafayette website (http://lafayettepolygraph.com/). Heart rate (HR), breath rate and EDA were recorded for this study. The mock crime was conducted with the use of pictures of the city, location and weapon chosen as relevant items. The pictures were carefully selected from the browser and inserted in a Power Point presentation. Due to the nature of this study's setting, a projector and a white screen were used, in order to visually stimulate participants through the Power Point presentation. The Power Point presentation was created in advanced by the researcher, which contained a series of pictures carefully chosen from the browser. (see Chapter 3, section 3.5).

7.5 Questionnaire Structure

The questionnaire used was the same used for the Study III. It contained 46 questions in total, with two series of questions for each crime area (6 series in total). Every series of question contained seven questions: one irrelevant question, one buffer question, four control questions and one relevant question. The City series was the first presented, followed by the Location and finally the Weapon series (see Appendix H).

7.6Procedure

The sample was distributed into an experimental and a control group. The experimental condition consisted of a discussion with the research assistant followed by the administration of CIT. The control group did not have any discussion. Participants were provided with full details of the study, including the implications of participating. Participants were encouraged to sign an informed decision, without any pressure or coercion, they were notified of their right to withdraw from the study at any point during the examination. Furthermore, they were informed of their right to have their information and data immediately removed and destroyed from the study. In addition, every participant needed to indicate any possible physical conditions that could have affected the results of the test.

7.6.1 STAGE ONE:

Experimental Group:

The structure of the priming stage of this study is as the one used for Study III (see Chapter 6, section 6.6).

Four participants formed every group. Participants were allocated to the group with a "first come first served" basis: ten discussion slots were made available, so participants could decide which day and time were the most suitable for them.

In order to create a psychological scenario in which participants could lie about something, they were asked to plan an attack. The assistant researcher directed the four participants to a classroom, in which they needed to sign the consent forms. During this initial stage, participants were provided with all the information they needed. They were instructed in participate to a group discussion with the assistant researcher. In order to the analysis of the results at group level, the three crime options were chosen by the research team, and then instructed to every participant from the experimental group. Although, the polygraph examination was conducted by a blind examiner (For more information regarding the relevant options chosen, see Chapter 3, section 3.5).

As participants did not know each other, it was important to create a "group memory", in which members of the group could feel comfortable enough to share that experience with the group. As suggested by the transactive theory (Wegner, 1987), it is important that members of the group share cognitions by knowing each other's memory expertise. In this way, every member can identify the others based on their roles and on the information they kept. Similarly, every member of the group can internalize his own role associated with the information assigned to him/her.

Therefore, in order to ensure effectiveness in the mock groups, the research group prepared five tickets, which specified the roles for each group and placed them in a little black bag; therefore, roles were assigned by a random draw conducted by participants.

Role 1 (the leader): The role of the leader had to lead the group discussion, trying to ensure that participants could discuss their own roles and the aspect of the plan associated with it.

Role 2: The second role involved the city, Paris. This member was required to provide suggestions regarding the city (considering possible ways to access the city; any advantages and disadvantages associated with Paris as potential city for an attack).

Role 3: The third role involved the target, the Eiffel Tower. This member was required to provide suggestions regarding the place in which the attack was supposed to occur (considering the best location for the attack; the type of target chosen for the attack; advantages and disadvantages associated with the Eiffel Tower as potential location for the attack).

Role 4: The fourth role involved the method of the attack, gun shooting. This member was required to provide suggestions regarding the weapon that would be used during the attack (considering the best type of guns, how to bring the weapons on the chosen location; advantages and disadvantages associated with the use of guns for the attack).

As in the Study III, participants were asked to plan the attack providing details and suggestions around the main elements of the plan. This process would help the members of the group to discuss details of the plan and consolidate the significance around the relevant options. All the roles were interlinked, which means that the leader had to lead the discussion and ensure that every aspect of the mock plan was covered during the discussion. Therefore, the other three members were invited to justifying the city of the attack, the location and the method of the attack. At the end of the priming stage, each member was instructed to write a summary of the plan as a final form of memory reinforcement.

As in the Study III, three images were shown to the group. The images pictured the three crime-related information (Paris, the Eiffel Tower, Guns). The same pictures were then shown in the PowerPoint presentation during the CIT examination. This procedure was conducted in order to increase the effectiveness of the priming process, making the recognition of the items stronger. Each group discussion lasted around 30 minutes.

Every member of the group was instructed not to reveal the relevant information discussed during the priming stage; rather, they were instructed to deny the knowledge of information related to the attack during the polygraph examination by answering “No” to all the questions.

Once the priming stage was concluded, every member of the group joined the polygraph researcher individually in the Psychology Lab for the administration of the polygraph. The order of the examination was: Leader, Role 2, Role 3 and Role 4.

The priming stage took approximately 10 minutes. Once the priming stage was concluded, every participant was instructed by the research assistant not to reveal the object used during the tasks. Rather, they were instructed to deny the use of the real object during the polygraph examination by answering “No” to all the questions. After this information, every participant joined the polygraph examiner in the Psychology Lab for the administration of the polygraph. In order to avoid any judgment bias during the examination and during the analysis of the data, a blind examiner conducted the polygraph examination.

Control Group

After having signed the consent forms, participants from the control group were instructed to enter the Psychology Lab for the administration of the polygraph. This group of participants was not primed with the discussion.

7.6.2 STAGE TWO

In the second stage of the experiment, participants from both experimental and control conditions were subjected to the administration of the polygraph. Although interviews with groups can be conducted also in groups (simultaneously), it was suggested that individual interviewing of groups is more effective as it is possible to examine the consistency of every member of the group and the recollection of the information is stronger, due to the absence of social inhibition (Basden et al., 1997; Granhag et al., 2015; Vernham et al., 2016). Therefore, for this study, every member of the groups was interviewed individually.

While the questions were verbally asked, a Microsoft PowerPoint presentation displayed the images related to each item in the question through a projector on a white board size screen. Every question lasted for the duration of the formulation of the question until the participant’s answer has been expressed. After the answer, the picture was followed by a blank slide, which lasted for 25 seconds; this amount of time corresponding to the period time in which the physiological levels need to come back to normality after the stimulation. It was ensured that the illumination level of the images was appropriate and that the participant was looking at the pictures during the examination (Iacono, 1991).

The test took approximately 20 minutes, plus 20 minutes for presentation and priming discussion and 5 minutes for debriefing (for more details regarding the general examination process, see Chapter 3, Section 3.5). The data was stored on the laptop through the Data Recording System.

As participants from the control group in Study III reacted significantly to the relevant items too, it was important to examine if personal significance for the items presented can be considered as a limitation for these studies. In fact, after the polygraph examination, participants were asked to fill a form in which they needed to indicate which of the items presented during the examination were relevant to them (e.g., which of the places displayed they visited before). This procedure was applied in accordance with the principle that it is important to verify that participants do not perceive certain information as more familiar or more relevant compared to others (Krapohl, et al., 2006).

7.7 Results

Participants (N=40) from Study II were divided into an experimental group (N=20 *M=1.28, 95% CI [.87,1.32], SD=.873*) and a control group (N=20 *M=.66, 95% CI [.65,.1.9], SD=.801*).

Effect sizes (Cohen's d) comparing the experimental and control groups averaged were computed. Cohen's d = (5.1142 - 0.003471)/0.436134 = 11.718256, showing a large effect size.

7.7.1 Preliminary analysis conducted on the EDA with the Lykken's approach

7.7.1.1 Descriptive data

Table 24. *Overall psychological responses using Lykken's Method*

Conditions	*Overall number of responses*	*Highest EDA*	*Second highest EDA*	*Lowest or zero EDA*
Experimental group City	80	67	13	13

Experimental group Location	80	58	22	22
Experimental group Weapon	80	48	32	32
Control group City	80	15	24	41
Control Group Location	80	13	16	42
Control group Weapon	80	9	18	53

Table 25. *Frequencies for the Terrorist Attack Plan from the experimental group*

Topic	*M*	*SD*	*Median*	*Mode*	*Variance*	*Sum*
City1	1.43	.813	2	2	.661	61
Location1	1.53	.751	1	2	.563	57
Weapon1	1.33	.859	1	2	.738	53
City2	1.50	.784	2	2	.615	60
Location2	1.15	.921	1.50	2	.849	46
Weapon2	.75	.899	0	0	.808	30

Table 26. *Percentages of detection rate among the series for the City, Location and Weapon topics in the experimental group*

Topic	*No Detection*	*Detection*

City1	15%	85%
Location1	20%	80%
Weapon1	25%	75%
City2	17%	83%
Location2	35%	65%
Weapon2	55%	45%

In regards to the City Topic, the analysis of frequencies allows for a possible comparison between the different levels of accuracy rates detected in the two series presented to the experimental group. In Series 1, the CIT was able to detect 34 participants out of 40 as deceptive (*M=1.43, SD=.813*), which accuracy rate accounted for 85%; in Series2, 33 participants (*M= 1.50, SD= .784*) were detected with an accuracy rate of 83%.

The same analysis of frequencies was conducted on the Location Topic. In Series1, the CIT was able to detect 30 participants out of 40 (*M=1.53, SD=.751*), which accuracy rate accounted for 80%; similarly, in Series 2, 33 participants (*M= 1.50, SD= .784*) were detected correctly with an accuracy rate of 83%.

Finally, for the Weapon Topic, 30 participants (*M=1.33, SD=.859*) were detected as deceptive, with 70% of accuracy rate. While in Series2, 18 participants (*M=.74, SD=.899*) were detected, with an accuracy rate of just 45%.

Table 27. *Frequencies for the Terrorist Attack Plan in the Control Group*

Topic	*M*	*SD*	*Median*	*Mode*	*Variance*	*Sum*
City1	.78	.909	0	0	.826	32
Location1	.49	.746	0	0	.556	20

Weapon1	.61	.862	0	0	.744	25
City2	.71	.716	1	0	.512	29
Location2	.93	.818	1	0	.670	38
Weapon2	.44	.673	0	0	.452	18

Table 28. Percentages of detection rate among the series for the City, Location and Weapon topics in the Control Group

Topic	***No Detection***	***Detection***
City1	54%	47%
Location1	66%	34%
Weapon1	64%	37%
City2	44%	56%
Location2	37%	63%
Weapon2	66%	35%

In regards to the City Topic, the analysis of frequencies allows for a possible comparison between the different levels of accuracy detected in the two series presented to the control group. In Series1, the CIT was able to detect 23 participants out of 40 (*M=.78, SD=.909*) as truth tellers, which accuracy rate accounted for 54%; in Series2, 18 participants (*M= .71, SD= .716*) were detected correctly with an accuracy rate of 44%.

The same analysis of frequencies was conducted on Location Topic. In Series1, the CIT was able to detect 27 participants out of 40 (*M=.49, SD=.746*), with 66% of accuracy rate; in Series2, 15 participants (*M= .93, SD= .818*) were detected as truth tellers with an accuracy rate of 37%.

In regards to the Weapon Topic, 26 participants (*M=.61, SD=.862*) were detected correctly, with 63% of accuracy rate. While in Series2, 27 participants (*M=.44, SD=.673*) were detected, with an accuracy rate of just 66%.

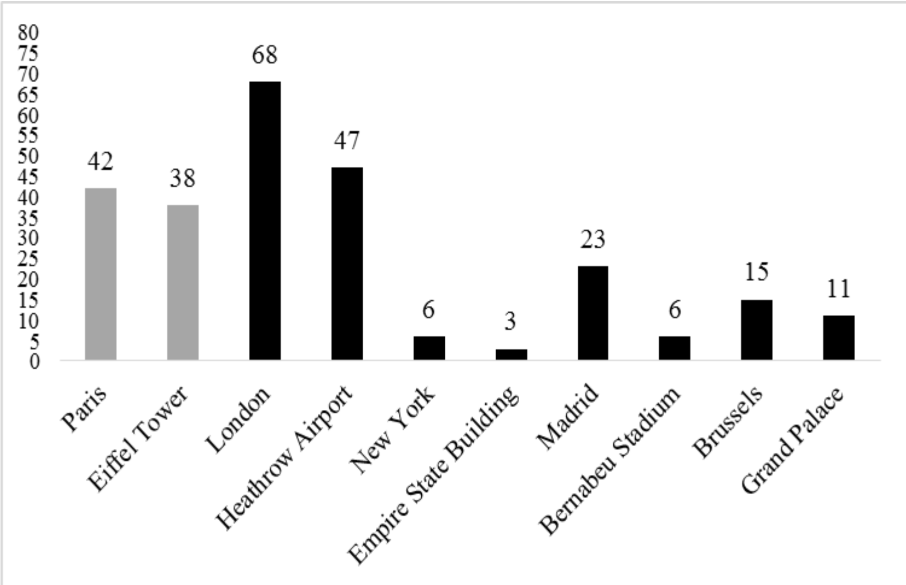

Figure 22. Personal significance with the items presented during the examination. The grey bars represent the relevant items, the black bars the control items.

In order to investigate if personal associations with the items presented could affect the results of the test, demographic information included a section were participants from both the experimental and control groups needed to specify if they have ever been in one of the places indicated. Descriptive data shown London and Heathrow as the most relevant places for the all sample.

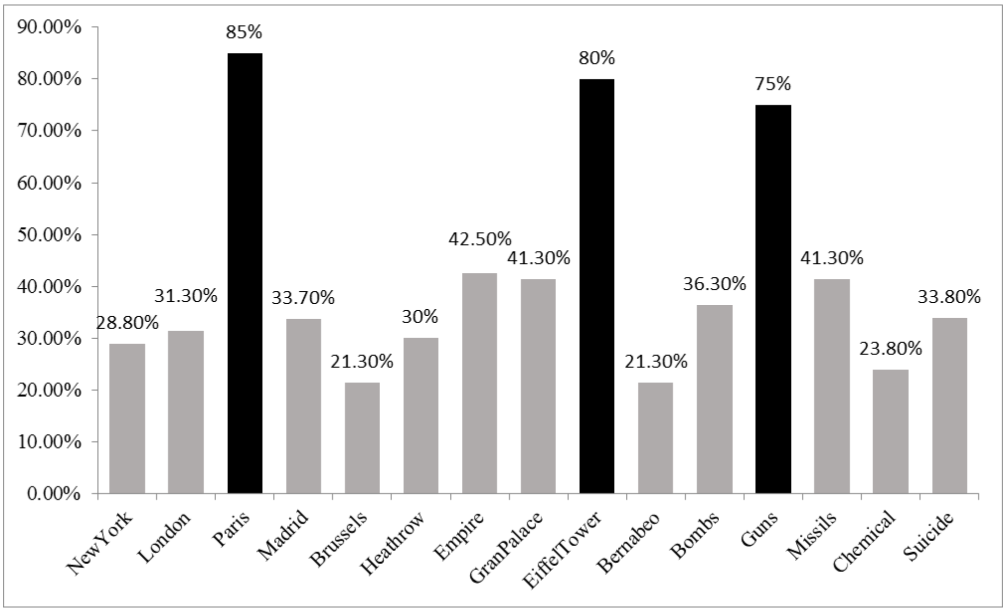

Figure 23. Physiological reactions' levels to all stimuli presented in both tasks from the experimental group

Through descriptive analysis, it was also examined the reaction levels of participants from the experimental group towards every relevant and irrelevant item presented during the CIT. Figure 23 shows that the highest physiological responses occurred to the relevant items (Paris for the City, the Eiffel Tower for the Location and Guns for the Weapon).

Similarly, the same analysis was applied to the control group. The reaction levels of participants towards every relevant and irrelevant item presented during the CIT was measured. Figure 24 shows the highest physiological responses occurred to some irrelevant items (New York for the City, the Grand Palace for the Location and Chemical weapons for the Weapon).

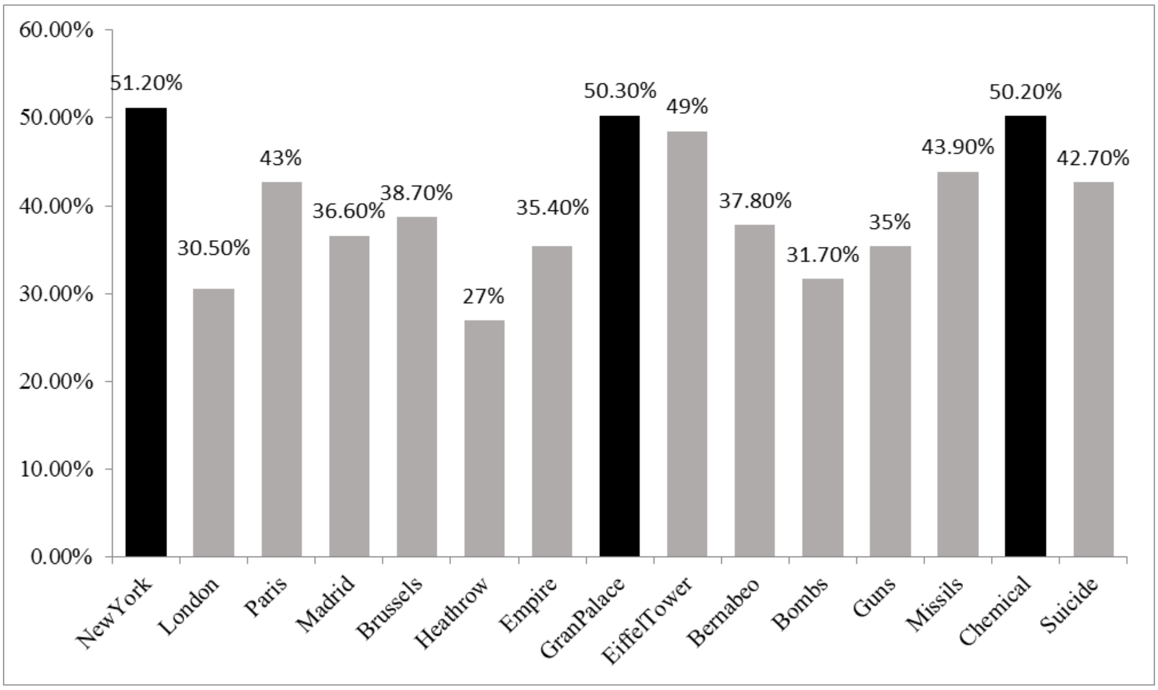

Figure24. Physiological reactions' levels to all stimuli presented to the control group

7.7.1.2 Inferential analysis

It is important, initially to determine if the data form this study is normally distributed, by running a test for normality. In order to obtain this analysis, a single variable was created which included both experimental and control conditions. The analysis contained also a grouping variable, which specified the conditions of each participant. The analysis suggests that the overall level of deception from experimental and control group, D=240=.339, p <.005, was significantly non-normal. However, even if the data is not normally distributed (probably due to the difference between the control and the experimental group), it was suggested that it is possible to run ANOVAs, considered to be robust tests to deviations from normality (Maxwell & Delaney, 2004).

The assumption of homogeneity of variances can be assumed, as assessed by Levene's test for equality of variances, *F(13) =1.345 p=.214*. Mauchly's test was no significant for the main effect of the topics (p=.981); no results were shown for the main effect of the Series as there are only two levels. A 2 (Group conditions: control and experimental group) by 3 (Topic: City, Location, Weapon) by 2 (Series: two series for each topic) Mixed Factorial ANOVA was conducted.

Factorial ANOVA revealed a significant main effect of group conditions, $F(1.79) = 47.123, p<.005\ \eta^2=.38$. Furthermore, post-hoc tests with Bonferroni corrections applied for multiple comparison (a=.017) revealed a significant difference between control group (p=.001) and experimental group (p=.001).

Analysis also revealed a significant main effect for the topics, $F(2.158) = 8.168, p<.005, \eta^2 =.094,$ which means that the deception levels were significantly differed over the three topics. Post-hoc tests with Bonferroni corrections applied for multiple comparisons *(a=.017)* revealed that participants reacted differently to the three topics: City *(p=.001),* Location *(p=.002)* and Weapon *(p=001).*

Finally, non-significant results were shown from the main effect of Series $F(1.79) =4.257, p=.042, \eta^2=.051.$ These results suggest that no habituation effect had significantly affected participants' physiological reactions during the test.

The last part of the analysis is focused on determine if the roles that participants played during the discussion significantly affected their physiological reactions, by reacting more to the questions that referred more to their roles. The most appropriate test for this analysis is the Pearson's Chi Square. Table 18 shows the results from the analysis, which indicate that there was not significant relationship between role and the questions referring to their specific topic.

Table 29. *Chi Square analysis of the relationship between roles to significant answer options*

	χ2 Value	**df**	**p-value**
Leader	5.928	4	.205
CityRole	2.200	4	.699
LocationRole	4.350	4	.368
WeaponRole	9.894	4	.042

7.7.2 Follow up analysis conducted on SCR, RLL and Cardio Using Z-Scores

7.7.2.1 Descriptive data

The first analysis consisted in the classification obtained through the Lykken's method (see Table 29), which shows that SCR was the measure that accurately detected more participants, compared to the RLL and Cardio. Also, from this method, it was possible to obtain deception rates for each measure, where the highest was obtained by the SCR, the second highest by the RLL and the lowest by the Cardio.

