DETECTING DECEIT - GUESSING OR ASSESSING? STUDY ON THE APPLICABILITY OF VERACITY ASSESSMENT METHODS IN HUMAN INTELLIGENCE

Thesis

Captain
Marko Uotinen

General Staff Officer Course 57
Army

July 2015
Intelligence from a human source, that is falsely thought to be true, is potentially more harmful than a total lack of it. The veracity assessment of the gathered intelligence is one of the most important phases of the intelligence process. Lie detection and veracity assessment methods have been studied widely but a comprehensive analysis of these methods’ applicability is lacking.

There are some problems related to the efficacy of lie detection and veracity assessment. According to a conventional belief an almighty lie detection method, that is almost 100% accurate and suitable for any social encounter, exists. However, scientific studies have shown that this is not the case, and popular approaches are often over simplified. The main research question of this study was: What is the applicability of veracity assessment methods, which are reliable and are based on scientific proof, in terms of the following criteria?

- **Accuracy**, i.e. probability of detecting deception successfully
- **Ease of Use**, i.e. easiness to apply the method correctly
- **Time Required** to apply the method reliably
- **No Need for Special Equipment**
- **Unobtrusiveness** of the method

In order to get an answer to the main research question, the following supporting research questions were answered first: What kinds of interviewing and interrogation techniques exist and how could they be used in the intelligence interview context, what kinds of lie detection and veracity assessment methods exist that are reliable and are based on scientific proof and what kind of uncertainty and other limitations are included in these methods?
Two major databases, Google Scholar and Science Direct, were used to search and collect existing topic related studies and other papers. After the search phase, the understanding of the existing lie detection and veracity assessment methods was established through a meta-analysis. Multi Criteria Analysis utilizing Analytic Hierarchy Process was conducted to compare scientifically valid lie detection and veracity assessment methods in terms of the assessment criteria. In addition, a field study was arranged to get a firsthand experience of the applicability of different lie detection and veracity assessment methods.

The Studied Features of Discourse and the Studied Features of Nonverbal Communication gained the highest ranking in overall applicability. They were assessed to be the easiest and fastest to apply, and to have required temporal and contextual sensitivity. The Plausibility and Inner Logic of the Statement, the Method for Assessing the Credibility of Evidence and the Criteria Based Content Analysis were also found to be useful, but with some limitations. The Discourse Analysis and the Polygraph were assessed to be the least applicable. Results from the field study support these findings. However, it was also discovered that the most applicable methods are not entirely trouble-free either.

In addition, this study highlighted that three channels of information, Content, Discourse and Nonverbal Communication, can be subjected to veracity assessment methods that are scientifically defensible. There is at least one reliable and applicable veracity assessment method for each of the three channels. All of the methods require disciplined application and a scientific working approach. There are no quick gains if high accuracy and reliability is desired.

Since most of the current lie detection studies are concentrated around a scenario, where roughly half of the assessed people are totally truthful and the other half are liars who present a well prepared cover story, it is proposed that in future studies lie detection and veracity assessment methods are tested against partially truthful human sources. This kind of test setup would highlight new challenges and opportunities for the use of existing and widely studied lie detection methods, as well as for the modern ones that are still under development.

**KEY WORDS**

Lie detection, veracity assessment, applicability, intelligence interview, interrogation, human intelligence, HUMINT, human source, deception.
DETECTING DECEIT - GUESSING OR ASSESSING? STUDY ON THE APPLICABILITY OF VERACITY ASSESSMENT METHODS IN HUMAN INTELLIGENCE