Table 30. *Detection rates computed for the three physiological measures and across measures in Study 2 for the experimental group using the Lykken's method*

Measures	n. of detection	n. of detection (control)	Percentage detected
RLL	14	13	67
SCR	17	16	75
Cardio	13	13	62

In a first step, the mean z-standardised response differences between relevant and irrelevant items were compared between the two experimental groups. Afterwards the mean of these measures was computed as an overall index of the differential responsivity in each physiological measure, expecting negative values for RLL and Cardio and a positive value for SCR as indication of knowledge of relevant information. As expected, the typical pattern of differential responses to relevant and irrelevant items was found for all the measures in the group of guilty participants, whereas these differences tended to zero for innocents (see Table 30).

Table 31. Means and standard deviation of the standardised physiological response differences between relevant and irrelevant items within each experimental condition

Measure	Guilty (20)		Innocent (20)	
	M	SD	M	SD
RLL	-0.73	0.12	-1.27	0.89
SCR	1.15	0.69	2.22	0.57
Cardio	-0.1	0.93	2.58	0.51

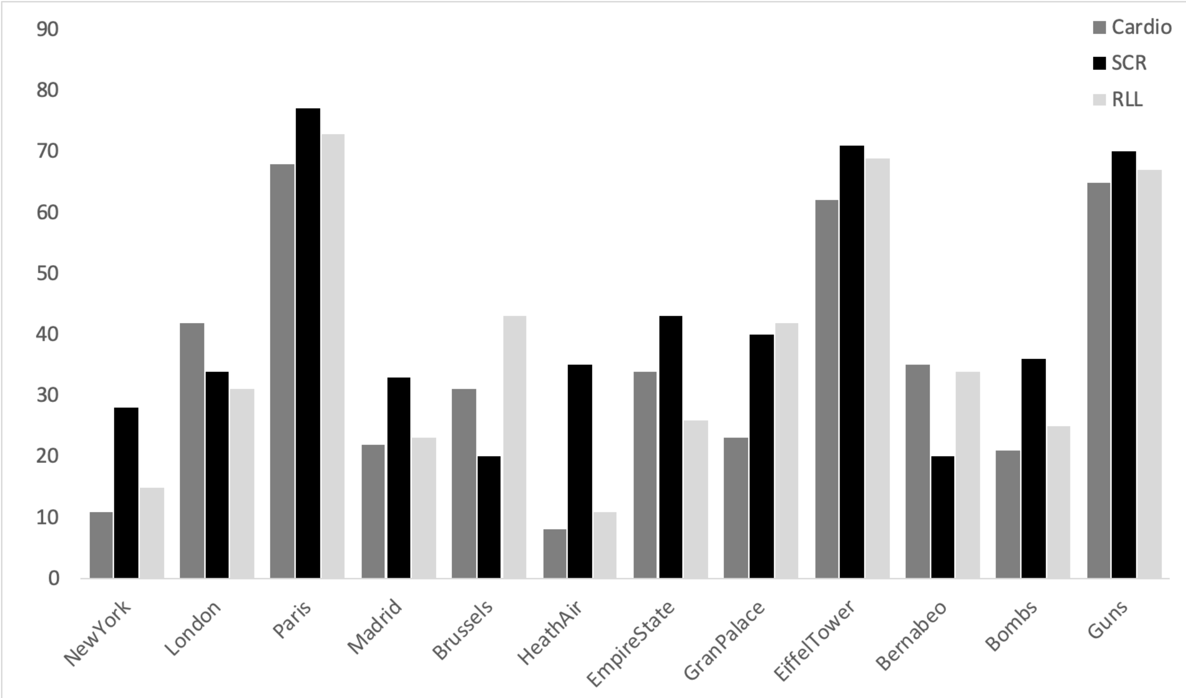

Figure 25. Physiological reactions' levels to all stimuli presented to the experimental group

The reaction levels of participants from the experimental group towards every relevant and irrelevant item presented during the CIT was examined. Figure 25 shows that the highest physiological responses from SCR, RLL and Cardio occurred on the relevant items (Paris, Eiffel Tower, Guns).

The same analysis was applied to the control group. Figure 26 shows the reaction levels of participants towards every relevant and irrelevant item presented during the CIT was measured. As expected, the highest physiological responses occurred on some irrelevant items (for the RLL it occurred on the Empire State building; for SCR it occurred on Brussels; for Cardio it occurred on London)

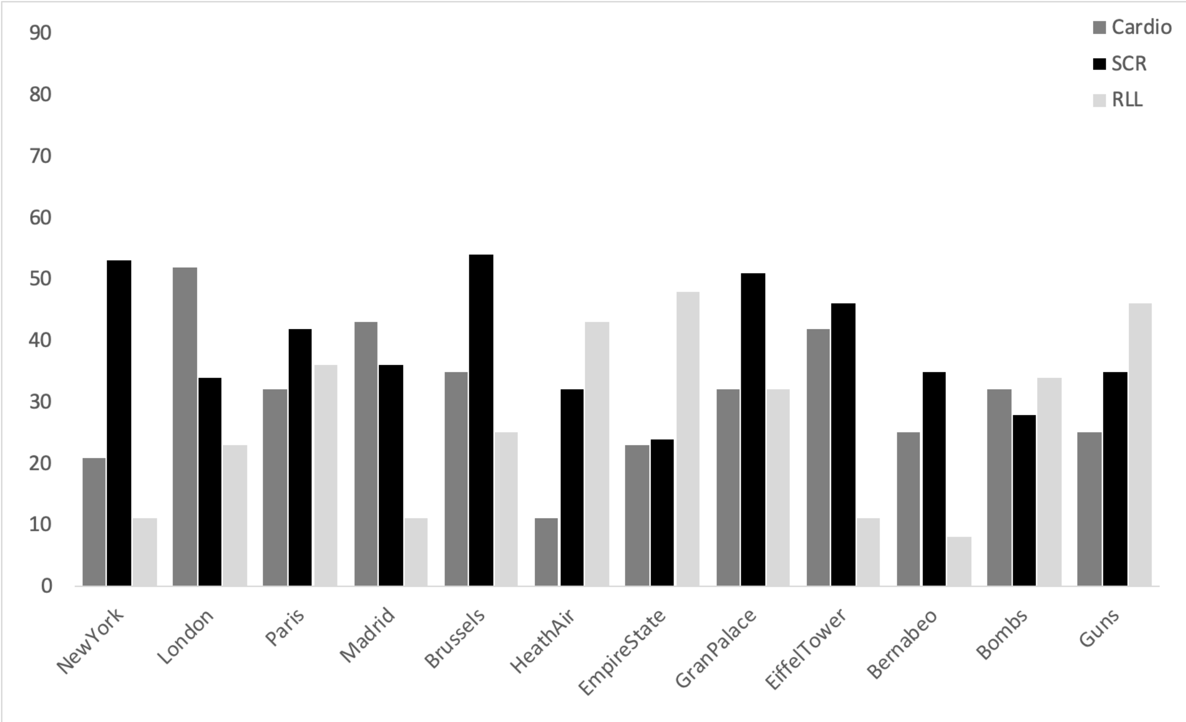

Figure26. Physiological reactions' levels to all stimuli presented to the control group

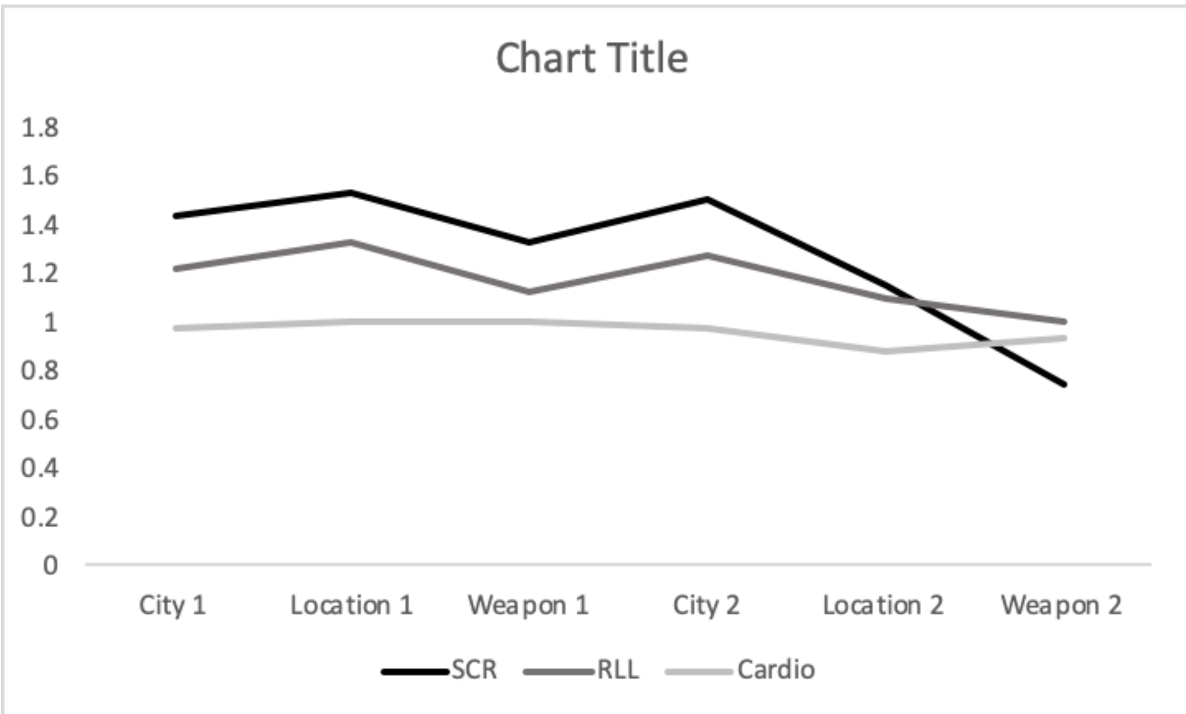

Figure 27. Means of the relevant items presented during the test from SCR, Cardio and RLL

From figure 27 it is possible to see how the means of each measure develop along the duration of the polygraph test. The figure suggests a clear decrease of the SCR and RLL. The most dramatic decline is visible on the SCR for Series 1 of the Weapon, then it increases for the Series 2 of the City, followed by a sharp decrease for the Series 2 of the Weapon. This suggests that the habituation effect affected SCR. However, as these are only preliminary results, the significance of this effect will need to be confirmed by inferential analysis. RLL levels presented a similar movement of the SCR, but smoother. Finally, Cardio levels appear relatively stable during the whole duration of the test.

7.7.2.2 Inferential analysis

It is important, initially to determine if the data form this study is normally distributed, by running a test for normality. In order to obtain this analysis, a single variable was created which included both experimental and control conditions. The analysis contained also a grouping variable, which specified the conditions of each participant. The analysis suggests that the overall level of deception from experimental and control group, $D=648=.42, p<.005$, was significantly non-normal. However, even if the data is not normally distributed (likely due to the difference between the control and the experimental group), it was suggested that it is possible to run ANOVAs, considered to be robust tests to deviations from normality (Maxwell & Delaney, 2004).

The assumption of homogeneity of variance can be assumed, as assessed by Levene's test for equality of variance, [*SCR* $F(1,51)= 25.9$, $p>.05$; *Cardio* $F(1,51)=.455$, $p>.05$; *RLL* $F(1,51)=.219$, $p>.005$]. Mauchly's test was no significant for the main effect of the topics [*SCR* $(p=.878)$; RLL $(p=.655)$; Cardio $(p=.920)$. No results were shown for the main effect of Time as there were only two levels.
A Mixed 2 (guilty-innocent) x 3 (City, Location, Weapon) x 2 (Series 1-Series 2) Factorial ANOVA was conducted for each measure (SCR, RLL and Cardio).

There was a statistically significant main effect for the Topics on every measure [SCR $F(1,102) = .532, p=.002, \eta^2=.062$; RLL $F(2,102) = 3.92\ p=.023, \eta^2=.071$; Cardio $F(2,102)$

$= 3.496, p=.034, \eta^2=.064$. This significance was investigated further with post-hoc tests with Bonferroni corrections applied for multiple comparison (a=.017) for every measure showing a significant difference between the City topic and the Weapon topic ($p<0.001$).

There was no statistically significant main effect of Time for any of the measures [SCR $F(1,51) = .378, p=.541\ \eta^2=.007$; RLL $F(1,51) = 7.963, p=(0.74)\ \eta^2=.13$. , Cardio F$(1,51) = .927, p=.340, \eta^2=.018$. Therefore, the habituation effect did not significantly affect the results of this study.

Finally, ANOVA revealed a significant main effect of groups conditions for each measure [SCR $F(1.70) = .356, p=012;$RLL $F(2.167) =.687, p=0.01;$ Cardio $F(2.13) = 1.346, p=0.002$]. In fact, post-hoc tests revealed significant differences between control group (p<.001) and experimental group (p<.001) for each measure.

7.8 Discussion

7.8.1 Accuracy Rate

The aim of this fourth part of the study was to examine the validity of the CIT by using a mock crime scenario. For this study, as suggested by the previous literature (Vrij et al., 2013), it was expected that the accuracy rate of the test would be above the level of chance. The different accuracy rates of Study IV represent the highest obtained so far compared to Studies I, II and III. This first research question has been confirmed, as all three measures used for this study detected levels of accuracy above the level of chance. SCR represents the measure with more detection (17 out of 20), while Cardio the one with less (13 out of 20). From the preliminary analysis conducted with SCR, the highest accuracy rate was found in the first series of questions, with an 85% accuracy rate for the City Topic and 80% for the Location Topic, while Series 3 of the Weapon Topic produced the lowest detection rates, with 45%. As previously observed in the previous studies, higher levels of deception for Series 1 of each measure might be explained by the fact that participants could present higher pre-existing levels at the beginning of the questionnaire.

The accuracy of the CIT was also confirmed by the analysis of participants' physiological reactions in relation to all the stimuli presented during the test. Results from all the measures showed that in the experimental group participants reacted more to the two relevant items (City=Paris; Location=The Eiffel Tower; Weapon=Gun), which confirms the previous results. These results are in line with previous studies which supported the idea that guilty participants will be able to recognize and discriminate relevant and familiar items among irrelevant and unfamiliar items (Bashore & Rapp, 1993; Lewis & Cuppari, 2009). In the same way, the physiological responses of the control group were analysed; results showed that participants reacted more to the control items, such as New York, the Grand Palace and Chemical Weapons.

Preliminary analysis found that 18% of false positives were detected from this study, which was higher compared to the previous studies. The second analysis using z-scores confirmed these results. In fact, SCR was able to correctly detect 16 participants, while RLL and Cardio were able to detect 13 participants. Therefore, SCR has the strongest detection rate for the control group. The occurrence of a false positive effect in this study could be caused by the same factors that affected Study III, such as the influence of the habituation effect and, more particularly, by the relevance of the questions asked. For this reason, the procedure of this study included an evaluation of possible personal associations to the items for both experimental and control groups. Results showed that most of the participants indicated London and Heathrow Airport as the most visited place. Therefore, if personal relevance should determine higher physiological reactions, results from the experimental group should have shown higher levels of deception for London and Heathrow Airport. Conversely, analysis from this experiment showed that the CIT was able to detect 75% of the sample's physiological reactions associated with the correct answers (Paris and the Eiffel Tower). In addition, through the initial analysis conducted, it was possible to see a clear significant difference between the control group and the experimental group in relation to the accuracy rate detected during the test. This was supported by the analysis of the Cohen's *d*, which reported a large effect size. Therefore, part of the aim of Study IVas achieved.

From the overall index which indicates the differential responsivity of each measure, RLL and Cardio showed negative values, as both measures resulted in suppression and decrease of the physiological levels when stimulated by relevant items. Conversely, SCR's index produced a positive value, as this measurement increased when stimulated by relevant items.

As expected, the typical pattern of differential responses to relevant and irrelevant items was found for all the measures in the group of guilty participants, whereas these differences tended to zero for innocents.

7.8.2 Tasks differences

The second part of the analysis tried to determine if the nature of the topics could have affected participants' physiological reactions in a different way. Preliminary analysis conducted on EDA showed that there was a statistically significant effect on the three topics. According to the results, the City Topic was the one to which participants reacted more. In fact, during both series of questions related to the City Topic, the test performed the highest accuracy rates (Series 1= 85%; Series 2=83%); in other words, participants' physiological reactions were stronger during the questions related to the city of the attack. These results were confirmed by inferential analysis conducted on the z-scores, which also found an effect on the topics in the other two, RLL and Cardio. However, the main significant difference was found between the City and the Weapon Topic. This confirms preliminary results on SCR, reflecting the same outcome also for RLL and Cardio. Therefore, questions about the City have affected more of the participants, compared to those questions related to the Weapon. Perhaps the contents discussed during the group discussion about the relevant items (Paris) had a greater impact on participants, compared to the other topics. Therefore, names and images of the cities might have been simply more salient and significant for the participants than names of the countries and weapons. In support of this explanation, results from a similar study conducted with groups demonstrated the same conclusions (Maijer, Bente, Ben-Shakhar & Shumacher, 2013).

7.8.3 Habituation Effect

The present study tried to determine the occurrence of the habituation effect and its level of influence. Both preliminary results and analysis conducted with the z-scores showed that no habituation effect affected the results of this study, which indicates that the relevance for the items was constantly effective. Due to the reduction of the repetitions applied since Study III, it was expected to find reduced levels of the habituation effect. Therefore, this confirms that the repetitions of the series of questions should be kept to a minimum for ensuring standardisation.

7.8.4 In-Group Setting

The aim of this experiment was to examine whether the CIT was able to detect intentions of undertaking a mock attack in a group setting. Both preliminary results and analysis conducted with z-scores showed that levels of accuracy of the CIT were higher in this study. Therefore, this improvement could be associated with the nature of the priming stage, which involved a group discussion. In fact, in line with previous studies (Hintz, 1990; Wegner, 1987), information encoded and stored in groups is stronger than when processed individually. Although the learning process of creating fake information occurs in the same way as when doing it individually, deceitful groups create a common fake memory (Vernham et al., 2016). In the case of this study, the fake memory is represented by the criminal plan discussed together. Perhaps the awareness of this experience allowed participants from the mock crime to feel, even if temporarily, part of a unique shared experience. The unique 'common memory' created during the discussion was reinforced by the division of roles, which increased the knowledge of the group about the general information related to the plan and about the single members' duties (Wegner, 1987). In addition, the accuracy rate could also have increased because members of the group were examined individually. In fact, it was demonstrated that group situations are cognitively more demanding, which consequently generates more cues to deceit (Vrij et al., 2010). In addition, the examinee needs to take into consideration not only what to reveal during the test, but also what the other members of the group could decide to reveal. All these dynamics increased the presence of deceitful elements in the group setting.

The study also examined any possible relevance of roles for each member of the mock group. Results showed that there was no correlation between the relevant options presented and the participant role. This means members of the group reacted equally to all the relevant items presented during the examination, independently by the role assigned. This demonstrates that the test does not need to examine a key group figure for detecting relevant information. Therefore, the CIT is able to achieve the same levels of accuracy by also interviewing those members of the group with less or indirect involvement with the plan.

In conclusion, the CIT was able to detect participants' intentions to plan a terrorist attack also from a group setting, which is in line with previous studies (E. H. Meijer et al., 2010; Meixner & Rosenfeld, 2011). Therefore, this study demonstrates a real practical application of the CIT as a forensic tool in preventing crime with individual offenders and criminal groups.

7.8.5 Limitations

This final section will discuss the limitations encountered in Study IV, although a more detailed section regarding the limitations of the studies will be discussed in Chapter 8 (section 8.2).

Study IV presents a major limitation related to the nature of the questions. Analysis showed that the questions related to the "City" trigged higher physiological responses compared to the other topics. This was also confirmed by inferential analysis with both types of analysis and on all three physiological measures. Similarly to Study III, this result could be the product of personal associations. For this reason, the present study administrated a post-test questionnaire, to investigate whether participants had personal associations with one or more items of the test. Although results found no clear support for this limitation, participants from the control group indicated some cities and locations to be relevant to them. Particularly, they expressed higher familiarity with the "City" topic. Although personal associations did not strongly affect the results of this study, this factor might be problematic for future studies, causing false positive rates. Therefore, this element should be taken into consideration for future experiments.