TABLE OF CONTENTS

1 INTRODUCTION .................................................................................................................. 1
  1.1 THE IMPORTANCE OF CORRECTLY ASSESSING THE VERACITY OF A HUMAN SOURCE ........ 1
  1.2 RESEARCH PROBLEM ................................................................................................. 3
  1.3 SCOPE, CONTEXT AND EXCLUSIONS ..................................................................... 4
  1.4 DUAL PROCESS THEORY AND INTELLIGENCE INTERVIEW ..................................... 6
  1.5 RELEVANT TERMINOLOGY ....................................................................................... 8
2 RESEARCH METHODS ....................................................................................................... 11
  2.1 META-ANALYSIS ....................................................................................................... 11
  2.2 MULTI CRITERIA ANALYSIS ................................................................................... 12
  2.3 FIELD STUDY ........................................................................................................... 14
3 INTERVIEWING AND INTERROGATION TECHNIQUES .................................................... 16
  3.1 QUESTIONING TECHNIQUES DO MATTER ............................................................... 16
  3.2 DIRECT QUESTIONS AND REPEATED QUESTIONING ............................................ 17
  3.3 ACCUSATORIAL QUESTIONING (REID TECHNIQUE) .............................................. 19
  3.4 KINESIC INTERVIEW AND INTERROGATION ......................................................... 22
  3.5 ELICITATION ............................................................................................................ 23
  3.6 INFORMATION-GATHERING QUESTIONING .......................................................... 24
  3.7 COGNITIVE INTERVIEW .......................................................................................... 25
  3.8 STRATEGIC USE OF EVIDENCE .......................................................................... 27
  3.9 POLYGRAPH, I.E. CONTROL QUESTION TEST AND CONCEALED INFORMATION TEST .. 28
  3.10 CONCLUSION ......................................................................................................... 31
4 VERACITY ASSESSMENT METHODS ............................................................................... 35
  4.1 MULTITUDE OF APPROACHES TO THE TRUTH ..................................................... 35
  4.2 ALL-SOURCE INTELLIGENCE GATHERING ............................................................ 40
  4.3 PLAUSIBILITY AND INNER LOGIC OF THE STATEMENT ...................................... 41
  4.4 CONTENT ................................................................................................................ 44
  4.5 DISCOURSE .............................................................................................................. 49
  4.6 NONVERBAL COMMUNICATION .............................................................................. 54
  4.7 CREDIBILITY AS A MEASURE OF VERACITY ....................................................... 59
  4.8 ADVANCED TECHNICAL AIDS ............................................................................. 64
  4.9 CONCLUSION .......................................................................................................... 69
5 APPLICABILITY OF THE VERACITY ASSESSMENT METHODS ...................................... 75
  5.1 OPERATIONAL ENVIRONMENT AND IMPLICATIONS TO APPLICABILITY CRITERIA .................................................. 75
  5.2 APPLICABILITY ANALYSIS ..................................................................................... 77
  5.2.1 ANALYSIS BASED ON EQUALLY WEIGHTED CRITERIA .................................. 77
  5.2.2 SENSITIVITY ANALYSIS, SCENARIO ONE ....................................................... 79
  5.2.3 SENSITIVITY ANALYSIS, SCENARIO TWO ..................................................... 80
  5.3 RESULTS FROM THE APPLICABILITY ANALYSIS .............................................. 81
6 FIELD STUDY .................................................................................................................. 85
  6.1 TEST SETUP .............................................................................................................. 85
REFERENCES

APPENDIXES
DETECTING DECEIT - GUESSING OR ASSESSING? STUDY ON THE APPLICABILITY OF VERACITY ASSESSMENT METHODS IN HUMAN INTELLIGENCE

"The use of the CIA’s enhanced interrogation techniques was not an effective means of obtaining accurate information or gaining detainee cooperation. … While being subjected to CIA’s enhanced interrogation techniques and afterwards, multiple CIA detainees fabricated information, resulting in faulty intelligence.”

Committee Study of CIA’s Detention and Interrogation Program, Findings and Conclusions. Senate Select Committee on Intelligence 2014, p. 2. (Author’s emphasis bolded.)

1 INTRODUCTION

1.1 The Importance of Correctly Assessing the Veracity of a Human Source

A professionally operating intelligence community is expected to produce reliable intelligence that is based on true facts. In many cases the gathered background information and other supporting data may be incomplete and the actual intelligence is based on deduction and assessments. Intelligence that is based on false facts or biased observations is misleading and can be more harmful than a total lack of it. Today, in the 21st century, plenty of different collection methods and intelligence disciplines are available. Some of them are more prone to the source’s deliberate manipulation than others and thus offer better chances for deception. The recognition of erroneous facts, distorted observations and biased assessments are critical steps towards more reliable and reality based intelligence. Still it must be recognized that 100% accuracy may never be achieved.