7.9 Conclusions

The purpose of this study was to evaluate the accuracy of the CIT through the application of a mock crime. The unique aspect of this study was the investigation of participants' involvement in planning a crime in a group setting. Study IV was able to demonstrate the validity of the CIT in different ways: the overall accuracy rate of the test was above the level of chance, which results are in line with previous studies. In order to determine if the group setting increased the accuracy rate of the polygraph, results from this study were compared to results from Study III. As hypothesised, in Study IV the accuracy of the CIT slightly increased, perhaps due to the influence of the group setting. Study IV also confirmed that the SCR was the most accurate measurements compared to the others used for recording the results; while Cardio was the less accurate.

In conclusion, results from Study IV have important implications for the utilization of the test a preventing tool for criminal groups. The combination of a group setting and the application of visual stimulation can be considered as an innovative addition to the standard setting of CIT examinations. Therefore, further studies should develop the results from this research more in order to improve the effectiveness of the CIT and its practical applicability in real contexts.

CHAPTER 8: CONCLUSIONS

8.1 Summary of the results

The current research investigated a number of aspects related to the validity of the Concealed Information Test (CIT). An attempt to provide a contribution to the understanding of the CIT mechanisms and demonstrate its potential applicability was conducted. Specifically, to achieve these aims, a series of four studies was employed that evaluated different aspects of the validity of the CIT.

In Study I, which is the first exploratory study, the participants' physiological responses were analysed with the application of a standard CIT examination. The analysis of the results demonstrated that the CIT was able to detect relevant information from guilty participants with an accuracy rate of 63/65% and correctly detect innocent participants with an accuracy of 86%. Different factors caused the occurrence of false positive and false negative rates, such as the influence of the habituation effect and the nature of the tasks from the priming stage. Thus, this study demonstrated that elementary factors like a strong sound could affect memory and the recollection of information. Therefore, the characteristics of the information that are encoded and stored could determine the process of recollection. Although a few limitations affected the results (presented in section 8.2), the CIT was able to correctly detect relevant information from participants in this study.

Study II tried to follow up some of the conclusions obtained from Study I. However, an additional methodological element was inserted in this new study: visual stimulation. This study demonstrated that additional visual stimulation increased the accuracy levels of the test. These conclusions propose that the combination of verbal and visual stimulation together might have a stronger impact on participants' physiological responses. The process of recognition and recollection of information during a CIT is perhaps stronger when stimulated visually. Although the results are not optimal due to the occurrence of false positive and false negative rates, this study proposed a methodological setting that had not been used before in similar CIT studies.

Study III was a development of the previous studies that retained some of the methodological elements from the previous designs, such as visual stimulation, the procedure and the administration of the CIT. However, the aim of this study was to apply all the findings from the previous studies (Studies I and II) in a more realistic forensic setting, by using a mock crime. Results showed that the CIT was able to detect concealed information with a 69% accuracy rate. Conclusions from this study showed that the CIT was able to detect crime-related information from a more realistic experimental context. In addition, the accuracy of the CIT increased compared to the previous studies. This suggests that the idea of being involved in a criminal act, even if in a controlled setting, might have a stronger impact on participants' memory.

Finally, Study IV applied all the conclusions from Studies I, II and III in a different design. Specifically, this study retained the methodological elements of Study III, but applied them in a group setting. The purpose of this study was to examine if crime-related information could be detected from a group of mock suspects that planned a mock crime together. The results that were compared to Study III demonstrated that participants from this study reacted more physiologically while trying to conceal the criminal information. This suggests that the increase might have been caused by the group setting, which could have affected the process of information encoding and recollection of the information. In addition, the group setting could have increased the motivation for lying, showing higher psychological reactions to the relevant items.

As presented in section 2.3.4, the phenomenon of attentional narrowing is extremely useful when trying to determine the effectiveness of certain information. In fact, it was suggested that attentional narrowing causes a superior encoding of central details, and undermines the encoding of peripheral details (Meijer et al., 2011) (for more information about memory functioning, see section 2.3.4). For this reason, the studies in this thesis tried to formulate the CITs based on the central details of the priming task (e.g. the object used for completing the task, the city chosen for the attack, etc). Results from this thesis suggest that emotional arousal caused an attentional narrowing, as in most of the studies it was apparent there was a clear difference between how guilty participants reacted physiologically to central items compared to irrelevant ones. Results from the four studies presented via graphs and descriptive data supported these conclusions: relevant items, such as the balloon (in Studies I and II) or the city of the attack (in Studies III and IV), clearly presented higher physiological responses compared to irrelevant items, such as the pin, the screwdriver or the city of London. These results are supported by other studies (e.g. Carmel et al., 2003; Gamer et al., 2010), which provided similar conclusions, suggesting that the CITs of the present thesis were effective.

In addition, these optimal results were obtained through the role of the "enactment effect" (Engelkamp, 1998). As discussed in section 3.4, Study II and Study III were developed following the principle of enactment effect, which has been demonstrated to increase the effect of storing and recollection processes. In fact, the first two studies of this thesis contained two action tasks. They involved the use of physical movements and gestures to improve the quantity of phrases that could be recalled, and increase the length of time for which they could be stored. Thus the completion of the tasks should have made accessing and processing this information easier (Engelkamp & Krumnacker, 1980) (for more information, see section 3.4). As this was not part of the aims of this thesis, the role of the enactment effect was not investigated. However, the purpose of the action tasks was to provide the most efficient priming task, allowing a systematic processing of the central information. As the results from these studies were positive, we can assume that the tasks were effectively developed, producing strong physiological reactions.

Overall, the CIT was able to detect concealed information from examinees with a moderate level of accuracy. Although some of the results were not optimal, this study contributed to a deeper understanding of the CIT mechanisms and its potential applications. Various factors limited the full accuracy rate in this study, causing the occurrence of false positive and false negative rates. Those factors were generally common to all studies (I, II, III and IV). A detailed discussion is provided in the limitations section.

8.2 Limitations

A number of limitations were identified while conducting the previous studies. Specifically, in the current section limitations related to methodological issues, false positive, external validity, countermeasures, examiner bias, and habituation effect are discussed.

8.2.1 Methodological Limitations

When administrating the polygraph, verbal and visual stimulation were provided to the participant until they could answer the question. After every participant's answer, a fixed interval of 15 seconds occurred. This mechanism was put in place to allow the examinee's physiological levels to come back to a baseline level after the previous response. Perhaps, after a few questions, some of the participants could have become aware of the interval between the questions. Consequently, deceptive participants could have tried to physiologically control themselves before the beginning of a new stimulation. Determining whether this was a real limitation for this study, or not, was not identified. However, future studies could try to replace the fixed interval with a random inter-stimulus interval, which could run between random ranges of seconds (e.g. 16 to 24 seconds). This application might help to improve the efficacy of the CIT and the validity of its results.

Furthermore, the CIT format requires only a few limited options per question (see Chapter 3, section 3.3). This is considered one of the requirements that limit the practical applicability of the CIT. Consequently, it would be possible to identify locations or other details of the crime only if the exact information is included in the set of test questions. When trying to determine the place, the type or the time of a criminal attack, a number of infinitive possibilities are available when guessing. For this reason, it would be useful to apply another methodology that includes more options. An example could be the partial test. Specifically, the structure of this test allows the presentation of more generic location options (e.g. North, Midwest), which then break down into smaller and smaller partitions until the locations are identified. This approach could help include more options, making the application of the CIT more suitable for realistic situations.

Another limitation is related to the scoring procedure applied in this study. The present research applied a more traditional standard coding procedure, which followed Lykken's method (1998) applied on raw and standardized scores (Ben-Shakhar, 1985). The Lykken's score procedure is considered the most traditional method for scoring CIT and it has been widely used in the CIT literature (e.g. Meixner & Rosenfeld, 2011; Meijer et al., 2010, Elaad & Ben-Shakhar, 2006; Gamer, Vossel, & Godert, 2006; Selle, Verschuere, Kindt, Meijer, & Ben-Shakhar, 2016; Vandenbosh, Verchuere, Crombex, & Clercq, 2008). As explained in section 3.7 this procedure is extremely practical. In fact, the conditional distribution of the scores obtained via the Lykken method for unknowledgeable examinees is already known. Therefore, the occurrence of false-positive results can be easily calculated (Timm, 1989). In addition, this method can be used without quantification and parameter estimations. Finally, in cases where the habituation effect occurs during the test, correction is not required when using the Lykken method, even if physiological levels change. In fact, the Lykken method allows computation of the scores by ranking them for each block, avoiding any alteration in the analysis (Matsuda, Nitton, & Allen, 2011). In line with the guidelines and indications by Maijer et al. (2011), the present study decided to use the Lykken method alongside the within-individual standardisation procedure (for more information, see section 3.7).

Although some advantages have been highlighted, both the Lykken's scoring method and the z-scoring averaging have several disadvantages. The Lykken method does not consider quantitative differences between responses to relevant and non-relevant items, scoring the same response even if three times as large as the next one (Meijer et al., 2011); while the z-scoring procedure ignores individual differences in response patterns among examinees (Matsuda et al., 2006). In addition, this last method does not consider the differences in general accuracies among measures; as every measure is weighted equally the level of accuracy cannot be assessed (Matsuda et al., 2012). For this reason, future studies should try to overcome these disadvantages by applying other scoring procedures, such as logistic regression discrimination, latent class discrimination, Bayesian classification, multivariate normal distribution discrimination and dynamic mixture distribution discrimination (Matsuda et al., 2012). The present study adopted the scoring procedures suggested by previous research conducted on the validity CIT (Vandenbosch, Verschuere, Crombez, & De Clercq, 2009; Selle, Verschuere, Kindt, Meijer, & Ben-Shakhar, 2016, iMotions, 2017; Kraphol, 2007). However, it needs to be considered that many other scoring methods could have been applied when analysing results from CIT examinations, which might have produced different results. Similarly, the type of filtering process used for cleaning the noise of the signal of this data set could be considered as a potential limitation of this study. The present study employed certain filtering types and specific parameters suggested by some of the most recent studies (e.g. Kircher & Raskin, 1988; National Research Council, 2003; Nelson, 2018). However, other filters and parameters could also be applied to produce valid results within the same data set, although it is possible that different signal processing designs as well as different filter procedures might result in different scores and different results (Nelson, 2018). Therefore, this could cause frustration to those researchers who believe that different signal processing models should always perform equivalently or uniformly.

For these reasons, future research could focus on assessing the differences between the scoring methods and the application of different filters by using the data used for this study. This will allow comparison of the validity and differences of other filter and scoring approaches, determining which one is the most reliable.

8.2.2 False positive

One of the most significant limitations impacting the validity of the CIT is the false positive effect, which can be caused by many factors, such as interviewees' knowledge of and personal relevance to items they are being questioned about. Based on the results reported in Studies III and IV, participants from the control groups reacted physiologically when asked about some of the relevant items during the test (e.g. Paris, the Eiffel Tower). This caused an increase in the false positive rate. It is not very clear if this element entirely affected the results of the studies; however, in order to avoid this limitation, future studies could conduct a pre-test before undertaking the CIT test. Before the administration of the CIT, the examiner could ask the examinees whether the items listed in a paper possess personal relevance to them. If the examinees indicate personal associations with certain items, the examiner could replace those items (Krapohl et al., 2006). The current study did not apply this proposed procedure, as the aim of the study was to analyse the data at the group level and it was therefore necessary that participants were primed with the same stimuli. However, it would be interesting to evaluate if false positive rates could be reduced by ensuring that the key information was not personally relevant to the innocent examinees.

8.2.3 External validity

One of the major problems of studies investigating polygraph accuracy is the issue of external validity. As none of the present studies occurred in a real context, to a certain degree the study lacks ecological validity. All the studies were designed in a laboratory setting; thus, the motivation to lie could have been affected negatively. Suspects in real-life situations might be more strongly motivated to lie, due to risks and consequences related to the results of the test (Bull et al., 2004). Consequently, they would be more engaged in trying to convince the examiner that they are innocent. Therefore, the validity of the results of the present study could have been affected by the nature of the laboratory setting.

In addition, the tasks and more specifically the mock crimes developed were inherently limited due to ethical restraints. In fact, when asking participants to plan a criminal attack, the assistant researcher tried to keep the discussion at a very generic level. This factor surely affected the intention to create a more realistic context. Furthermore, in order to ensure the standardisation of the test and allow an appropriate analysis of the results at the group level, the relevant options were decided a priori by the research team, with the results being the same for each participant. In other words, participants did not choose to plan a criminal act or voluntarily seek information about it. Their participation was based on the general information that was decided for them (e.g. Paris as the city, the Eiffel Tower as a location and guns as the weapons). This necessary methodological choice does not reflect realistic situations, in which individuals take a decision based on their opinions, after a process that can occur individually or in a group. Consequently, this factor might also have had a negative influence on their motivation to lie. Therefore, future studies should apply a different methodology that can ensure standardisation and efficacy at the same time; perhaps future studies could let participants decide to commit a mock crime or to inform themselves about it.

Another factor that might have affected the generalisation of the results concerns the type of sample used, which did not include any offender or potential criminal. As discussed in Chapter 6 (section 6.1), the present study was not interested in the examination of how demographic information or criminal experiences could affect the results of the test. Therefore, this methodological choice was intentional and justified. However, it needs to be considered that if this study would have included participants with a criminal background, results from the test could have been different. For this reason, it would be interesting to investigate how a criminal background could affect deceptive strategies. This would increase the generalisation of the results and their application in real contexts.

Another problem for the external validity of the results relates to the timing. For example, in the present study, participants were invited to complete a mock crime in 10 minutes and then perform the CIT after a short period of time. In a realistic context, the process of planning a criminal attack would generally take months of preparation. In fact, encoding, storing and recollecting key information within an event tends to diminish over time (Elaad, 1997). Due to the short time period between the stimulation and the test, it is considered that these factors did not affect the results of this study. For experimental reasons, the present study could not replicate the typical conditions present in a criminal context. Therefore, it must be taken into consideration that a long process of planning a criminal act might elicit different results.

8.2.4 Countermeasures

Another limitation that involves all the studies with the polygraph is the use of countermeasures for decreasing physiological reactions, to thus cheat the test (see Chapter 2, section 2.2.3). Although the present study did not investigate the effect of countermeasures, it is possible that participants used mental or physical strategies for influencing the results. Some of the countermeasures used by participants in this study might have involved attempting to daydream, controlling breathing rhythm, using relaxation techniques, counting internally, attempting to persuade themselves in using other items to perform the tasks, and intensely visualising that they did not perform the tasks.

In addition, during the visual stimulation, participants were supposed to look at the pictures during the examination and not elsewhere. However, it is possible that some participants tried not to look at the pictures when they were displayed, avoiding the recognition of the relevant item. This could have helped them to prevent an increase in their physiological reactions. It is not clear if this countermeasure played a significant role in this study. However, it would be interesting to investigate the efficacy of this strategy in future studies, with the support of eye tracker equipment. This additional equipment would help researchers to determine whether participants try to avoid the visual stimulation when examined.

8.2.5 Examiner Bias

In the present thesis, all the studies were conducted with a blind examiner (Chapter 3, section 3.5). This methodological procedure was to ensure that no examiner bias could affect the results during the test or during the scoring and analysis of data (Ben-Shakhar & Furedy, 1990; Rosenthal & Fode, 1996a; Elaad, 1997). However, a small limitation related to the examiner's role was encountered anyway. As the relevant items were the same for all the participants, after a few interviews in which participants reacted to the same items, the examiner could make an educated guess about which the relevant items were. Although the same blind examiner coded the data with the assistance of a blind assistant researcher, it was not possible to determine if this factor affected the present results. As suggested by Meijer et al. (2011), any bias or expectations from the experimenter could be communicated to the examinees in many unintentional ways, significantly affecting the validity of the examination. Therefore, it is essential that these factors are minimised. The best way to conduct the CIT and avoid any experimenter bias is by conducting a blind examination, like in the case of this thesis. However, in cases where 1) this protocol cannot be applied, 2) the experimenter becomes aware of the CIT items, or 3) the experimenter becomes aware of whether the examinee is innocent or guilty, the test should be automated by using a computer to present the questions and the items (Meijer et al., 2011), ensuring the best level of objectivity and reliability.

8.2.6 Habituation effect

One of the limitations that affected some of the studies is the habituation effect (see Chapter 2, section 2.1). The CIT is based on the concepts of orienting response and stimulus significance. In accordance with this theoretical framework, innocent examinees perceive all the items with the equivalent stimulus significance, eliciting orienting responses of equal magnitude that will eventually habituate. Conversely, guilty examinees' orienting responses will be elicited by the critical items, producing similar magnitudes (Meijer et al., 2011) (for more information, see section 2.1). Laboratory research has demonstrated that novel stimuli can indeed enhance an OR (e.g. Turpin, 1986). This mechanism was clearly demonstrated in the results of the studies of this thesis, showing clear physiological reactions of similar magnitude for the relevant questions (for the experimental groups) and similar physiological reactions for most of the items in the innocent groups. However, as predicted by Sokolov (1963), when repeated presentations of the same stimulus occur, the magnitude of the response gradually declined, producing a habituation effect. In line with this theoretical framework, it was found that participants potentially decreased their physiological reactions due to the repeated exposure to the relevant item, which decreased the effect upon SCR reactions after the second repetition (Barry, 2006; Barry, 2009; Ben-Shakhar, Lieblich, & Kugelmass, 1975; Furedy, Gigliotti, & Ben-Shakhar, 1994). This confirms that this thesis could not demonstrate an absolute absence of habituation effect for the studies conducted. On the other hand, this conclusive argument finds its place along with the general literature, which could not find any dishabituation effect in CIT studies (e.g. Ben-Shakhar et al., 2000). Conclusions from this thesis, therefore, confirm that this physiological phenomenon is inevitable in CIT examinations.

Despite the intent to manipulate the factors that could affect the results, the analysis showed that natural habituation of the physiological responses always occurs. Therefore, it is also possible that this effect was caused by other factors. For example, some researchers (Nakayama & Kizaki, 1990) state that SCR is more susceptible to habituation compared to other physiological measures. Therefore, future research should investigate this factor in more depth, trying to determine which physiological measure is more sensitive to this effect under similar methodological conditions. A follow-up study could perhaps use the same dataset from the present study and compare the validity and the sensitivity of the other physiological measures (e.g. RLL, blood pressure). Furthermore, another factor that could cause habituation effect concerns examinees' personal characteristics. In fact, certain individuals might be more susceptible to habituation than others (Ben-Shakhar & Gati, 2003). This suggests that future studies should further investigate how examinees' personal characteristics might affect the ability of the CIT to detect deception.

8.2.7 Number of Questions and Items

Studies have suggested precise requirements for the development of the CIT (Meijer et al., 2011) that could help to reduce an occurrence of the habituation effect and increase the detection accuracy rate of the test. One of these is the number of questions presented for each test. It was recommended that at least five questions need to be asked of the examinee in order to ensure high accuracy levels (Ben-Shakhar & Elaad, 2003). Following these guidelines, every test from the studies in this thesis were developed accordingly. In fact, each test presented a minimum of five questions per question category. This number appeared to have created a modest level of accuracy, as the CIT was able to detect relevant information with an accuracy rate that ranged between 70-80%. These results are in line with other studies (Ben-Shakhar & Elaad, 2003; Meijer et al., 2011; Verschuere et al., 2004) which adopted the same methodology and produced similar accuracy rates. In addition, some of these studies (e.g. Elaad & Ben-Shakhar, 1997; Elaad, 2002) emphasised the importance of question repetition, suggesting that detection efficiency increases with the number of repetitions of the CIT questions. For this reason, the CITs of this thesis always presented three repetitions (for Studies I and II) and two repetitions per question category (for Studies III and IV). Based on the conclusions from Ben-Shakhar and Elaad (2002) and Meijer et al., (2011), this helped the CITs of this thesis to reduce the risk of false positive, increasing the detection accuracy rate. However, results from the present studies clearly showed how the repetitions of questions inevitably elicited habituation effect after the first series of questions, decreasing the magnitude of the physiological responses. Although the magnitude of responses can differ and the detection rate is usually larger in field settings than in laboratory studies, this thesis presents results and conclusions that are in line with other studies in the literature (e.g. Vrij et al., 2013).

Another important element that needs to be discussed is the choice of the items in the CITs. In accordance with Meijer et al. (2011), to be able to develop an efficient test, the researcher must take into consideration that any salient item evokes a physiological response; therefore, if the critical items are very salient, an orienting response will also be evoked in innocent participants. For this reason, it is advised that the critical items should not be more salient or more plausible than the control items. However, they should not share common features with the critical item, thus the test should have items that can be easily differentiated (Ben-Shakhar et al., 1996). The studies from this thesis adhered to these principles, although Study I included a knife among other more neutral items (such as hand and hammer). In fact, innocent participants reacted to the item even if they were not primed (for more information, see section 3.5). These results confirm the validity of the guidelines and suggest that the critical items needed to be less critical/salient compared to the control items. In line with these guidelines, Study II changed the nature of the priming tasks (attacking a mannequin) and removed the knife from the test, including only those items that were distinct but plausible. In addition, the literature also suggests that the orienting/novelty response always occurs with the first item presented in the question (Meijer, at al., 2011). For this reason, none of the critical items were presented as the first option for any of the CITs developed for this test. Although the CITs of this thesis strictly followed the recommendations and guidelines suggested by the literature, there are certain methods that should be integrated in the process of developing a CIT to ensure that every item chosen is adequate. For instance, the researcher could test the items a priori with the Doob Kirshenbaum procedure (Doob & Kirshenbaum, 1973), which ensures that all the items chosen are plausible and equally salient. This method consists of presenting all the items to a group of volunteers, asking them to pick the item they find most plausible. If the critical item is chosen, then the item must be changed. Another method that can help with the choice of items is to show the test items to the examinee before the administration of the test. In this way, the suspect has the opportunity to admit knowledge of any correct item beforehand (Meijer et al., 2011). Although the choice of items did not cause any major methodological limitation for this thesis, for future studies it is advisable to apply one of these procedures to ensure the plausibility of the items.

8.2.8 Physiological Measures

It has been suggested that the orienting response is characterised by changes in a range of physiological measures: when stimulated by the relevant items, heart rate usually decelerates (Graham & Clifton, 1996), respiratory system suppresses (Lynn, 1996) and SCR increases (Selle, Kindt, Vershuere, Meijer, & Ben-Shakhar, 2016). This pattern is caused by the coactivation of the sympathetic and the vagal branch of the autonomic nervous system, partially related to the OR mechanism (for more information, see section 2.1). Results from the studies of this thesis supported and confirmed the behaviour of these physiological responses (see section 3.2); in particular, results showed that SCR was the one with the highest detection accuracy, perfectly in line with previous studies (Selle, Kindt, Vershuere, Meijer, & Ben-Shakhar, 2016; Hirota, Ogawa, Matsuda, & Takasawa, 2009; Bradley et al., 2011; Council, 2003b; Holden, 2000; Saxe, 1994; Society, 1986; Ben-Shakhar & Elaas, 2004). Although preliminary results were only conducted using EDA, the follow up results used the combination of SCR, RLL and heart rate. In fact, it was suggested that CIT validity can be increased above the best single measure by combining the measures together (e.g. Ben-Shakhar & Dolev, 1999; Ben-Shakhar & Elaad, 2002). This analysis was conducted by averaging response differences between relevant and control items across physiological channels (for more information, see section 3.2). This combination was necessary as the contributions of the single physiological measures is not reliable enough (Gamer, 2011). Gamer et al (2006) suggested that the weighted combination of the SCR scores, RLLs and heart rate responses generated a larger validity rate compared to the best scores from any single physiological measure. For this reason, after the preliminary analysis, this thesis wanted to combine the physiological measures for enhancing the CIT validity of the studies. Although results from this thesis were able to provide more accurate detection rates due to the combination of the measures, there are certain limitations that can affect the validity of the results, such as differences in physiological responsiveness. In addition, it is important to consider the difference between conclusions drawn in a laboratory context compared to a field setting. In fact, studies have suggested how certain physiological measures do not uniformly respond to CIT examinations in every setting; on the contrary, certain physiological patterns might occur in a laboratory or in a real setting (Suzuki et al., 2004). Similarly, certain physiological measures might follow different habituation patterns (Gamer et al., 2008). Therefore, it is important to take these limitations into consideration when considering the conclusions from this thesis, especially regarding the validity of the single physiological measures but also their combination as the best way to enhance the validity of the test.

8.3 Recommendations and Suggestions for Further Research

The results from this study contribute to the general literature around the CIT, which support the validity of the test for assessing concealed information. However, as listed in the Limitations Section (see Chapter 8, section 8.2), the test presents various threats to its validity. Therefore, there are ample opportunities for improving the problematic aspects of the test.

8.3.1 Improving the influence of countermeasures

It was suggested that participants from this study could have attempted to use certain physiological strategies for cheating the results of the test. Although the present study did not clarify whether these strategies affected the results, this limitation remains one of the most problematic aspects of polygraph tests. For these reasons, future researchers could apply a broader approach to study the role of countermeasures. In fact, developing this work further would help to understand when the application of these countermeasures causes a misclassification of guilty suspects, increasing false positive outcomes (Ben-Shakhar et al., 2011). Future studies could also try to investigate whether countermeasures affect the physiological measures differently, determining which one is the most resistant or the most sensitive to the use of those techniques. It would be beneficial to develop new strategies for decreasing the effect of countermeasures, for example by not informing participants that they will be physiologically recorded. If participants are unaware of being tested, they might not attempt to conceal information. Developing this work further could help to decrease the false negative rates, improving the application of the CIT and reducing the risk of skewing the results.

8.3.2 Decreasing false positive rates

Results from the present study showed a consistent percentage of false positives across the four studies. Although the false rate was not at concerning levels, this percentage represents a threat to the validity of the test. The present study suggested that the occurrence of false positive rates could be caused by the influence of the habituation effect, the structure of the test and personal relevance to the items presented. However, many other factors could play a relevant role in this effect, such as the sensitivity of certain physiological measures. One novel contribution would be using the existing physiological measures with other equipment used for detecting deception, such as an eye tracking system, event-related brain potentials (ERPs), electrophysiological monitoring methods (EEG), and functional magnetic resonance imaging (fMRI). The combination of CIT with other sensors or equipment could be beneficial for increasing the reliability of the results and decreasing the percentage of false positive occurrences.

Similarly, the CIT could be combined with the analysis of the response latency (or reaction time - RT). Previous studies have shown that response latency measures successfully detect concealed information in deceptive participants (Ben-Shakhar, 2012). The most valuable use of the RT-based CIT aims to examine the underlying cognitive processes involved in deceptive behaviour. This analysis has been found to be very effective in discriminating concealed information from control information, with a very large effect size (Verschuere et al., 2010). This method could provide information regarding the cognitive processes involved during deception, such as incidental encoding of stimuli in the CIT (Gamer et al., 2010). Therefore, RT-based CIT could help to improve the accuracy of the test by decreasing false positive/negative rates.

Another factor that could affect the reliability of the CIT when examining innocent examinees concerns personal characteristics, such as anxiety and personality traits. Other researchers tried to investigate these factors but this was more than 30 years ago (e.g. Gudjonsson, 1982; Raskin & Hare, 1978). Therefore, research that is more recent is needed, with an emphasis on whether and how these factors could have a real impact on results. For example, future research could consider the application of new personality scales and questionnaires such as the Ten-item Personality Inventory (TIPI), developed by Gosling et al. (2003). The use of new instruments could evaluate in-depth the potential correlation between personality traits and deception. This suggestion could also be extended to the study of demographic factors, such as education level, age, social background, and criminal experience. This might establish whether there are any differences in the way in which individuals try to conceal information, or not.

8.3.3 Increasing the external validity of the CIT

One of the main concerns regarding the application of the CIT in practical contexts relates to the external validity of the test. As discussed in Chapter 2 (see section 2.2.2), most of the CIT studies were conducted in a laboratory setting, causing all the limitations already discussed (see Chapter 8, section 8.2). Consequently, future research should focus on improving the external validity of the test, for example by using more realistic forensic settings. CIT studies could try to use former offenders or the prison population, which would increase the generalisation of the results.

Another suggestion could involve the use of the mock-crime paradigm (Gamer et al., 2010; Seymour & Fraynt, 2009). This consists of instructing participants to choose the critical items they prefer. Furthermore, the administration of the CIT is delayed by a few weeks. In this way, the mock crime would introduce some of the typical conditions present in a realistic situation. This methodology should be able to avoid some of the limitations encountered in this study, improving the quality of external validity for future applications.

8.3.4 Investigating Social Interactions

The present study tried to determine if the group setting could influence the physiological responses of the participants when trying to conceal information. However, the purposes of this study were not focused on investigating social aspects that could improve the application of the CIT in a group context, for example the social interactions of the participants, or the communication and decision-making process. It would be interesting trying to evaluate how the social dynamics between the participants can affect deceptive behaviour during the CIT examinations.

Furthermore, the studies in the present study administrated the CIT individually to each participant. This methodological choice reflected the standard setting of the CIT. However, this choice was the consequence of a lack of equipment that could have supported the idea of administrating simultaneous interviews. For these reasons, future studies could try to investigate how multiple interviews could affect the physiological responses of the participants when trying to be deceptive. Results from these studies would provide a great contribution to the understanding of whether there is a correlation between social dynamics and deception, or not.

8.4 Implications of the Present Thesis

The main implication of the current research is a better understanding of how memory can be influenced by specific factors present during a particular event. The research demonstrated that specific elements, such as a sound or the shape of an object, could have an impact on examinees' memory as well as in the way these elements can be recollected. Inevitably, the result of this process has an influence on how physiological responses are provoked during a polygraph examination. Thus, this study suggested that specific memories associated with particular information are stronger than others are, due to the personal impact or personal associations they create in the memory. Although this was an exploratory study, the outcomes provided contributions that will be useful when developing a CIT. In addition, some of the methodological suggestions from this study could help improve the formulation of questions, answer alternatives and increase the efficacy of the test.

The introduction of visual stimulation represents another implication of this study. Even though this new methodological setting was explorative, results demonstrated that visual stimulation is able to increase examinees' responses, providing more valid results. However, this aspect will need to be investigated more in another study. The practical application of these conclusions will be extremely relevant and helpful for police examinations. In fact, the amount of photographic evidence available in forensic cases could be used as central information contained in the test. Consequently, this methodological support could help to conduct interviews with suspects, trying to investigate whether they possess knowledge of a crime. Similarly, this approach could be used with eyewitnesses, when trying to identify whether they recognise a suspect or an offender (e.g. with the use of line-ups).

The last implication of this study focused on trying to understand if the CIT was able to detect criminal knowledge from the participants. The study tried to understand if planning a crime individually or in a group have different impacts when trying to hide knowledge about an event. The innovative element of this study focuses on the application of the CIT in a setting that was poorly researched in previous literature. The CIT was able to detect criminal knowledge from a group of fake suspects. The implications for these results focus on applying the CIT to a group of criminals. This could be applied to any forms of organised crime, such as gangs or terrorism. Due to the escalation of this typology of crimes and their influence at international levels, the CIT could represent a useful tool able to support interviews and to prevent crimes.

8.5 Conclusions

The present study evaluated different aspects of the validity of the CIT. Three main conclusions can be drawn from the thesis. First, the CIT was able to detect relevant information from participants in all the studies with a moderate accuracy rate. This demonstrates that the theoretical mechanism of the CIT works, which can be used for detecting information that individuals try to conceal. Second, the application of visual stimulation can be considered as a useful supportive element during CIT examinations, which deserves more attention in future research. Finally, the CIT also demonstrated a moderate validity when trying to detect information from a group of examinees. This novel application has important implications for the world of organised groups.

In conclusion, this study suggests multiple applications of the CIT as a preventative tool for forensic settings, for solo or groups of offenders. With the support of future research, the CIT would be able to support investigations when trying to detect possible criminal information, prioritize and identify potential suspects involved in criminal activities. In this way, the still certain limitations of the CIT would not have irreversible consequences on examinees.

REFERENCES

Allen, J., Iacono, W.G., & Danielson, K.D. (1992). The identification of concealed memories using the event-related potential and implicit behavioural measures: a methodology for prediction in the face of individual differences. *Psychophysiology*, 29, 504-522.

Alloway, T. P., McCallum, F., Alloway, R. G., & Hoicka, E. (2015). Liar, liar, working memory on fire: Investigating the role of working memory in childhood verbal deception. *Journal of experimental child psychology*, 137, 30-38.

Alvarez, G. A., & Canavagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological Sciences, 15*, 106-111.

Amir, N., Leiner, A. S., & Bomyea, J. (2010). Implicit memory and posttraumatic stress symptoms. *Cognitive Therapy & Research, 34*, 49–58.

Anderson, C. A., Benjamin, A. J., & Bartholow, B. D. (1998).Does the gun pull the trigger? Automatic priming effects of weapon pictures and weapon names. *Psychology Science, 9*, 308-314.

Andreassi, J. (2007). *Psychoneuroimmunology* (4th Edn). New York: Elsevier.

Anolli, L., & Ciceri, R. (1997). The voice of deception: Vocal strategies of naive and able liars. *Journal of Nonverbal Behavior*, 259-284.

Aquinas, T. (1972). Question 110: Lying *Summa Theologiae* (Vol. 41). New York: McGraw-Hill.

Atkinson, R. C., & Shiffrin, M. R. (1968). Human memory: A proposed system and its control processes. In K. W. Spence, & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 2, pp. 89-195). New York: Academic Press.

Augustine, S. (1952). On Lying. In R. J. Deferrari (Ed.), Fathers of the Church (Vol. Volume 16, pp. 51–110). New York: Fathers of the Church.

Backster, C. (1963a). Polygraph professionalism through technique standardization. *Law & Order,* 11 (4), 63-64.

Backster, C. (1963b). New standards in polygraph chart interpretation - do the charts speak for themselves? *Law & Order,* 11 (6), 67-68.

Baddeley, A. D. (1986). *Working memory*. Oxford, UK:Clarendon Press:.

Baddeley, A. D. (2000). The episodic buffer: a new component of working memory. *Trends in Cognitive Science, 4*, 417-423.

Baddeley, A. D. (2012). Working memory: theories, models, and controversies. *Annual Review of Psychology, 63*, 1-29.

Baddeley, A. D., & Hitch, G. (1974). *Working memory:The psychology of learning and motivation: Advances in research and theory.* New York: Academic Press.

Banich, M. T. (1999). *Neuropsychology. The Neural Bases of Mental Function*. New York: Houghton Mifflin Company.

Bargh, J. A., & Chartrand, T. L. (2000). Studying the mind in the middle: A practical guide to priming and automaticity research. In H. Reis & C. Judd (Eds.), *Handbook of research methods in social psychology* (pp. 253–285). New York, NY: Cambridge University Press.

Barland, G. H. (1999). A message. *American Polygraph Association Newsletter, 32(3)*, 16-17.

Barland, G. H., & Raskin, D. C. (1975a). An evaluation of field techniques in detection of deception. *Psychophysiology, 12(3)*, 321-330.

Barland, G. H., & Raskin, D. C. (1975b). Psychopathy and detection of deception in criminal suspects. *Psychophysiology, 12*, 224.

Bartlett, F. C. (1932). *Remembering: A study in Experimental and Social Psychology* Cambridge: Cambridge University Press.

Barry, R. J. (2006). Promise versus reality in relation to the unitary orienting reflex: A case study examining the role of theory in psychophysiology. *International Journal of Psychophysiology*, 62(3), 353-366.

Barry, R. J. (2009). Habituation of the orienting reflex and the development of preliminary process theory. *Neurobiology of learning and memory,* 92(2), 235-242.

Basden, B. H., Basden, D. R., Bryner, S., & Thomas, I. R. L. (1997). A comparison of group and individual remembering: Does collaboration disrupt retrieval strategies? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23(5)*, 1176.

Bashore, T. R., & Rapp, P. E. (1993). Are there alternatives to traditional polygraph procedures? *Psychological Bulletin, 113(1)*. doi:10.1037/0033-2909.113.1.3.

Beardsley, T. (1999). Truth or ConsequencesA: polygraph screening program raises questions about the science of lie detection. *Scientific American, 140*, 21-24.

Ben-Shakhar, G. (1980). Habituation of the orienting response to complex sequences of stimuli. *Psychophysiology,*17(6), 524-534.

Ben-Shakhar, G. (1985). Standardization within individuals: A simple method to neutralize individual differences in skin conductance. *Psychophysiology,* 22(3), 292-299.

Ben-Shakhar, G. (2002). *A critical review of the control question test (CQT)* Kleiner, M. (Ed). London: Academic Press.

Ben-Shakhar, G. (2011). Countermeasures. In Verschuere, B; Ben-Shakhar, B. & Meijer, E.H. (Eds.), *Memory Detection*. New York: Cambridge University Press.

Ben-Shakhar, G., & Dolev, K. (1996). Psychophysiological detection through the guilty knowledge technique: Effect of mental countermeasures. *Journal of applied psychology, 81(3)*, 273.

Ben-Shakhar, G., & Elaad, E. (2002). Effects of questions' repetition and variation on the efficiency of the guilty knowledge test: A reexamination. *Journal of Applied Psychology, 87(5)*, 972.

Ben-Shakhar, G., & Elaad, E. (2002). The Guilty Knowledge Test (GKT) as an application of psychophysiology: Future prospects and obstacles. *Handbook of polygraph testing*, 87-102.

Ben-Shakhar, G., & Elaad, E. (2003). The validity of psychophysiological detection of information with the Guilty Knowledge Test: A meta-analytic review. *Journal of Applied Psychology*,88(1), 131.

Ben-Shakhar, G., & Furedy, J. J. (1990). *Theories and applications in the detection of deception: A psychophysiological and international perspective*. New York: Springer.

Ben-Shakhar, G., & Gati, I. (1987). Common and distinctive features of verbal and pictorial stimuli as determinants of psychophysiological responsivity.Journal of Experimental Psychology: General,116(2), 91.

Ben–Shakhar, G., & Gati, I. (2003). The effects of serial position and frequency of presentation of common stimulus features on orienting response reinstatement. *Psychophysiology,* 40(1), 139-145.

Ben- Shakhar, G., Lieblich, I., & Kugelmass, S. (1975). Detection of information and GSR habituation: An attempt to derive detection efficiency from two habituation curves. *Psychophysiology*, 12(3), 283-288.

Berntson, G. G., Bigger, J. T. J., Eckberg, D. L., Grossman, P., Kaufmann, P. G., Malik, M., van der Molen, M. W. (1997). Heart rate variability: origins, methods, and interpretive caveats. *Psychophysiology, 34(6)*, 623-648.

Biederman, I. (1987). Recognition-by-components: a theory of human image understanding. *Psychological Review*, 94(2), 115.

Bond, C. F., & DePaulo, B. M. (2006). Accuracy of deception judgments. *Personality and Social Psychology Review, 10(3)*, 214-234.

Burgoon, J. K., & Floyd, K. (2000). Testing for the motivation impairment effect during deceptive and truthful interaction. *Western Journal of Communication (includes Communication Reports)*,64(3), 243-267.

Bradley, M. T., & Ainsworth, D. (1984). Alcohol and the psychophysiological detection of deception. *Psychophysiology, 21(1)*, 63-71.

Bradley, M. T., Barefoot, C. A., & Arsenault, A. M. (2011). Leakage of information to innocent suspects. In Verschuere, B., Ben-Shakhar, B. & Meijer, E.H. (Eds.), *Memory Detection*. New York: Cambridge University Press.

Bradley, M. T., & Cullen, M. C. (1993). Polygraph lie detection on real events in a laboratory setting. *Perceptual and motor skills, 76(3)*, 1051-1058.

Bradley, M. T., & Klohn, K. I. (1987). Machiavellianism, the Control Question Test and the detection of deception. *Perceptual and Motor Skills, 64(3)*, 747-757.

Bradley, M. T., MacDonald, P., & Fleming, I. (1989). Amnesia, feelings of knowing, and the guilty knowledge test. *Canadian Journal of Behavioural Science/Revue Canadienne Des Sciences Du Comportement, 21(2)*, 224-231. doi:10.1037/h0079810

Bradley, M. T., MacLaren, V. V., & Carle, S. B. (1996). Deception and nondeception in guilty knowledge and guilty actions polygraph tests. *Journal of Applied Psychology, 81(2)*, 153-160. doi:10.1037/0021-9010.81.2.153

Brady, T. F., Konkle, T., & Alvarez, G. A. (2011). A review of visual memory capacity: Beyond individual items and toward structured representations. *Journal of Vision 11(5):4*. doi:10.1167/11.5.4.

Bradley, M. T. & Rettinger, J. (1992). Awareness of crime-relevant information and the guilty knowledge test. *Journal of Applied Psychology*. 77 (1), 55 - 59.

Braithwaite, J. J., Watson, D. G., Jones, R., & Rowe, M. (2013). A guide for analysing electrodermal activity (EDA) & skin conductance responses (SCRs) for psychological experiments. Psychophysiology, 49(1), 1017-1034.

Brannan, D. W., Eslerm, P. F., & Anders Strindberg, N. T. (2001). Talking to "terrorists": Towards an independent analytic framework for the study of violent substate activism.*Studies in Conflict and Terrorism*, 24, 3-24.

Boucsein, W. (2012). *Electrodermal activity.* Springer Science & Business Media.

Bryant, F. B., Grimm, L. G., & Yarnold, P. R. (2000). *Reading and understanding more multivariate statistics*. Washington DC: American Psychological Association.