When defining the value of intelligence from a human source, in this case accuracy and usability, it is essential to assess the truthfulness of the source itself and the veracity of the infor-

1 Schum and Morris 2007, 249. ‘Fabricated accounts from a single human source more or less led to the U.S. invasion of Iraq in 2003. Afterwards it was discovered that “Curveball’s” testimony about Saddam Hussein’s WMD stockpiles was false.’

2 Field Manual 2-0, Intelligence 2010 (FM 2-0 2010), ii and iii. US Army Field Manual 2-0, Intelligence lists eight different intelligence disciplines.
information obtained from him or her. Collected intelligence can be incomplete or misleading for various reasons that are not only related to the human source’s deception. A human source might possess only a limited amount of information, might recall it incompletely or the scheme could be somehow distorted. These limitations are naturally occurring due to the nature of memory, recollection processes, high level of stress or overload of cognitive processes. The interviewer is also a possible source of error. Correct observations can lead to incorrect interpretations if the interviewer’s training is insufficient and if the information assessment methods are based on beliefs rather than knowledge. The interviewer’s pre assumptions, coercive methods and manipulation can lead to either totally false statements or true statements of which the veracity cannot be correctly assessed.

Sometimes interviewing and interrogation techniques, lie detection techniques and veracity assessment methods are confused to be synonyms or are thought to serve the same purpose. That is not necessary the case, though closer examination has revealed that in some cases they are overlapping. There have also been multiple efforts to create a Swiss Knife type of technique that combines all of the above mentioned areas into one. In a way these efforts reflect the need to diminish the ever rising working load that is inherent to human intelligence collection and related veracity assessment. Despite the efforts, such an overly efficient and accurate method has not yet been developed.

A lot of scientific studies have been conducted to address the challenges that are related to different interviewing and interrogation techniques and veracity assessment methods. The truthfulness of the human source or veracity of the information can be assessed with numerous methods. Based on their validity and usability in different information collection situations they are not seen as equal. Some of them are more valid and reliable than others and some of them are more prone to errors which derive from the interviewer’s lack of skills. The level of required expertise and other supporting technical equipment also varies depending on the method and situation.

Although accuracy and reliability of different veracity assessment methods have been studied broadly, a thorough analysis of these methods’ applicability is lacking. The aim of this study is to examine different methods for assessing the veracity of the information obtained from a

---

5 Neuman and Salinas-Serrano 2006, 200.
6 Hart, Fillmore and Griffith 2009, 135.
human source and to analyze these methods in terms of applicability in the context of the intelligence interview.

1.2 Research Problem

A few problems are related to the discussion of the different veracity assessment methods related to human intelligence. There is a conventional belief that an almighty, almost 100% accurate method, or methods, to detect lies exists. It is also believed that these methods can be applied rapidly during any social encounter. A layperson often adopts those methods that they feel are easiest to adopt and most convenient to reach. Nowadays very popular television series such as *Lie to Me* and *The Mentalist* give an easy access to the secrets of lie detection. Both of these television series are in many respects professionally made and are mostly based on actual behavioral science although methods are sometimes presented as quick wins with inadequate background information\(^7\).

From the scientific point of view popular approaches are often over simplified. Cutting the edges straight make methods easier to apply but in the end oversimplification can start to work against the original goal – getting information that is based on true facts.

The main research question of this study is:

- What is the applicability of veracity assessment methods, which are reliable and are based on scientific proof, in terms of the following criteria?
  
  o *Accuracy*, i.e. probability of detecting deception successfully
  o *Ease of Use*, i.e. easiness to apply the method correctly
  o *Time Required* to apply the method reliably
  o *No Need for Special Equipment*
  o *Unobtrusiveness* of the method

In order to get an answer for the main research question, the following supporting research questions are first answered:

- What kinds of interviewing and interrogation techniques exist and how could they be used in the intelligence interview context?

\(^7\) Heckman and Happel 2006, 64 and 75.

\(^8\) They are also subjected to critique because character’s success in different cases is simplified so that it makes good entertainment. See also: http://www.slate.com/blogs/quora/2012/11/01/are_the
- What kinds of lie detection and veracity assessment methods exist that are reliable and are based on scientific proof?
- What kind of uncertainty and other limitations are included in these methods?