Buckley, J. P., & Senese, L. C. (1991). The Influence of Race and Gender on Blind Polygraph Charts Analyses. *Polygraph, 20(4)*, 247-258.

Bull, R., Baron, H., Gudjonsson, G., Hampson, S., Rippon, G., & Vrij, A. (2004). *A review of the current scientific status and fields of application of polygraphic deception detection*. Final report from the BPS Working Party. London: British Psychological Society.

Butterworth, S. (1930). On the theory of filter amplifiers. Wireless Engineer, 7(6), 536-541.

Canetti, D., & Pedahzur, A. (2002). The effects of contextual and psychological variables on extreme right-wing sentiments.*Social Behavior & Personality: an international journal*,30(4).

Carmel, D., Dayan, E., Naveh, A., Raveh, O., & Ben-Shakhar, G. (2003). Estimating the validity of the guilty knowledge test from simulated experiments: the external validity of mock crime studies.*Journal of Experimental Psychology Applied, 9*(4), 261.

Carlson, N. R. (2010). *Psychology the science of behavior*: Canada: Pearson Canada Inc.

Carson, T. L. (2009). Lying, Deception, and related Concepts. In C. Martin (Ed.), *The Philosophy of Deception*. New York: Oxford University Press.

Christianson, S. A., & Loftus, E. (1990). Some characteristics of people's traumatic memories. *Bulletin of the Psychonomic Society., 28*, 195–198.

Clark, T. (1987). Echoic memory explored and applied. *Journal of Consumer Marketing, 4 (1)*, 39–46.

Cohen, R. L. (1981). On the generality of some memory laws. *Scandinavian Journal of Psychology, 22(1)*, 267-281.

Cohen, T. R., Gunia, B. C., Kim-Jun, S. Y., & Murnighan, J. K. (2009). Do groups lie more than individuals? Honesty and deception as a function of strategic self-interest. *Journal of Experimental Social Psychology, 45(6)*, 1321-1324.

Coltheart, M. (1980). Iconic memory and visible persistence. *Perception & Psychophysics, 27 (3)*, 183–228. doi:10.3758/BF03204258

Corteen, R. S. (1969). Skin conductance changes and word recall. *British Journal of Psychology,* 60(1), 81-84.

Council, N. R. (2003a). *The polygraph and lie detection* Washington D.C.: The National Academies Press.

Council, N. R. (2003b). *The Polygraph and Lie Detection. Committee to Review the Scientific Evidence on the Polygraph. Division of Behavioral and Social Sciences and Education*. Washington, DC.: The National Academies Press.

Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of verbal learning and verbal behavior, 11(6)*, 671-684.

Crenshaw, M. (2000). The psychology of terrorism: An agenda for the 21st century.*Political psychology*, 21(2), 405-420.

Crider, A., & Lunn, R. (1971). Electrodermal lability as a personality dimension. *Journal of Experimental Research in Personality*.

Curran, T., Debuse, C., Woroch, B., & Hirshman, E. (2006). Combined Pharmacological and Electrophysiological Dissociation of Familiarity and Recollection. *Journal of Neuroscience, 26 (7)*, 1979–1985.

Daniels, C. W. (2002). Legal aspects of polygraph admissibility in the United States. In M. Kleiner (Ed.), *The Handbook of Polygraph Testing* (pp. 327-338). London: Academic Press.

Darley, J. M., & Gross, P. H. (1983). A hypothesis-confirming bias in labeling effects. *Journal of Personality and Social Psychology, 44(1)*, 20.

Darrow, C. W. (1929). Differences in the physiological reactions to sensory and ideational stimuli. *Psychological Bulletin, 26*, 185–201.

Davis, R. C. (1961). *Physiological responses as a means of evaluating information. The manipulation of human behavior*. New York: Wiley.

DePaulo, B. M., Kashy, D. A., Kirkendol, S. E., Wyer, M. M., & Epstein, J. A. (1996a). Lying in everyday life. *Journal of Personality and Social Psychology,, 70(5)*, 979-995.

DePaulo, B. M., Kirkendol, S. E., Tang, J., & O'Brien, T. P. (1988). The motivational impairment effect in the communication of deception: Replications and extensions. *Journal of Nonverbal Behavior, 12(3)*, 177-202.

DePaulo, B. M., Lindsay, J. I., Malone, B. E., Muhlenbruck, L., CharIton, K., & Cooper, H. (2003). Cues to deception. *Psychological Bulletin, 129(1)*, 74-118. doi:10.1037/0033-2909.129.1.74

DePaulo, B. M., & Pfeifer, R. L. (1986). On-the-job experience and skill at detecting deception. *Journal of Applied Social Psychology, 16*, 249-267.

DePaulo, B. M., & Rosenthal, R. (1979). Telling lies. *Journal of Personality and Social Psychology,* 37 (10), 1713 -1722.

Dick, A. O. (1974). Iconic memory and its relation to perceptual processing and other memory mechanisms. *Perception & Psychophysics, 16 (3)*, 575–596. doi:10.3758/BF03198590

Dohm, T. E., & Iacono, W. G. (1993). *Design and pilot of a polygraph filed validation study with appendices* (D.o.D. P.Institute Ed.). Fort McClellan, AL.

Dudai, Y. (2004). The Neurobiology of Consolidations, Or, How Stable is the Engram. *Annual Review of Psychology, 55*, 51–86. doi:10.1146/annurev.psych.55.090902.142050. PMID 14744210

Ekman, P. (1988). Lying and nonverbal behavior: Theoretical issues and new findings. *Journal of Non verbal Behavior, 12(3)*, 163-175.

Ekman, P. (1992). *Telling Lies*. London: W.W. Norton & Company.

Ekman, P., & Friesen, W. V. (1969). The repertoire or nonverbal behaviour:categories, origins, usage, and coding. *Semiotica, 4*, 49-98.

Ekman, P., & O'Sullivan, M. (1991). Who can catch a liar? . *American Psychologist, 46*, 913-920.

El Haj, M., Antoine, P., & Nandrino, J. L. (2017). When deception influences memory: the implication of theory of mind. *The Quarterly Journal of Experimental Psychology, 70*(7), 1166-1173. doi:10.1080/17470218.2016.1173079

El Haj, M., Moroni, C., Luyat, M., Omigie, D., & Allain, P. (2014). To what extent does estination recall induce episodic reliving? Evidence from Alzheimer's disease. J*ournal of clinical and experimental neuropsychology, 36*, 127-136.

Elaad, E. (1990). Detection of guilty knowledge in real-life criminal investigations.*Journal of Applied Psychology*,75(5), 521.

Elaad, E. (1997). Polygraph examiner awareness of crime-relevant information and the guilty knowledge test. *Law and Human Behavior*, 21(1), 107-120.

Elaad, E. (1998). The challenge of the concealed knowledge polygraph test. *Expert Evidence, 6*, 161-187.

Elaad, E. (2011). Vaidity of the Concealed Information Test in realsitic contexts. In B. Verschuere, G. Ben-Shakhar, & E. H. Meijer (Eds.), *Memory Detection*. New York: Cambridge University Press.

Elaad, E. (2014). Differences in the readiness of guilty and informed innocent examinees to cooperate on the guilty action test. . *Psychophysiology, 51(1)*, 70-79. doi:10.1111/psyp.12146

Elaad, E., & Ben-Shakhar, G. (1989). Effects of motivation and verbal response type on psychophysiological detection of information. *Psychophysiology, 26(4)*, 442-451.

Elaad, E., Ginton, A., & Jungman, N. (1992). Detection measures in real-life criminal guilty knowledge tests.*Journal of Applied Psychology*,77(5), 757.

Elaad, E., Ginton, A., & Ben-Shakhar, G. (1994). The effects of prior expectations and outcome knowledge on polygraph examiners' decisions. *Journal of Behavioral Decision Making, 7(4)*, 279-292.

Elaad, E., Ginton, A., & Jungman, N. (1992). Detection measures in real-life criminal guilty knowledge tests. *Journal of Applied Psychology, 77*, 757-767.

Engelkamp, J. (1998). *Memory for actions*. Hove, U.K.: Psychology Press.

Engelkamp, J., & Krumnacker, H. (1980). Imaginale und motorische Prozesse beim Behalten verbalen Materials. *Zeitschrift für experimentelle und angewandte Psychologie, 27*, 511-533.

Engelkamp, J., & Zimmer, H. D. (1985). Motor programs and their relation to semantic memory. *German Journal of psychology*. (9), 239-254.

Engelkamp, J., Zimmer, H. D., Mohr, G., & Sellen, O. (1994). Memory of self-performed tasks: Self-performing during recognition. *Memory & Cognition*, 22(1), 34-39.

Engelkamp, J., & Zimmer, H. D. (1997). Sensory factors in memory for subject-performed tasks *Acta Psychologica, 96*, 43-60.

Engelkamp, J., & Zimmer, H. D. (1999). Levels of processing effects in subject-performed tasks. *Memory & Cognition 27(5)*, 907-914.

Engelkamp, J., Zimmer, H. D., Mohr, G., & Sellen, O. (1994). Memory of self-performed tasks: Self-performing during recognition. *Memory & Cognition, 22(1)*, 34-39.

Eysenck, M. W. (2012). *Fundamentals of cognition*. New York: Psychology Press.

Eysenck, M. W. (2015). *Cognitive Psychology: A Student's Handbook.* In M. T. Keane (Ed.), (7th Edition ed.). U.K: Psychology Press.

Fahy, F. L., Riches, I. P., & Brown, M. W. (1993). Neuronal activity related to visual recognition memory: long-term memory and the encoding of recency and familiarity information in the primate anterior and medial inferior temporal and rhinal cortex. *Experimental Brain Research, 96*, 457-472.

Farwell, L. A., & Donchin, E. (1991). The Truth Will Out: Interrogative Polygraphy ("Lie Detection") With Event-Related Brain Potentials. *Psychophysiology, 28(5)*, 531-547.

Fere, C., 1888. Note on changes in electrical resistance under the effect of sensory stimulation and emotion. *Comptes Rendus des Seances dela Societe de Biologie,* 5, 217 – 219.

Fiedler, K., Schmid, J., & Stahl, T. (2002). What is the current truth about polygraph lie detection? *Basic and Applied Social Psychology, 24(4)*, 313-324.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. London: Sage Publication.

Ford, C. (1996). *Lies! Lies! Lies! The Psychology of Deceit.* Washington, DC: American Psychiatric Press.

Forman, R. F., & McCauley, C. (1986). Validity of the positive control polygraph test using the field practice model. *Journal of Applied Psychology, 71 (4)*, 691-698.

Förster, J., Liberman, N., & Friedman, R. S. (2007). Seven principles of goal activation: A systematic approach to distinguishing goal priming from priming of non-goal constructs. *Personality and Social Psychology Review, 11(3)*, 211-233.

Fowles, D. C., Christie, M. J., Edelberg, R., Grings, W. W., Lykken, D. T., & Venables, P. H. (1981). Publication recommendations for electrodermal measurements. *Psychophysiology, 18(3)*, 232–239.

Fox, E., Russo, R., Bowles, R., & Dotton, K. J. E. P. G., 130:681-700. 6. (2001). Do threatening stimuli draw or hold attention in visual attention in subclinical anxiety. . *Experimental Psychology 130*, 681-700.

Furedy, J. J., & Ben-Shakhar, G. (1991). The roles of deception, intention to deceive, and motivation to avoid detection in the psychophysiological detection of guilty knowledge. *Psychophysiology*, 28(2), 163-171.

Furedy, J. J., Gigliotti, F., & Ben-Shakhar, G. (1994). Electrodermal differentiation of deception: The effect of choice versus no choice of deceptive items. *International Journal of Psychophysiology*, 18(1), 13-22.

Gamer, M., Kosiol, D., & Vossel, G. (2010). Strength of memory encoding affects physiological responses in the Guilty Actions Test. *Biological Psychology,* 83(2), 101-107.

Gamer, M., Verschuere, B., Crombez, G., & Vossel, G. (2008). Combining physiological measures in the detection of concealed information. *Physiology & Behavior, 95(3)*, 333-340. doi:10.1016/j.physbeh.2008.06.011

Ganis, G., Kosslyn, S. M., Stose, S., Thompson, W. L., & Yurgelun-Todd, D. A. (2003). Neural correlates of different types of deception: an fMRI investigation. *Cerebral cortex*, 13(8), 830-836.

Gati, I., & Ben-Shakhar, G. (1990). Novelty and significance in orientation and habituation: A feature-matching approach. *Journal of Experimental Psychology General,*119(3), 251.

Giesen, M., & Rollison, M. A. (1980). Guilty knowledge versus innocent associations: Effects of trait anxiety and stimulus context on skin conductance. *Journal of Research in Personality, 14(1)*, 1-11.

Gilson, E. Q., & Baddeley, A. D. (1969). Tactile short-term memory. *Quarterly Journal of Experimental Psychology, 21*, 180–184. doi:10.1080/14640746908400211

Gödert, H. W., Gamer, M., Rill, H. G., & Vossel, G. (2005). Statement validity assessment: Inter-rater reliability of criteria-based content analysis in the mock-crime paradigm. *Legal and Criminological Psychology, 10*, 225–245. doi:10.1348/135532505X52680

Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends Neuroscience, 15 (1)*(20-5).

Gosling, S. D., Rentfrow, P. J., & Swann Jr, W. B. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in personality*, 37(6), 504-528.

Granhag, P. A., & Hartwig, M. (2008). A new theoretical perspective on deception detection: On the psychology of instrumental mind-reading. *Psychology, Crime & Law, 14(3)*, 189-200.

Granhag, P. A., Rangmar, J., & Strömwall, L. A. (2015). Small cells of suspects: Eliciting cues to deception by strategic interviewing. *Journal of Investigative Psychology and Offender Profiling, 12(2)*, 127-141.

Grosso, A., Cambiaghi, M., Concina, G., Sacco, T., & Sacchetti, B. (2015). Auditory cortex involvement in emotional learning and memory. *Neuroscience, 23*, 45-55.

Grubin, D., & Madsen, L. (2005). Lie detection and the polygraph: A historical review. *The Journal of Forensic Psychiatry & Psychology, 16*(2), 357-369. doi:10.1080/14789940412331337353

Grubin, D., Madsen, L., Parsons, S., Sosnowski, D., & Warberg, B. (2004). A prospective study of the impact of polygraph on high-risk behaviors in adult sex offenders. *Sexual Abuse: A Journal of Research and Treatment, 16*, 209-222.

Gudjonsson, G. H. (1982). Some psychological crucials of electrodermal responses to deception. *Personality and Individual Differences, 3(4)*, 381-391.

Gustafson, L. A., & Orne, M. T. (1963). Effects of heightened motivation on the detection of deception. *Journal of Applied Psychology,*47(6), 408.

Hancock, J. T., Thorn-Santelli, J., & Ritchie, T. (2004). *Deception and design: The impact of communication technology on lying behavior.* Paper presented at the SIGCH/Conference on Human Factors in Computing Systems ACM.

Handler, M., Honts, C. R., Krapohl, D. J., Nelson, R., & Griffin, S. (2009). Integration of pre-employment polygraph screening into the police selection process. *Journal of Police and Criminal Psychology*, 24(2), 69-86.

Hemenway, D., Vriniotis, M., & Miller, M. (2006). Is an armed society a polite society? Guns and road rage. *Accident And Prevent, 38* 687-695.

Herr, P. M., Sherman, S. J., & Fazio, R. H. (1984). On the consequences of priming:assimilation and contrast effects. *Journal of Experimental Social Psychology, 19*, 323-340.

Higgins, E. T., King, G. A., & Mavin, G. H. (1982). Individual construct accessibility and subjective impressions and recall. *Journal of Personality and Social Psychology, 43(1)*, 35.

Hintz, V. B. (1990). Cognitive and consensus processes in group recognitition memory performance. *Journal of Personality and Social Psychology, 59*, 705-718.

Hira, S., & Furumitsu, I. (2002). Polygraphic examinations in Japan: Application of the guilty knowledge test in forensic investigations. I*nternational Journal of Police Science & Management*,4(1), 16-27.

Holden, E. J. (2000). Pre- and post-conviction polygraph: building blocks for the future - procedures, principles, and practices. *Polygraph, 29 (1)*, 69 - 97.

Hollingworth, A. (2005). The relationship between online visual representation of a scene and long-term scene memory. *Journal of experimental Psychology: Learning, Memory and Cognition, 31*, 396-411.

Honts, C. R. (1991). The emperor's new clothes: The application of the polygraph tests in the American workplace. *Forensic Reports, 4*, 91–116.

Honts, C. R., Amato, S. L., & Gordon, A. K. (2001). Effects of spontaneous countermeasures used against the comparison question test. *Polygraph, 30(1)*, 1-9.

Honts, C. R., Raskin, D. C., & Kircher, J. C. (1987). Effects of physical countermeasures and their electromyographic detection during polygraph tests for deception. *Journal of Psychophysiology, 55, 346-266.*

Honts, C. R., Devitt, M. K., Winbush, M., & Kircher, J. C. (1996). Mental and physical countermeasuresr educe the accuracy of the concealed knowledge test. . *Psychophysiology, 33*, 84 - 92.

Honts, C. R., & Perry, M. V. (1992). Polygraph admissibility: Changes and challenges. *Law and Human Behavior, 16(3)*, 357.

Honts, C. R., & Quick, B. D. (1995). The polygraph in 1995: progress in science and the law. . *North Dakota Law Review, 71*, 987 - 1020.

Honts, C. R., Raskin, D., & Kircher, J. C. (1994). Mental and physical countermeasuresr educe the accuracy of polygraph tests. *Journal of Applied Psychology and Aging, 79 (2)*, 252 - 259.

Honts, C. R., Raskin, D. C., & Kircher, J. C. (1985). Effects of socialization on the physiological detection of deception. *Journal of Research in Personality, 19(4)*, 373-385.

Honts, C. R., Raskin, D. C., & Kircher, J. C. (1987). Effects of physical countermeasures and their electromyographic detection during polygraph tests for deception. *Journal of Psychophysiology*.

Horvath, F. S., & Reid, J. E. (1972). The polygraph silent answer test. *Journal of Criminal Law, Criminology and Polive Science, 63(2)*, 285-293.

Hummel, J. E. (2001). Complementary solutions to the binding problem in vision: Implications for shape perception and object recognition. *Visual cognition*, 8(3-5), 489-517.

Iacono, W. G. (1991). Can we determine the accuracy of polygraph tests? . In J. R. Jennings, P. K. Ackles, & M. G. H. Coles (Eds.), *Advances in psychophysiology* (Vol. Vol. 4, pp. 201-207). London: Jessica Kingsley.

Iacono, W. G., Boisvenu, G. A., & Fleming, J. A. (1984). Effects of diazepam and methylphenidate on the electrodermal detection of guilty knowledge. *Journal of Applied Psychology, 69(2)*, 289-299.

Iacono, W. G., Cerri, A. M., Patrick, C. J., & Fleming, J. A. E. (1992). Use of antianxiety drugs as countermeasures in the detection of guilty knowledge. *Journal of Applied Psychology, 77(1)*, 60-64. doi:10.1037/0021-9010.77.1.60

Iacono, W. G., & Lykken, D. T. (1997). The validity of the lie detector: Two surveys of scientific opinion.*Journal of Applied Psychology*,82(3), 426.

Iacono, W. G., & Patrick, C. J. (1997). *Polygraph and integrity testing*. In R. Rogers (Ed.), (pp. 252-281). New York: The Guildford Press.

IMotions (2017). Galvanic Skin Response: The complete Pocket Guide.

Instrument, L. (2006). *LXSoftware User Manual* (Vol.11.1). U.S.A.

Intraub, H., & Nicklos, S. (1985). Levels of processing and picture memory: the physical superiority effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11(2)*, 284.

Jain, A., Ramta, B., Kumar, A., & Singh, K. D. (2015). A comparative study of visual and auditory reaction times on the basis of gender and physical activity levels of medical first year students. *International Journal of Applied Basic Medical Research, 5(2)*, 124-127.

Johansson, G., & Wrestling, C. (1993). Memory Representations Underlying Motor Commands Used During Manipulation of Common and Novel Objects. *Journal of Neurophysiology, 69 (6)*, 1789–1796.

Jones, E. E., & Sigall, H. (1971). The bogus pipeline: A new paradigm for measuring affect and attitude. *Psychological Bulletin, 76(5)*, 349.

Kahneman, D. (1973).*Attention and effort* (Vol. 1063). Englewood Cliffs, NJ: Prentice-Hall.

Kashy, D. A., & DePaulo, B. M. (1996). Who lies? *Journal of Personality and Social Psychology, 70*, 1037 -1051.

Kassin, S. M. (1997). The psychology of confession evidence. *American Psychologist,* 52(3), 221.

Kassin, S. M., & Gudjonsson, G. H. (2004). The psychology of confessions: a review of the literature and issues. *Psychology Science Public Interest, 5*, 33-44.

Kemp, B. J. (1973). Reaction time of young and elderly subjects in relation to perceptual deprivation and signal-on versus signal-off condition. *Developmental Psychology, 8*, 268–272.

Kircher, J. C., & Raskin, D. C. (1988). Human versus computerized evaluations of polygraph data in a laboratory setting. *Journal ofApplied Psychology, 73 (2)*, 291-302.

Kirkpatrick, L., & Cahill, L. (2003). Amygdala modulation of parahippocampal and frontal regions during emotionally influenced memory storage. *Neuroimage, 20*, 2091-2099.

Kleiner, M. (2002). *Handbook of polygraph testing*. New York: Academic Press

Kohr, R. L., & Games, P. A. (1974). Robustness of the analysis of variance, the Welch procedure and a Box procedure to heterogeneous variances. *The Journal of Experimental Education, 43(1)*, 61-69.

Kolb, B., & Whishaw, I. (2008). *Fundamentals of Human Neuropsychology* (6th edition ed.). New York: Worth Publishers.

Kozel, F. A., Johnson, K. A., Mu, Q., Grenesko, E. L., Laken, S. J., & George, M. S. (2005). Detecting deception using functional magnetic resonance imaging. *Biological psychiatry, 58(8)*, 605-613.

Kozel, F. A., Revell, L. J., Lorberbaum, J. P., Shastri, A., Elhai, J. D., Horner, M. D., & George, M. S. (2004). A pilot study of functional magnetic resonance imaging brain correlates of deception in healthy young men.*The Journal of Neuropsychiatry and Clinical Neurosciences*, 16(3), 295-305.

Krapohl, D. J. (2011). Limitations of the concealed information test in criminal cases. *Memory detection: Theory and application of the Concealed Information Test,* 151-170.

Krapohl, D., & Rosales, T. (2014). Decision Accuracy for the Relevant-Irrelevant Screening Test: A Partial Replication *Polygraph,* 43(1), 20-29.

Krapohl, D., Senter, S., & Stern, B. (2005). An exploration of methods for the analysis of multipleissue relevant/irrelevant screening data. *Polygraph, 34(1)*, 47-61.

Krapohl, D. J., McCloughan, J. B., & Senter, S. M. (2006). How to Use the Concealed Information Test. *Polygraph, 35(3)*, 34-49.

Kraut, R. E., & Poe, D. (1980). On the line: The deception judgments of customs inspectors and laymen. *Journal of Personality and Social Psychology, 39*, 784-798.

Kubis, J. F. (1962). *Studies in Lie Detection:Computer Feasibility Consideration* (AF 30(602)-2270). Retrieved from Fordham University:

Lacey, J. I., & Lacey, B. C. (1958). The relationship of resting autonomic activity to motor impulsivity. In The brain and human behavior *Proceedings of the Association for Research in Nervous and Mental Disease* (pp. 1257–1290). Baltimore, MD: Williams & Wilkins.

Lafayette Instrument Company (2009). Retrieved from http://lafayettepolygraph.com/

Langleben, D. D., Schroeder, L., Maldjian, J. A., Gur, R. C., McDonald, S., & Ragland, J. D. (2002). Brain Activity during Simulated Deception: An Event-Related Functional Magnetic Resonance Study. *Neuroimage, 15(3)*, 727-732.

Larson, J. A. (1932). *Lying and its Detection*. Chicago: University of Chicago Press.

Lavie, N. (1994). Perceptual load as a major crucial of the locus of selection in visual attention. *Perception & Psychophysics,* 56 (2), 183-197.

Lavie, N., Lin, Z., Zokaei, N., & Thoma, V. (2009). The role of perceptual load in object recognition. *Journal of Experimental Psychology. Human Perception Perform, 35*(5), 1346-1358. doi:10.1037/a0016454

Lawson, A. (1988). *An Analysis of Love and Betrayal*. New York: Basic Books.

Lederman, S. J., & Klatzky, R. L. (2009). Attention, Perception, & Psychophysics. Haptic Perception: A tutorial. *Psychophysics*. (71), 1439-1459.

Leo, R. A. (1992). From coercion to deception: the changing nature of police interrogation in America. *Crime, Law and Social Change,* 18, 35-59.

Lewis, J. A., & Cuppari, M. (2009). The polygraph: The truth lies within. *The journal of psychiatry & law,* 37(1), 85-92.

Lieblich, I., Kugelmass, S., & Ben-Shakhar, G. (1970). Efficiency of GSR detection of information with repeated presentation of series of stimuli in two motivational states. *Journal of Applied Psychology*, 59, 113-115.

Loftus, E. F. (1993). The reality of repressed memories. *American Psychologist, 48(5)*, 518-537.

Logie, R. H. (1986). Visuospatial processing in working memory. *Quarterly Journal of Experimental(38A)*, 229-247.

Logie, R. H., Cocchini, G., Della Sala, S., & Baddeley, A. D. (2004). Is there a specific executive capacity for dual-task coordination? Evidence from Alzheimer's disease. *Neuropsychology, 18*, 504- 513.

Lubow, R. E., & Fein, O. (1996). Pupillary size in response to a visual guilty knowledge test: new technique for the detection of deception. *Journal of Applied Psychology, 2*, 164–177.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjuctions. *Nature, 390*, 279-281.

Lykken, D. T. (1974). Psychology and the lie detector industry. *American Psychologist, 29*, 725-739.

Lykken, D. T. (1988). Detection of guilty knowledge: A comment on Forman and McCauley. *Journal of Applied Psychology, 73(2)*, 303-304. doi:10.1037/0021-9010.73.2.303

Lykken, D. T. (1998). *A tremor in the blood: Uses and abuses of lie detection* (2nd Ed). New York: Plenum.

Lynch, M. P. (2009). Deception and the Nature of Truth. In C. Martin (Ed.), *The Philosophy of Deception*. New York: Oxford University Press.

Lynn, R. (1966) *Attention, Arousal and the Orientation Reaction.* Oxford, U.K.: Pergamon Press.

MacKay, D. G., & Ahmetzanov, M. V. (2005). Emotion, memory, and attention in the taboo Stroop paradigm. *Psychological Sciences,* 16(1), 25-32.

MacLaren, V. V. (2001). A quantitative review of the Guilty Knowledge Test *Journal of Applied Psychology, 86*, 674–683

Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review, 87*, 252–271.

Mandler, G. (2008). Familiarity breeds attempts: A critical review of dual process theories of recognition. *Perspectives in Psychological Science, 3 (5)*, 392–401.

Marr, D. (1982). *Vision*. San Francisco: Freeman.

Marr, D., & Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society of London. Serie B, Biological Sciences, 200*, 269-294.

Martin, C. (2009). *The philosophy of deception* New York: Oxford University Press.

Maschke, G. W., & Scalabrini, G. J. (2000). *The Lie Behind the Lie Detector.*Available from: https://antipolygraph.org/lie-behind-the-lie-detector.pdf

Matsuda, I., Nittono, H., & Allen, J. (2012). The Current and Future Status of the Concealed Information Test for Field Use. *Frontiers in Psychology, 3*(532). doi:10.3389/fpsyg.2012.00532

Matte, J. A. (1996). *Forensic Psychophysiology using the Polygraph: Scientific truth verification and lie detection*. J.A.M.: Williamshire.

Matte, J. A., & Reuss, R. M. (1992). A study of the relative effectiveness of physiological data in field polygraph examinations. *Polygraph, 21(1)*, 1-22.

Marston, W. M. (1938). The Lie Detector Test. New York: Richard R. Smith.

Maxwell, S. E., & Delaney, H. D. (2004). *Designing experiments and analyzing data: A model comparison perspective*: (2nd Ed). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

McGaugh, J. L. S. (2000). Memory a century of consolidation. *Science, 287*, 248-251.

McGloin, J. M., Sullivan, C. J., Piquero, A. R., & Bacon, S. (2008). Investigating the stability of co-offending and co-offenders among a sample of youthful offenders. *Criminology, 46*, 155-188.

McGuire, W. J. (1969). The nature of attitudes and attitude changes. In G. Lindzey & H. Aronson (Eds.), *In The Handbook of Social Psychology* (second edition ed., Vol. Vol. 3.). Reading, MA: Addison-Wesley.

Mead, J. (1960). Control of respiratory frequency. *Journal of Applied Physiology, 15(3),* 325-336.

Medina, J. J. (2008). The Biology of Recognition Memory. *Psychiatric Times*.

Meijer, E. H., Bente, G., Ben-Shakhar, G., & Schumacher, A. (2013). Detecting concealed information from groups using a dynamic questioning approach: simultaneous skin conductance measurement and immediate feedback. *Front Psychology, 4(68)*, 1-6. doi:10.3389/fpsyg.2013.00068

Meijer, E. H., Smulders, F. T., & Wolf, A. (2009). The contribution of mere recognition to the P300 effect in a concealed information test. *Applied psychophysiology and biofeedback*, 34(3), 221-226.

Meijer, E. H., Smulders, F. T., & Merckelbach, H. L. (2010). Extracting concealed information from groups. *Journal of Forensic Sciences, 55*(6), 1607-1609. doi:10.1111/j.1556-4029.2010.01474.x

Meijer, E. H., & Verschuere, B. (2010). The polygraph and the detection of deception *Journal of Forensic Psychology Practice, 10(4)*, 325-338.

Meijer, E., Verschuere, B., & Ben-Shakhar, G. (2011). 16 Practical guidelines for developing a CIT.*Memory Detection: Theory and Application of the Concealed Information Test*, 293.

Meijer, E. H., Verschuere, B., & Merckelbach, H. (2010). Detecting criminal intent with the concealed information test. *Open Criminology Journal, 3*, 44-47.

Meixner, J. B., & Rosenfeld, J. P. (2011). A mock terrorism application of the P300-based concealed information test. *Psychophysiology, 48*(2), 149-154. doi:10.1111/j.1469-8986.2010.01050.x

Meyer, R. G., & Youngjohn, J. R. (1991). Effects of feedback and validity expectancy on responses in a lie detector interview. *Forensic Reports*.

Mitchell, R. W., & Thompson, N. S. (1986). *Deception, Perspectives on Human and Nonhuman Deceit*. Albany*:*Suny Press*,* 21–29.

Mohamed, M. M., & Sloane, B. F. (2006). Cysteine cathepsins: multifunctional enzymes in cancer. *Nature Reviews Cancer, 6(10)*, 764-775.

Mulligan, N. W., & Hornstein, S. L. (2003). Memory for actions: Self-performed tasks and the reenactment effect. *Memory and Cognition, 31(3)*, 412-421.

Nahari, G., & Ben-Shakhar, G. (2011). Psychophysiological and behavioral measures for detecting concealed information: the role of memory for crime details. *Psychophysiology, 48*(6), 733-744. doi:10.1111/j.1469-8986.2010.01148.x

Nakayama, M., & Kizaki, H. (1990). Usefulness of the repeated presentation of questions on the psychophysiological detection of deception. Shinrigaku kenkyu: *The Japanese journal of psychology*, 60(6), 390-393.

Nickerson, R. S. (1965). Short-term memory for complex meaningful visual configurations: A demonstration of capacity. *Canadian Journal of Psychology, 19*, 155-160.

O'Toole, D., Yuille, J. C., Patrick, C. J., & Iacono, W. G. (1994). Alcohol and the physiological detection of deception: Arousal and memory influences. *Psychophysiology, 31(3)*, 253-263.

Office of Technology Assessment (1983). *Scientific Validity of Polygraph Testing: A Research Review and Evaluation*. Washington, D. C.: U.S. Congress.

Öhman, A. (1992). Orienting and attention: Preferred preattentive processing of potentially phobic stimuli. Attention and information processing in infants and adults. *Perspectives from human and animal research,* 263-295.

Onitsuka, Y., Shiotsuka, Y., & Iramina, K. (2015). *The difference in the kinds of stimulations in CIT based on ERP*. Paper presented at the 8th Biomedical Engineering International Conference

Osugi, A. (2011). Daily application of the concealed information test: Japan. In B. Verschuere, G. Ben-Shakhar, & E. H. Meijer (Eds.), *Memory Detection: Theory and Application of the Concealed Information Test* (pp. 253-275). Cambridge: Cambridge University Press.

Packard, M. G., & Knowlton, B. (2002). Learning and Memory Functions of the Basal Ganglia. *Annual Review of Neuroscience, 25*, 563–593.

Pain, M. T., & Hibbs, A. (2007). Sprint starts and the minimum auditory reaction time.. *Journal of Sports Sciences, 25(1)*, 79-86.

Patrick, C. J., & Iacono, W. G. (1989). Psychopathy, threat, and polygraph test accuracy. . *Journal of Applied Psychology, 74 (2)*, 347-355.

Patrick, C. J., & Iacono, W. G. (1991a). A comparison of field and laboratory polygraphs in the detection of deception. *Psychophysiology, 28 (6)*, 632-638.

Patrick, C. J. (2011). Science on the rise: birth and development of the Concealed Information Test.*Memory Detection: Theory and Application of the Concealed Information Test.* Cambridge University Press, New York, New York, 3-11.

Patterson, D. A., & Hennessy, J. L. (2013).*Computer Organization and Design MIPS Edition: The Hardware/Software Interface.* Newnes Morgan Kaufmann.

Patterson, J., & Kim, P. (1991). *The Day America Told the Truth*. New York: Prentice Hall.

Pavlov, I. P. (1927). Conditional reflexes: an investigation of the physiological activity of the cerebral cortex.Oxford: Charandon Press.

Podlesny, J. A. (2003). A paucity of operable case facts restricts applicability of the guilty knowledge technique in FBI criminal polygraph examinations.*Forensic Science Communications*,5(3).

Persuh, M., Genzer, B., & Melara, R. D. (2012). Iconic memory requires attention. *Frontiers in Human Neuroscience, 6*.

Podlesny, J. A., & Raskin, D. C. (1977). Physiological measures and the detection of deception. *Psychological Bulletin, 84(4)*, 782-799. doi:10.1037/0033-2909.84.4.782

Pollina, D. A., Dollins, A. B., Senter, S. M., Krapohl, D. J., & Ryan, A. H. (2004). Comparison of polygraph data obtained from individuals involved in mock and actual criminal investigations. *Journal of Applied Psychology, 89 (6)*, 1099-1105.

Porter, G. (2007). Visual culture in forensic science. *Australian Journal of Forensic Sciences, 39(2)*, 81-91.

Quigley-Fernandez, B., & Tedeschi, J. T. (1978). The bogus pipeline as lie detector: two validity studies. *Journal of Personality and Social Psychology*, 36(3), 247.

R Core Team (2017). R: A language and environment for statistical computing. R Found. Stat. Comput. Vienna, Austria. URL http://www. R-project. org/., page R Foundation for Statistical Computing.

Radvansky, G. (2005). *Human Memory*. Boston: Allyn and Bacon.

Rajaram, S. (2011). Collaboration both hurts and helps memory: A cognitive perspective. *Current Directions in Psychological Science, 20(2)*, 76-81.

Raskin, D., Barland, G. H., & Podlesny, J. A. (1976). *Validity and reliability of detection of deception: Final report (Government Study No. Contract# 75-NI-00-0001)*. Retrieved from Salt Lake City, Utah:

Raskin, D. C., & Hare, R. D. (1978). Psychopathy and detection of deception in a prison population. *Psychophysiology, 15(2)*, 126-136.

Reed, S. D. (1993). *Effect of demographic variables on psychophysiological detection of deception outcome accuracies.* . Retrieved from USA:

Reid, J. E., & Inbau, F. E. (1977). *Truth and Deception* (2nd Ed.). Baltimore: Williams& Wilkins.

Remington, R. W., Johnston, J. C., & Yantis, S. (1992). Involuntary attentional capture by abrupt onsets. *Perception & Psychophysocs, 51(3)*, 279–290.

Reynolds, G. D., & Richards, J. E. (2008). Infant Heart Rate: A Developmental Psychophysiological Perspective. In L. A. Schmidt & S. J. Segalowitz (Eds.), *Developmental Psychophysiology*. UK: Camdridge University Press.

Riddoch, M., & Humphreys, G. H. (2001). *Handbook of Cognitive Neuropsychology* (H. P. Press Ed.).

Rosenfeld, J. P., Labkovsky, E., Winograd, M., Lui, M. A., Vandenboom, C., & Chedid, E. (2008). The Complex Trial Protocol (CTP): A new, countermeasure-resistant, accurate, P300-based method for detection of concealed information. *Psychophysiology, 45(6)*, 906-919.

Rosenthal, R., & Fode, K. L. (1996). *Dull Rats and Bright Rats. In Introducing Psychological Research* (pp. 402-407). Palgrave, London.

Royet, J. P., Zald, D., Versace, R., Costes, N., Lavenne, F., Koenig, O., & Gervais, R. (2000). Emotional responses to pleasant and unpleasant olfactory, visual, and auditory stimuli: a positron emission tomography study. *Neuroscience, 20*, 7752–7759.

Ruffman, T., Murray, J., Halberstadt, J., & Vater, T. (2012). Age-related differences in deception. *Psychology and Aging, 27(3)*, 543-549.

Sacco, T., & Sacchetti, B. (2010). Role of secondary sensory cortices in emotional memory storage and retrieval in rats. *Science, 329(5992)*, 649-656.

Saxe, L. (1994). Detection of deception: Polygraph and integrity tests. *Current Directions in Psychological Science, 3(3)*, 69–73. doi:10.1111/1467-8721.ep10770416

Saxe, L., Dougherty, D., & Cross, T. (1985). The validity of polygraph testing. *American Psychologist, March*, 355-366.

Schacter, D. L. (1987). Implicit memory: history and current status. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*, 501–518.

Schacter, D. L., & Tulving, E. (1994). What are the memory systems of 1994? *Memory systems*. Cambridge: MA: MIT Press.

Schmidt, K., Patnaik, P., & Kensinger, E. A. (2011). Emotion's influence on memory for spatial and temporal context. *Cognitive Emotion, 25(2)*, 229-243.

Schmidt, L. A., & Segalowitz, S. J. (2008). *Developmental Psychophysiology*. UK: Cambridge University Press.

Schwartz, S., Vuilleumier, P., Hutton, C., Maravita, A., Dolan, R. J., & Driver, J. (2005). Attentional load and sensory competition in human vision: Modulation of fMRI responses by load at fixation during task-irrelevant stimulation in the peripheral visual field. *Cerebral Cortex, 15*, 770 –786.

Segrave, K. (2004). *Lie Detectors: A social history*. North Carolina: McFarland & Company.

Selle, N., Verschuere, B., Kindt, M., Meijer, E., & Ben-Shakhar, G. (2016). *Orienting versus inhibition in the Concealed Information Test: Different cognitive processes drive different physiological measures*. Psychophysiology, 53(4), 579-590.

Serota, K. B., Levine, T. R., & Boster, F. J. (2010). The Prevalence of Lying in America: Three Studies of Self-Reported Lies. *Human Communication Research, 36*, 2-25.

Serota, K. B., Levine, T. R., & Boster, F. J. (2010). The prevalence of lying in America:Three studies of self-reported lies. *Human Communication Research, 36(1)*, 2-25. doi:10.1111/j.1468-2958.2009.01366.x

Seymour, T. L., & Fraynt, B. R. (2009). Time and encoding effects in the concealed knowledge test. *Applied Psychophysiology and Biofeedback*, 34(3), 177-187.

Seymour, T. L., & Kerlin, J. R. (2007). Successful detection of verbal and visual concealed knowledge using an RT-based paradigm. *Applied Cognitive Psychology, 22*, 475-490.

Seymour, T. L., Seifert, C. M., Shafto, M. G., & Mosmann, A. L. (2000). Using response time measures to assess" guilty knowledge". *Journal of Applied Psychology*, 85(1), 30.

Shepard, R. N. (1967). Recognition memory for words, sentences, and pictures. . *Journal of Verbal Learning and Verbal Behaviour, 6*, 156-163.

Sigurdsson, J. F., & Gudjonsson, G. H. (1996). The psychological characteristics of 'false confessors'. A study among Icelandic prison inmates and juvenile offenders. *Personality and Individual Differences, 20(3)*, 321-329.

Society, B. P. (1986). Report of the working group on the use of the polygraph in criminal investigation and personnel screening. *Bulletin of the British Psychological Society, 39*, 81 - 94.

Sokolov, E. N. (1963). *Perception and the conditioned reflex*. New York: Macmillan.

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs, 74*, 1–29. doi:10.1037/h0093759

Standing, L. (1973). Learning 10000 pictures. *The Quarterly Journal of experimental psychology*, 25(2), 207-222.

Standing, L., Conezio, J., & Haber, R. N. (1970). Perception and memory for pictures: Single-trial learning of 2500 visual stimuli. *Psychonomic Science, 19*, 73-74.