Regarding the applicability of the veracity assessment methods the first hypothesis is that no single method is accurate and reliable by itself and higher accuracy can be achieved by using multiple methods side by side. The second hypothesis is that overall accuracy could be improved with the latest technical aids but their implementation is possibly obtrusive and requires special equipment with operating personnel.

1.3 Scope, Context and Exclusions

In this study the information collection is examined within the context of the intelligence interview. Intelligence interview as a term was chosen because emphasis is put on the non-coercive methods. Although officially interrogation in both law enforcement and intelligence refers to something that essentially is a dialog between two people, the term interrogation carries a slightly coercive connotation. Events in the history, questionable accusatory or guilt-presumptive interrogation methods in law enforcement and enhanced interrogation methods used in Iraq and Afghanistan are perhaps the main reasons for that⁹.

The research problem is approached from the social interaction’s point of view. The intelligence interview takes place in a face to face situation where direct social interaction is the primary means of communication. The interviewer can be supported or accompanied by other people but in the context of this study they do not interfere with the interview. This means that the successful interviewer is forced to multitask constantly in order to lead the conversation, listen to the interviewee, monitor the interviewee’s reactions, control his or her own behavior and plan ahead the interview based on the intelligence requirements and interviewee’s feedback.

⁹ Meissner et al. 2010, 9.
Different detailed scenarios, where intelligence interviews could take place, are excluded from this study. Emphasis is on the methods, their logic of operation and especially on the analysis of their applicability. The results and findings of this study may be applied to different kind of situations but that consideration is left for the reader or future studies.

Situations, where interviewees are antagonistic, mute or otherwise uncommunicative, are excluded. They form a different kind of challenge for the intelligence interview and are not in the focus of this study. Persuasion and other methods of influence that could be used to enhance communication between the interviewer and interviewee are also excluded.

Since this study is about detecting deception it is fair to assume that interviewees are not totally honest in their statements. The motives behind lying are not discussed. It is recognized that there are multiple reasons for interviewees not to tell the truth or make his or her version of the truth look as nice as possible\textsuperscript{10}. However, it is worth mentioning that some motives behind lying include higher stakes and stronger emotions than others and that way may have a more noticeable effect on the interviewee’s communication\textsuperscript{11}. It must also be emphasized that some people are comfortable with lying and show very little traditional signs of deceit\textsuperscript{12}.

\textsuperscript{10} Fein 2009, 51.
\textsuperscript{11} Frank and Svetieva 2012, 131.
1.4 Dual Process Theory and Intelligence Interview

During social interaction the detection of deception is based on perceptions that are mainly audial and visual. The person who is making assessments on the other one’s veracity receives audial information that the speech transmits. Not only the content, the words and sentences, but also the way the content is delivered conveys information. Expressions and gestures form the visual part of the information that is connected with the verbal message. Audial and visual information can be perceived to be in concert or in contradiction with each other.

Several studies have indicated that a human brain processes information on two levels that are more or less separated. A rather neutral scientific naming convention for intuition and reasoning has been adopted to make a distinction between these two mental operations. System 1, which refers to intuition, starts to process perceptions automatically with little effort and deliberate control. System 1 is slow-learning, associative, fast to apply and operates in parallel with other mental processes. System 2, which refers to reasoning, is comparatively slow, requires mental effort to apply and operates under the control of one’s mind. In addition to be controlled by rational thinking, System 2 is flexible, analytical, rule-governed and has also some control over System 1.13

It is easy to see how System 1 could be in all perspectives favorable compared to System 2. System 1 is unconsciously applied, fast, straightforward and meets the needs of anyone who seeks uncomplicated and rapid assessments of perceptions. Independence of working memory, independence of general intelligence, rapidness, low effort and automatic operation are the kind of traits that would be very usable in an intelligence interview context. On the other hand analytical, reflective, logical and rule based approaches are also welcome, but they come with the expense of low capacity, high effort and sequential operations which are slow to apply.