Steblay, N. M. (1992). A meta-analytic review of the weapon focus effect. *Law and Human Behavior, 16(4)*, 413–424.

Sternberg, R. J., & Sternberg, K. (2016). Cognitive Psychology. In C. Learning (Ed.), (7th Edition ed.).

Subra, B., Muller, D., Bègue, L., Bushman, B. J., & Delmas, F. (2010). Automatic effects of alcohol and aggressive cues on aggressive thoughts and behaviors. *Personality Social Psycholical Bullettin 36*, 1052-1057.

Sullivan, E. (2001). *The Concise Book of Lying*. New York: Picador.

Swets, J. A. (1996). *Signal Detection Theory and the ROC Analysis in Psychology and Diagnostics*. Mahwah, NJ: Erlbaum.

Synnott, J., Dietzel, D., & Ioannou, M. (2015). A review of the polygraph: history, methodology and current status. *Crime Psychology Review, 1*(1), 59-83. doi:10.1080/23744006.2015.1060080

Tarchanoff, J., 1890. Galvanic phenomena in the human skin during stimulation of the sensory organs and during various forms of mental activity. Pfl¨uger's Archiv f¨ur die gesamte *Physiologogie des Menschen und der Tiere,* 46, 46–55.

Terence, M. R., & Daniels, D. (2003). Motivation. In C. W. Borman, D. R. Ilgen, & R. J. Klimoski (Eds.), *Handbook of Psychology* (Vol. volume 12): John Wiley & Sons.

Timm, H. W. (1982). Analyzing deception from respiration patterns. *Journal of Police Science & Administration*. *10*(1), 47-51.

The British Psychology Society (2004). A review of the current scientific status and fields of application of Polygraphic Deception Detection. Final report from the BPS Working Party. Leicester, UK: British Psychological Society

Thompson, P. D., Colebatch, J. G., Brown, P., Rothwell, J. C., Day, B. L., Obeso, J. A., & Marsden, C. D. (1992). Voluntary stimulus-sensitive jerks and jumps mimicking myoclonus or pathological startle syndromes. *Movement Disorders*, 257-262.

Thompson, T. J., & Amy, B. (2009). The Habituation Effect in Personnel Security Polygraph Screening. *Polygraph, 38*, 218-222.

Tourangeau, R., Smith, T. W., & Rasinski, K. A. (1997). Motivation to Report Sensitive Behaviors on Surveys: Evidence From a Bogus Pipeline Experiment 1. *Journal of Applied Social Psychology*, 27(3), 209-222.

Trovillo, P. V. (1939). A history of lie detection. *Journal of Criminal Law and Criminology, 29,848-881.*

Tulving, E. (2000). *Concepts of memory*. New York: Oxford University Press.

Tulving, E., & Markowitsch, H. J. (1998). Episodic and Declarative Memory: role of the hippocampus. *Hippocampus, 8 (3)*, 198-204.

United States Congress, H. C. o. G. O., Subcommittee on Government Information and Foreign Operations. (1965). *Use of polygraphs as "lie detectors" by the federal government* (198). Retrieved from Washington, DC: U. S.

United States. Congress. Office of Technology Assessment (1983). *Technology, innovation, and regional economic development*. US Congress. Washington D.C:Office of Technology Assessment.

Vaidya, C. J., Zhao, M., Desmond, J. E., & Gabrieli, J. D. (2002). Evidence for cortical encoding specificity in episodic memory: memory-induced re-activation of picture processing areas. *Neuropsychologia, 40(12)*, 2136-2143.

Van Mastrigt, S. B., & Farrington, D. P. (2011). Prevalence and characteristics of co-offending recruiters. *Justice Quarterly, 28(2)*, 325-359.

Vandenbosch, K., Verschuere, B., Crombez, G., & De Clercq, A. (2009). The validity of finger pulse line length for the detection of concealed information. *International Journal of Psychophysiology,* 71(2), 118-123.

Vernham, Z., Granhag, P. A., & Mac Giolla, E. (2016). Detecting deception within small groups: A literature review. *Frontiers in psychology, 7*, 1012.

Verschuere, B., Crombez, G., De Clercq, A., & Koster, E. H. W. (2004). Autonomic and behavioral responding to concealed information: Differentiating orienting and defensive responses. *Psychophysiology, 41*(3), 461-466. doi:10.1111/j.1469-8986.00167.x

Verschuere, B., Crombez, G., Koster, E. H. W., & De Clercq, A. (2007). Antisociality, underarousal and the validity of the Concealed Information Polygraph Test. *Biological Psychology, 74*, 309-318.

Verschuere, B., Crombez, G., Koster, E. H., Van Bockstaele, B., & De Clercq, A. (2007). Startling secrets: Startle eye blink modulation by concealed crime information.*Biological Psychology*,76(1-2), 52-60.

Verschuere, B., Ben-Shakhar, G., & Meijer, E. (2011). *Memory detection: Theory and application of the Concealed Information Test.* Cambridge University Press.

Victoroff, J. (2005). The mind of the terrorist: A review and critique of psychological approaches.*Journal of Conflict resolution*,49(1), 3-42.

Vrij, A. (2000). *Detecting lies and deceit: The psychology of lying and the implications for professional practice*. Chichester: Wiley.

Vrij, A. (2008). *Detecting Lies and Deceit: Pitfalls and Opportunities* (2nd ed.). UK: John Wiley & Sons, Inc.

Vrij, A., Akehurst, L., & Morris, P. (1997). Individual differences in hand movements during deception. *Journal of Nonverbal Behavior, 21(2)*, 87.

Vrij, A., Fisher, R., Mann, S., & Lean, S. (2006). Detecting deception by manipulating cognitive load. *Trends in Cognitive Science, 10(4)*, 141-142.

Vrij, A., Jundi, S., Hope, L., Hillman, J., Gahr, E., Leal, S., & Granhag, P. A. (2012). Collective interviewing of suspects. *Journal of Applied Research in Memory and Cognition, 1(1)*, 41-44.

Vrij, A., Mann, S., & Leal, S. (2013). Deception traits in psychological interviewing. *Journal of Police and Criminal Psychology, 28(2)*, 115-126. doi:10.1007/s11896-013-9125-y

Vrij, A., Mann, S., Leal, S., & Granhag, P. (2010). Getting into the minds of pairs of liars and truth tellers: An examination of their strategies. *Open Criminology Journal, 3(1)*, 17-22.

Waid, W.M., Orne, E. C., Cook, M.R. and Orne, M.T. (1978). Effects of attention, as indexed by subsequent memory, on electrodermal detection of information. *Journal of Applied Psychology*, 18, 470-476.

Waid, W. M., Orne, E. C., Cook, M. R., & Orne, M. T. (1981). Meprobamate reduces accuracy of physiological detection of deception. *Science, 212(4490)*, 71-73.

Waid, W. M., & Orne, M. T. (1980). Individual differences in electrodermal lability and the detection of information and deception. *Journal of Applied Psychology, 65(1)*, 1.

Walsh, John E. (1996).*Unraveling Piltdown: The Science Fraud of the Century and its Solution*. Random House.

Ward, J. (2009). *The Student's Guide to Cognitive Neuroscience*: Psychology Press.

Warr, M. (2002). *Companions in Crime: The Social Aspects of Criminal Conduct.* . Cambridge: Cambridge University Press.

Wegner, D. M. (1987). Transactive memory: A contemporary analysis of the group mind. In B. Mullen & G. R. Goethals (Eds.), *In Theories of group behavior.* (pp. 185-208). New York, NY: Springer.

Weingarten, E., Chen, Q., McAdams, M., Yi, J., Hepler, J., & Albarracín, D. (2016). From primed concepts to action: A meta-analysis of the behavioral effects of incidentally presented words. *Psychological Bulletin, 142*, 472–497.

Wheeler, B. (2009). Monkeys crying wolf? Tufted. *The Royal Society, 276*, 3013–3018.

Williams, H. L., & Conway, M. A. (2008). *Memory in the Real World* (3rd ed.). UK.

Yagi, Y., Coburn, K. L., Estes, K. M., & Arruda, J. E. (1999). Effects of aerobic exercise and gender on visual and auditory P300, reaction time, and accuracy. *Eurean Journal of Applied Physiolological Occupational Physiology, 80(5)*, 402-408.

Yantis, S., & Jonides, J. (1996). Attentional Capture by Abrupt Onsets: New Perceptual Objects or Visual Masking? *Journal of Experimental Psychology: Human Perception and Performance, 22 (6)*, 1505–1513.

APPENDICES

Appendix A. Leaflet for recruiting participant around campus and city centre

University of
HUDDERSFIELD

DO YOU THINK YOU CAN YOU CHEAT THE LIE DETECTOR????

Postgraduate students from the department of Psychology are currently conducting a study regarding people’s ability to cheat the polygraph and are in need of participants.

If you decide to participate and give us 30 min of your time:

- **You will have a free coffee and biscuits (after the experiment)**

- **You will be able to know if your persuasion skills can defeat the polygraph**

- **You will know the extent of your ability to control your physical responses when lying**

For more information contact:

Anita Fumagalli U1259371@hud.ac.uk

Appendix B. Participant Information Sheet

Detection of Deception with the Concealed Information Test

PROJECT INFORMATION SHEET

You are being invited to take part in this study, which involves the polygraph machine. Before you decide to take part it is important that you understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with me if you wish. Please do not hesitate to ask if there is anything that is not clear or if you would like more information.

What is the study about?

The purpose of this study is to test if the polygraph or otherwise lie detector is able to detect deception. The polygraph machine examines four physiological responses; these are heart rate, blood pressure, breathing, and sweat levels known as galvanic skin response (GSR). The general opinion is that the polygraph machine is able to detect deception by measuring these four physiological responses and the changes that occur during the test. For this study we will be using the Concealed Information Test to examine the validity of the polygraph, which is used to assess whether suspects conceal 'guilty knowledge' by asking them questions with multiple choice answers while measuring the physiological response to each possible answer read out. The polygraph assumes that if there is a wavering in polygraph readings when an answer is read out, the suspect possesses a 'guilty knowledge'.

Do I have to take part?

It is your decision whether or not you take part. If you decide to take part you will be asked to sign a consent form, and you will be free to withdraw at any time and without giving a reason. A decision to withdraw at any time, or a decision not to take part, will not affect you.

What will I need to do?

If you agree to take part in the research you will be asked to complete two tasks, the first of which is smashing an egg with an object, the next is bursting a balloon with an object. You will then be attached to the polygraph machine that includes a blood pressure cuff on your arm, two breathing receivers, one on top of the chest and one just under the chest. Three electrodes will be attached to your fingers two measuring pulse one measuring skin response. The second researcher, who will not be aware what object you used for the tasks, will enter the room and it is your job to deceive him/her. You will be shown images on a screen in front of you and asked various questions. You MUST answer NO to all questions regardless of being deceitful or truthful, this will take approximately 25-30 minutes. Your physiological response will be recorded to examine your reaction after seeing the visual stimulus and after being asked the question by the researcher.

Will my identity be disclosed?

All information disclosed within the interview will be kept confidential, except where legal obligations would necessitate disclosure by the researchers to appropriate personnel.

What will happen to the information?

All information collected from you during this research will be kept secure and any identifying material, such as names will be removed in order to ensure anonymity. It is anticipated that the research may, at some point, be published in a journal or report. However, should this happen, your anonymity will be ensured.

Who can I contact for further information?

If you require any further information about the research, please contact me on:

The researcher: Anita Fumagalli

E-mail: U1259371@unimail.hud.ac.uk

Dissertation Supervisor: John Synnott

J.P.Synnott@hud.ac.uk

Appendix C. Participant Consent Form

Participant Consent form:

Summary:

The aim of this research is to examine the validity of the polygraph machine in detecting deception. Participants after completing the priming stage, will be attached to the polygraph machine via blood pressure cuff, pneumographs (breath rate monitors), heart rate monitors and a galvanic skin response measure on the fingertips and they will be asked a series of questions to which the participant must answer “No” to all. The experiment will take approximately 30 minutes to complete.

Please tick boxes that apply:

	Yes	No
I have read and understood the Participant Information Sheet		
Questions about your participation have been answered to your satisfaction		
You are fit to participate		
You are taking part in this study voluntarily		
I understand that my identity will be protected and that all data will be anonymous		

I understand that I can withdraw from the study at any time without having to give an explanation		
I agree to the data (in line with conditions outlined above) being archived and used by other bona fide researchers		

To be completed by the **participant**:

Name: ____________________________

Age: _______________________________

Address any medical conditions _____________________

To be completed by the **researcher**:

Name: _____________________________

Signature: ___________________________

Date: ________________________________

Appendix D. Debrief Study I and II

<u>Debrief</u>

Project Title – Investigating the validity of the polygraph.

Thank you for participating in this study. The study was conducted for my PhD project on the detection of knowledge of a task by using the Concealed Information Test (CIT).

The main aim of this study was to explore whether the CIT was able to detect relevant information after the performance of a simple task.

It was hypothesised that an individual's physiological arousal will increase when exposed to the stimuli. Subsequently, the polygraph test then measured your physiological changes, including your heart rate, respiratory rate and Galvanic Skin Response (GSR).

Should you feel distressed about the nature of this study, I have provided contact information for a service you might find useful below:

University of Huddersfield's wellbeing service:

Telephone: 01484 472675

Email: studentwellbeing@hud.ac.uk

Thank you once again for your participation in this study.

If you have any further queries, please do not hesitate to contact me:

Anita.fumagalli@hud.ac.uk

Appendix E. Debrief Study IV

Debrief

Project Title - Detecting Knowledge of a Mock Crime using the Concealed Information Test

Thank you for participating in this study. The study was conducted for my PhD project on the detection of knowledge of a mock crime using the concealed information test in a group setting.

The main aim of this study was to explore whether the role of a group could have an influence on physiological reactions. Similarly, an additional aim was to investigate the validity of the polygraph test within a criminal investigation and examine and recognise issues surrounding the validity of the polygraph examination, with applicability to the Concealed Information Test question series.

The stimuli you were exposed to be related to planning a mock criminal attack within your small groups. It was hypothesised that an individual's physiological arousal will increase when exposed to the stimuli. Subsequently, the polygraph test then measured your physiological changes, including your heart rate, respiratory rate and Galvanic Skin Response (GSR).

Please circle any of the below choices if they are significant to you in anyway or if you have previously visited them.

London, UK	Heathrow Airport
New York, USA	Empire State Building
Paris, France	Eiffel Tower
Madrid, Spain	Bernabeu Stadium
Brussels, Belgium	Brussels Grand Palace

Should you feel distressed about the nature of this study, I have provided contact information for a service you might find useful below:

University of Huddersfield's wellbeing service:

Telephone: 01484 472675

Email: studentwellbeing@hud.ac.uk

Thank you once again for your participation in this study.

If you have any further queries, please do not hesitate to contact me:

Anita.Fumagalli@hud.ac.uk

Appendix F. Study I questionnaire

1. Do you attend the University of Huddersfield?
2. Concerning the object used for the task, was it an earring?
3. Was it a fork?
4. Was it a pen?
5. Was it a scissor?
6. Was it a paperclip?
7. Was it a pin?
8. Was it a screwdriver?
9. Is it sunny today?
10. Concerning the object used for the task, was it a stone?
11. Was it a hammer?
12. Was it a glass bottle?
13. Was it a screwdriver?
14. Was it a wooden roll?
15. Was it a knife?
16. Was it a gun?
17. Did you go to University today?
18. Concerning the object used for the task, was it a key?
19. Was it a screwdriver?
20. Was it a paperclip?
21. Was it a pen?
22. Was it a pin?
23. Was it a fork?
24. Was it a scissor?
25. Have you ever been in Ramsden Building?
26. Concerning the object used for the task, was it a stone?
27. Was it a gun?
28. Was it a wooden roll?
29. Was it a screwdriver?
30. Was it a knife?
31. Was it a hammer?
32. Was it a wooden roll?
33. Do you live in England?
34. Concerning the object used for the task, was it an earring?
35. Was it a scissor?

36. Was it a fork?
37. Was it a pin?
38. Was it a screwdriver?
39. Was it a pen?
40. Was it a paperclip?
41. Are you an undergraduate student?
42. Concerning the object used for the task, was it a stone?
43. Was it a wooden roll?
44. Was it a hammer?
45. Was it a glass bottle?
46. Was it a gun?
47. Was it a screwdriver?
48. Was it a gun?

Appendix G. Study II questionnaire

1. Do you attend the University of Huddersfield?
2. Concerning the object used for the task, was it an earring?
3. Was it a fork?
4. Was it a pen?
5. Was it a scissor?
6. Was it a paperclip?
7. Was it a pin?
8. Was it a screwdriver?
9. Is it sunny today?
10. Concerning the object used for the task, was it a shoe?
11. Was it a hammer?
12. Was it a book?
13. Was it a wooden roll?
14. Was it a hand?
15. Was it a knife?
16. Was it a bottle of water
17. Did you go to University today?
18. Concerning the object used for the task, was it a key?
19. Was it a screwdriver?
20. Was it a paperclip?
21. Was it with a pen?
22. Was it with a pin?
23. Was it with a fork?
24. Was it a scissor?
25. Have you ever been in Ramsden Building?
26. Concerning the object used for the task, was it a can?
27. Was it a bottle of water?
28. Was it a hand?
29. Was it a book?
30. Was it a knife?
31. Was it a hammer?
32. Was it a wooden roll?
33. Do you live in England?
34. Concerning the object used for the task, was it an earring?
35. Was it a scissor?
36. Was it a fork?
37. Was it a pin?
38. Was it a screwdriver?
39. Was it a pen?
40. Was it a paperclip?
41. Are you an undergraduate student?
42. Concerning the object used for the task, was it a shoe?
43. Was it a knife?
44. Was it a hammer?
45. Was it a bottle of water?
46. Was it a wooden roll?

47. Was it a book?
48. Was it with a hand?

Appendix H. Study III and IV questionnaire

1. Do you attend the University of Huddersfield?
2. Concerning the criminal attack, was it planned in Rome?
3. Was it planned in New York?
4. Was it planned in London?
5. Was it planned in Paris?
6. Was it planned in Madrid?
7. Was it planned in Brussels?
8. Is it sunny today?
9. Concerning the target of the attack, was it planned in the St Peter's Basilica?
10. Was it planned in the Heathrow Airport?
11. Was it planned in the Empire State Building?
12. Was it planned in the Brussels Grand Palace?
13. Was it planned in the Eiffel Tower?
14. Was it planned in the Bernabeu Stadium?
15. Did you go to University today?
16. Concerning the weapons used, were they used knifes?
17. Were they used bombs?
18. Were they used guns?
19. Were they used nuclear missiles?
20. Were they used chemical weapons?
21. Were they used suicide vests?
22. Have you ever been in the Ramsden Building?
23. Concerning the criminal attack, was it planned in Rome?
24. Was it planned in Madrid?
25. Was it planned in Paris?
26. Was it planned in Brussels?
27. Was it planned in London?
28. Was it planned in New York?
29. Do you live in England?
30. Concerning the target of the attack, was it planned in the St Peter's Basilica?
31. Was it planned in the Brussels Grand Palace?
32. Was it planned in the Heathrow Airport?
33. Was it planned in the Bernabeu Stadium?
34. Was it planned in the Empire State Building?
35. Was it planned in the Eiffel Tower?
36. Are you an undergraduate student?
37. Concerning the weapons used, were they used knifes?
38. Were they used chemical weapons?
39. Were they used bombs?

40. Were they used suicide vests?
41. Were they used nuclear missiles?
42. Were they used guns?

Appendix I: Expanded Section: Short and Long-term memory

Tulving (2000) suggested that memory has different sub memory systems. This idea was also supported by Sternberg and Sternberg (2016), which defined these subsystems as memory operations. Encoding, storage and retrieval, each of which has its stage in processing information. Encoding is described as that operation able to transform physical and sensory input into a representation that can be inserted into the memory. These physical and sensory inputs can be received in four different ways: acoustic, tactile, visual and semantic (Sternberg & Sternberg, 2016). Once encoding has occurred, information needs to be stored into the memory system (Eysenck, 2015), which leads to the second operation, storage. This process refers to how you store encoded information in memory. The final operation consists in the retrieval, which describes how you gain access to information stored into the memory (Sternberg & Sternberg, 2016).

Atkinson and Shiffrin (1968) supported this memory mechanism, developing a multi-model of the memory structure. They proposed three different memory units: sensory store, which consists of the persistence of sensory information for a moment after its perception. Then, the short-term store, where the information from the sensory store is sent. This type of store unit has a limited capacity that consists of 5-7 items. Finally, the last store is defined as the long-term, which has a more significant capacity than the short-term store. In fact, the long-term one can store a bigger amount of information for an unlimited duration (Atkinson & Shiffrin, 1968; Brady, Konkle, & Alvarez, 2011).

Although this represents a critical approach, different critiques have been reported. The structure of this model, in fact, was considered an oversimplification. For instance, it has been found that the long-term memory is made by different sub-components, like the episodic and semantic memory (Schacter & Tulving, 1994). The episodic memory contains all the autobiographical events, such as times, places and other contextual information. The semantic memory refers to general knowledge that we have acquired during our lives (Tulving & Markowitsch, 1998). These two sub-memories together make the declarative memory (or explicit memory). The other main component of the long-term memory consists in the procedural memory (or implicit memory), which it is usually captured and used unconsciously (Schacter, 1987). One of the main processes that help to consolidate memory information in the long-term memory is memory consolidation. This process can stabilise information after the acquisition through two sub-processes: synaptic consolidation, occurring during the first hours after learning, and systems consolidation, which consolidates the information over a period of weeks to years (Dudai, 2004).

Baddeley and Hitch (1974) have also applied different improvements to the concept of short-term memory. They replaced and developed a new model, called working memory, providing a full explanation of the structure and functions of the short-term memory. Their new model includes three-component system: the central executive, which is assisted by two subsystems that are involved with acoustic and verbal information. Then the phonological loop, which deals with acoustic information and finally the visual-spatial scratchpad, which concerned with visual and spatial information (Baddeley & Hitch, 1974).

The central executive is considered the most complex component of the working memory. This component is regarded as the control system because of its several attentional processes. In fact, it was assumed that the executive capacities involve the ability to divide attention between two essential stimulus streams (Baddeley & Hitch, 1974). Furthermore, it can switch between tasks, increasing or reducing the attention concerning the complexity of the task presented (Baddeley, 1986). Similarly, the phonological loop is divided into two components, a phonological store and an articulatory rehearsal system. The phonological store is associated with the perception of speech. It can store auditory stimuli in the order in which they are heard. This information can decay quickly if they do not actively refresh it. The refreshing process, thus, is controlled by the articulatory rehearsal system, strongly associated with the production of speech (Baddeley & Hitch, 1974).

Another important component of the working memory is the episodic buffer. It is the responsible for holding integrated episodes or chunks in a multidimensional code (Baddeley, 2012). Its main duty consists in linking information, to form integrated units of visual-spatial and verbal information in sequential order (Baddeley, 2000). When working as buffer store, the episodic buffer also links the working memory to the long-term memory. This role is related to its capacity of holding multidimensional representations, even if it has a limited capacity (Baddeley, 2012). This component is crucial for deceptive processes, as it allows to manage visual and verbal information and their sequence, which can be vital when trying to building a false story.

The last component of the working memory is the visuospatial sketchpad, which can temporarily maintain and manipulate visuospatial information. It provides a virtual environment for physical simulation, calculation visualisation and optical memory recall (Baddeley & Hitch, 1974). Furthermore, it plays a crucial role in spatial orientation and the explication of the visuospatial problems (Baddeley, 1986). As part of the working memory, it holds the information acquired during the initial process; the visual information, such as a place, a person's face, can be recollected by the long-term memory (Baddeley, 2012). Furthermore, the sketchpad creates an interface between spatial and visual information, which can be accessed through the long-term memory or the senses. In essence, it allows a range of visible information channels to be bound together with similar information of a motor, tactile or haptic nature (Baddeley, 2012).

One of the most interesting aspects researched over the years concerns the intention to establish if a separation of visual and spatial components exists. Although it is difficult to evaluate this thin border, several researchers suggested that it is possible to talk about this differentiation (Logie, 1986; Logie, Cocchini, Della Sala, & Baddeley, 2004). Regarding this, Logie (1986) provided an overview of this approach demonstrating a dissociation of visual and spatial working memory. In fact, he suggested that two different elements constitute the visuospatial sketchpad. The visual cache, which its capacity concerns the storing of information about colour and shape, and the inner scribe, which deals with movement and spatial information.

While this approach tries to analyse memory systems, generalising over representational contents, research on visual perception attempted to determine what is being represented, generalising across processes (Schacter & Tulving, 1994). Therefore, the intersection between memory and vision become a fascinating field of research, because the analysis concerns both the nature of stored visual representations and memory processes (Brady et al., 2011).

Appendix J. <u>Expanded Section: Visual-Working memory</u>

Research about visual working memory has focused on the capacity of the system to contain visual information. The interest is related to the idea that there is a relationship between the limited capacity of this system and individual differences in processing information, such as fluid intelligence, reading comprehension and academic achievement. Therefore, the working memory may be a core cognitive ability that underlies our ability to process information (Brady et al., 2011). In order to analyse the capacity of a memory system, its dimension needs to be evaluated concerning the number of items that can be stored and regarding the fidelity with which each item can be stored (Brady et al., 2011).

In order to analyse the capacity of visual working memory, Luck and Vogel (Luck & Vogel, 1997) conducted a study with the aim of estimate the working memory capacity for features and conjunction of elements. Observers were asked to complete a change detection task, which consists of a presentation of an array of coloured squares, which needed to be remembered. On each trial, the squares disappeared for about 1 second and reappeared with all the items the same as either before or with some changes. Observes had to determine whether the square displayed was the same as the previous or whether it had changed. Results show that in those cases in which there were fewer three or four items on display, observers were able to detect changes most of the time. However, when the number of items increased beyond four, the observers' performance decreased (Luck & Vogel, 1997). As previously mentioned, the standard by which how well each object can be remembered has its importance. In fact, it is essential to understand not just how many items can be stored, but also the quantity of information about them can be remembered.

With the purpose of analysing this aspect, Alvarez and Cavanagh (Alvarez & Canavagh, 2004) conducted a study in which observers were asked to remember objects from categorically different colours (low information load) to perceptually similar 3D cubes (high information load). Results showed that observes could quickly detect a change between cubes when only a single cube was remembered, but they could not identify the same difference when they tried to remember four cubes. This result suggests that encoding additional items reduced the resolution with which each item could be recalled, confirming the idea that there is an information limit on the working memory (Brady et al., 2011).

Thus, this demonstrated that there is an exchange between the number of objects that can be stored and the fidelity with which each item in the store; the information load per item determines this relationship. In other words, the more information that had to be remembered from an individual item, the fewer the total number of items that could be stored with sufficient resolution (Alvarez & Canavagh, 2004).

Appendix K. Expanded Section: Visual-Recognition memory

Visual recognition memory is a fundamental form of memory, which emerges in early infancy. It is a subcomponent of the declarative memory. Its functions relate to defining features namely and reflecting a core biological adaptation. However, its central role consists in responding to novelty and recognising previously encountered events, objects or people (Medina, 2008). This memory has two different processes. The first one is the recollection, which consists in the retrieval of information associated with a previously experienced event. The second one relates to familiarity, which is the feeling that the event was already experienced, without recollection occurrence. Familiarity is a fast-automatic process, while the recollection is slow and controlled (Mandler, 1980).

The neurons involved in these two processes are located in a specific brain area called perirhinal cortex, which is part of the temporal lobe, close to the hippocampus (Mandler, 1980). There are two typologies of neurons: the novelty neurons, which the main characteristic consists of responding to novel items that have not seen before. However, they respond much less to familiar items. This reduction in the neurons 'activity occurs rapidly and can last for an extended period of time. The second category of neurons is familiarity with neurons. They lose their responsiveness when an image is seen at different times (Fahy, Riches, & Brown, 1993).

The dual-process mode/theory supports the consideration of familiarity and recollection as two distinct processes. However, a debate developed in this area of research, which starts from the idea that recollection is merely a stronger version of familiarity. Thus, single-process models consider recognition as a continuum, ranging from weak to strong memories (Curran, Debuse, Woroch, & Hirshman, 2006). Conversely, the dual process theories differentiate two types of recognition; the first one relates to the ability to recognise that some object or event has been encountered before, while the second one can identify what that object or event was (Mandler, 2008). Although this represents an interesting research area, it is considered controversial, because of the difficulty in trying to obtain a separate empirical estimation of these two processes. At the moment, neuroscientific research has not solved this controversy, although the dual-process is considered the most popular and valid interpretation (Medina, 2008).

Concerning visual perception of physical objects or people, the discussion tends to be slightly different. The ventral visual-processing stream consists of the areas of the occipital, occipitotemporal and temporal regions that are solely devoted to processing visual stimuli and are unresponsive to information from other modalities. Information from the three-dimensional world is projected into the retinas, which are just two-dimensional planes (Banich, 1999).

Therefore, the brain must reconstruct and put the third dimension back into the pictures solely by 2-D information. Although the same object can be projected into the retina in different ways, the brain must interpret the object as being the same, regardless of variations in retinal size, retinal position and orientation.

In order to recognise objects, the brain must be able to not only distinguish among objects but also identify an object regardless of variations in the condition under which it is viewed. The response of cells in inferotemporal cortex is unaffected by changes in retinal position, retinal size, and orientation: this represents the neuronal mechanism that allows us to recognise an object from various perspectives (Banich, 1999).

The area of visual space to which the cell is sensitive is called receptive field. If a stimulus falls within the receptive field, the cell fires. On the other hand, if the same stimulus falls outside the receptive field, the cell does not fire. Each section of the primary visual cortex is responsible for receiving from only a precise region in space. A large receptive field that surrounds almost all of the visual field is used for object recognition because it allows the identification of an object regardless of where it is located in space. Therefore, this large receptive field enables the cell to react to objects from their global shape, rather than just other specific features (Banich, 1999).

Having a receptive field that always includes the central region of visual space also enables object recognition. The part of the retina, which possesses the highest density of sensory receptors, is the fovea. It also receives information about the central area of visual space. This region provides for the highest possible resolution when attempting to recognise an object. Finally, the colour plays an important role that aids in object recognition. It represents an essential attribute of cells because it allows us to separate an object from the background in which it is surrounded. This process is defined as Figure-Ground Separation (Banich, 1999).

Recognizing is one of the underlying mechanisms at the base of our cognitive functions. It was suggested by an early and influential class of models that the fundamental goal of vision was to reconstruct the three- dimensional structure of objects and their spatial relationships (Marr & Nishihara, 1978). This approach, which was later elaborated by Biederman (1987), assumed that every given object could be described concerning generic three-dimensional components and their spatial relationships. Therefore, the primary goal of vision is basic-level recognition, without respect to image characterises that can be raised from lighting, viewpoint and other variables.

One of the most influential theories of vision is from Marr (Marr, 1982). He theorised that the brain constructs a series of representations of the world by making inferences derived from necessary retinal information. He summarises the vision process into different stages:

- -Primal Sketch: A scene is perceived with its fundamental elements such as edges, contract, orientation and bars.
- -2½ D Sketch: Information like figure-ground discrimination and surface texture are added. At this stage, the perception of the stimulus is still viewer-centred in that internal representation of the observer's viewpoint.
- -3D Sketch: An object-centred representation of the object is established. This representation is independent of the specific view. Also, a full structural description and knowledge of the object become available.
- -Semantic Interpretation: Meaning is attributed to the stimulus.

The recognition of an object it is an operation that needs to be inserted in a greater context. Firstly, the individual needs to select information that is considered relevant from those that will remain unsolved during the flow of information processing. The locus of selection is the primary concern of attention research, and it has to stimulate different theories and studies.

Lavie (1994) tried to solve the dispute whether attention affects information processing at early stages of perception or only at later stages. The initial selection approach states that perception is a limited process, which requires selective attention to proceed. According to this view, selection occurs early after rudimentary analysis of physical features that are used to distinguish between selected and non-selected stimuli. Therefore, in this case, unattended stimuli are not fully perceived. In contrast, the late-selection approach claims that perception is an extensive process that performs automatically, without the need for selection. In this case, selection only occurs late in process, after full perception.

Lavie's perceptual load theory combines the central assumptions of these two approaches. The theory stipulates that perception has a limited capacity, but it operates automatically on all information within its capacity. Therefore, several objects can be recognised before perception gets overloaded. Capacity limits can be reached either early or late in the processing, depending on the level of perceptual load in the task (Lavie, Lin, Zokaei, & Thoma, 2009). The locus of selection is another aspect that was clarified, the instruction to ignore task-irrelevant distracter objects may not be enough to make these objects unattended. According to this approach, the allocation of attention is automatic. Therefore, it cannot be restrained at will.referably, the level of perceptual load in any given tasks can determine whether task-irrelevant objects can be successful ignored (Lavie, 1994). However, the type of task could determine the level of perceptual load. In functions that involve only low perceptual load, spare capacity from task-relevant processing will unintentionally spill over to the perception of task-irrelevant objects, occurring a late selection. In this case, task-irrelevant objects will not be recognised. In contrast, when a task requires full perceptual capacity in situations of high perceptual load, no capacity of spill over to task-irrelevant processing will be left available, occurring an early selection (Lavie et al., 2009).

Different studies that have used different types of load manipulation have supported the perceptual load model (Lavie, 1994). Similar results have been found, increasing the level of perceptual load in task-relevant processing reduces significantly the extent to which task-irrelevant stimuli are perceived, unconsciously and consciously. Furthermore, they assumed that increased perceptual load minimizes the length to which task-irrelevant stimuli produce behavioural interference effects (Lavie et al., 2009).

The load theory was also supported by neuroimaging studies, which found that the level of perceptual load determines the level of visual cortex activity related to the presence of various task-irrelevant stimuli. Examples of these stimuli were meaningless checkerboards patterns, moving dots and meaningful pictures of an object (Schwartz et al., 2005).

On the other hand, the hybrid object recognition model gives a different perspective. Hummel (2001) considers the role of attention in object recognition, but he suggests that it has a more restricted role. According to the hybrid theory, attention is only needed to form “analytic” object representations that involve a structural and invariant description of the spatial relations of object parts. Contrarily to the load perception theory, this approach suggests that distracter object recognition would remain unaffected by the level of perceptual load in the task when the distracter objects are always repeated in the same view (Lavie et al., 2009).

As discussed above, different mechanisms are involved in the perception and recognition of objects. However, the last crucial stage consists in the elaboration of this information and its storing into the memory. According to a neuropsychological model proposed by Riddoch and Humphreys (Riddoch & Humphreys, 2001), the stages of an object recognition can be divided into four phases. During the first one, the primary object features are processed, such as shape and colour. Then, these essential features are grouped based on similarity. During the third stage, the visual representation of the object is matched with structural information contained in the memory. Finally, semantic attributes are applied to the visual representation, attributing meanings, thus recognition.

In according to early research in the literature (Nickerson, 1965; Shepard, 1967; Standing, Conezio, & Haber, 1970), our ability to recognise pictures that we have seen before, is impressive. Nickerson (1965) demonstrates that individuals can distinguish large sets of old images from new distractor pictures at high levels of accuracy. In the same way, Shepard (1967) asked participants to recognise 612 photographs; 98% of the answers were correct. Standing (1973) reported even more impressive results from his studies; he presented 10000 different pictures on a two alternative recognition test. The mean percentage of accuracy accounted for 83%, which mean that almost 6600 photos were recognised successfully. However, distractors used in these experiments were specially chosen to be profoundly different from studies images. Participants might have remembered the pictures shown by applying visual coding such as shape, colour, orientation, texture, and so on, or by maintaining verbal descriptions of picture content or even by keeping representations conceptual of picture identity (Hollingworth, 2005).

Appendix M. Extended section: Sensory Memory

Sensory information is vital for our everyday life as it allows individuals to experience the same stimulus through different sensory channels. This type of memory is considered to be an automatic response, therefore, outside the cognitive control (Carlson, 2010). Stimuli from eyes, nose, ears mouth and skin are received and send to the sensory cortex of the brain (Grosso, Cambiaghi, Concina, Sacco, & Sacchetti, 2015).

Sensory memory allows individuals to recall impressions of sensory information once the original stimulus has occurred (Coltheart, 1980). The typologies of sensory memories consistent of five elements, which represent the storage for all the information recorded through the main human senses: sight, hearing, taste, smell and touch. All the sensory information is stored with specific modality; for instance, echoic memory is used exclusively for storing auditory information, while the haptic memory is used for storing solely tactile information. There are a few common features in each sensory modality such as the storage capacity, which appears to be very limited and able to store just for a concise period of time. However, there are also quite essential differences that characterise them, such as temporal duration; in fact, every store has slightly different durations (Persuh, Genzer, & Melara, 2012).

Iconic memory is the sensory storage able to register visual information. It is part of optical memory system, which also includes visual short and long memory. This store is considered to be brief, pre-categorical, with a high capacity of deposit (Sperling, 1960). Its role is vital as it provides a comprehensible representation of the entire visual perception for a brief period of time (Dick, 1974). The iconic memory consists in two components:

- The visible persistence is the sensational impression that a visual image endures even after its physical offset, due to neural persistence. It provides a very brief (150ms) pre-categorical visual representation of the physical figure, which is generated by the sensory system. With this type of memory component, the longer the physical stimulus is given, the faster the visual representation decays in memory (Coltheart, 1980). Furthermore, the luminance of the stimulus can affect the visible persistence; therefore, due to the involvement of photoreceptors that needs to be activated in the visual cortex, when the brightness of a stimulus increases, the duration of the visible persistence decreases (Dick, 1974).

- The informational persistence consists of the actual information regarding the stimulus that remains after its physical offset. This component is highly affected by the duration of the stimulus; therefore, the longer the physical stimulus is presented, the longer the visual representation is kept in memory. The informational persistence also represents non-visual components, such as the general characteristic of the image and its spatial location (Coltheart, 1980).

The haptic memory can recollect information gained by touching after a stimulus has been presented. One of the primary functions consists of assessing the forces needed for gripping or interacting with a familiar object; this memory works better when stimuli are applied to those areas that are more sensitive to touch, such as hands (Johansson & Wrestling, 1993). It has been observed that the haptic memory stores similar to the visual sensory memory; in fact, information acquired by touching are memorized for a short period and tend to decay after two/three seconds, with a capacity of approximately four to five items (Lederman & Klatzky, 2009; Sperling, 1960). The reliability of the memory after the stimulus has been removed from the skin last approximately ten seconds; after this period the information traces become weak and start to be forgotten (Gilson & Baddeley, 1969).

The echoic memory is specialized in retaining auditory information. Compared to the iconic and haptic memory, the echoic can store for a more extended period. Information is acquired from the ears, then processed and understood (Carlson, 2010). Differently by visual memory, where eyes can scan the stimuli many times, the auditory stimuli are not able to repeat the scanning; also, they can be received one at a time before they can be processed and understood (Jain, Ramta, Kumar, & Singh, 2015). Echoic sensory can store large amounts of auditory information; however, they can behold just for three or four seconds. When a stimulus is presented, it resonates, and it is been replayed many times for a brief period of time after the presentation of the stimulation (Radvansky, 2005).

In order to investigate the levels of processing effect, Fergus Craik and Robert Lockhart (1972) described memory recall of stimuli as a function of the depth of mental processing. This function occurs when the increase recall value of a stimulus is presented in the method with which it was imputed. For instance, it was suggested that auditory stimuli have the highest recall value when spoken or heard. Conversely, visual stimuli have the highest recall value when images are presented (Vaidya, Zhao, Desmond, & Gabrieli, 2002). In fact, there are different sensory modes that by their nature produce higher recall value in certain senses than others. In accordance with Intraub and Nicklos, (1985), visual input creates the strongest recall value compare to the other senses, allowing the most extensive spectrum levels of processing modifiers. This outcome can be explained by the fact that pictures and visual stimuli have an excellent potential for both physical and semantic richness. Equally, tactile memory representations create powerful recollections. The auditory stimuli follow conventional levels of processing rules, while odour memory is the weaker capacity of recall memories (Vaidya et al., 2002). Experiments focused on object detection suggested that even if briefly, the presentation of pictures are rapidly understood. This is because semantic questions that ask to identify information about images or a situation may evoke processing that is redundant to automatically initiated processes. The physical questions, in fact, might direct attention to additional characteristics of the scene that the individual would not typically focus on. The physical questions trigger the memory for visual rich stimuli, resulting in a more detailed and distinctive code (Intraub & Nicklos, 1985).

A few differences regarding the sensory memories can also be found in studies investigating the reaction time, specifically between auditory and visual reaction time. Conversely to the previous conclusions, studies reported that auditory reaction time is faster than visible reaction time (Pain & Hibbs, 2007; Vaidya et al., 2002). Specifically, visual stimuli take approximately 180-200ms to be detected, while auditory stimuli take around 140-160ms (Thompson et al., 1992). In confirmation of this, another study estimated that auditory stimuli take only 8-10 to reach the brain, whereas visual stimuli take 20-40ms (Kemp, 1973). Therefore, according to this study, the auditory stimuli reaches the cortex faster than the visual stimuli. However, Yagi et al. (1999) obtained different results from their research. In fact, they found that visual stimuli are faster than auditory stimuli.

Copyright Statement