Systems 1 and 2 do not operate separately and out of control. They can be monitored and controlled by System 2 and other higher mental operations. Monitoring of System 1, System 2 and overt behavior requires mental effort and is exposed to external interferences like dual-tasking. People are not used to thinking hard and the output of System 1 is often satisfactory. In normal situations the monitoring of System 2 has been found to be lax. This highlights the

chance for erroneous judgments that are solely based on first impressions or intuition. However, intuitive skills can be improved with practice.\textsuperscript{14}

In spite of the many preferred qualities of intuition there are also drawbacks. As in perceptions, in intuition there is little room for uncertainty. Intuition processes information in relation to context and points of reference. This enables faster and more effortless assessments on the meaning of information. Compared to reasoning, that is better in the handling of absolute values and in analytical decision making, the intuition relies on the highly accessible features that are connected to ongoing or previous similar events. Unfortunately the most accessible features are not always the most relevant ones.\textsuperscript{15}

As it was mentioned before, System 2 has got control over both itself and System 1. However, System 2 can only apply corrective measures to the initial judgment by System 1 if biases or other flaws are detected. The effectiveness of System 2’s corrective measures is dependent on the accessibility of corrective thoughts, metacognitive knowledge and relevant statistical rules. In addition, System 2 can be impaired by time pressure, competing cognitive tasks, lowered alertness and being in a good mood.\textsuperscript{16}

This notion highlights the possibility to enhance System 2’s performance through training. Although System 1 cannot be taught or trained intentionally the same way as System 2, because it is slow-learning and out of reach of the deeper deliberate control, System 1’s performance can be improved with correct heuristics. Master chess players or experienced nurses, might not have developed their intuitive skills because they are natural talents in their respective fields of expertise, but instead they have established accurate heuristics through extensive exposure to various situations.\textsuperscript{17}

Overcorrection, high accessibility bias, incorrect heuristics or rules of reasoning and other heuristics related errors might still lead to incorrect judgments despite the best efforts to correct initial impressions or to control final reasoning. Practicing, access to relevant rules and statistical training potentially improve the performance of both systems.\textsuperscript{18}

\textsuperscript{14} Kahneman 2003, 699 and 700.
\textsuperscript{15} Kahneman 2003, 700, 701 and 703.
\textsuperscript{16} Kahneman 2003, 710 and 711.
\textsuperscript{17} Kahneman 2003, 699 and 700. Evans 2008, 267.
\textsuperscript{18} Kahneman 2003, 700, 707, 711 and 716. Evans 2008, 263, 267 and 268.
The evident easiness of System 1’s operation is appealing. In many cases System 1 produces assessments that are accurate enough or at least satisfy the observer’s imminent needs. System 1’s performance is enhanced through training or experience which improves heuristics, learned models or other accuracy related factors. As one anchor for System 1 is the baseline, it is very important that the baseline is constructed correctly. Only if perceptions are based on true facts, are interpreted correctly and are put to relevant context, it is possible to form a reliable basis for System 1’s operation. In a lie detection context, System 1’s fastness, automaticity and effortlessness are beneficial traits. Methods which would potentially utilize System 1 in lie detection are undoubtedly preferred. The threat in this kind of approach is that easily accessible assessments are not always relevant or accurate.

System 2 can be seen as a more scientific and comprehensive approach to lie detection. It does not only control and guide System 1, but it also performs higher mental processes that are necessary for in depth analysis. Lie detection methods that rely on System 2’s operation are less intuitive and require deliberate mental effort to apply. As processed information is subjected to more detailed analysis it is reasonable to argue that System 2 potentially produces information that is less prone to biases and the observer’s personal expectations. However, improved accuracy and diagnostic value comes at the expense of time and effort.

In human information processing there is a place for both of these systems. Intuition connects perceptions with reasoning. Despite the current commonly accepted model of System 1 and System 2 the phenomenon might not be that distinctive after all19. However, features that describe Systems 1 and 2 can be found to exist in cognitive processes. Intuition is appealing in everyday life and also in lie detection for clear reasons. Who would not like to be a master lie detector with a lightning fast sense for deception? Key to success could be a feasible combination of both intuition and reasoning to meet both practical and scientific requirements of respective situations.

1.5 Relevant Terminology

**Human intelligence** (HUMINT) is the collection of foreign information by a trained HUMINT collector. It uses human sources and a variety of collection methods, both passively and actively, to collect information including multimedia on threat characteristics20.

20 FM 2-0 2010, 1-22.
**Human source** is a person from whom information is received or elicited. Both in the intelligence and law enforcement communities human sources may vary by their status and willingness to cooperate. At best a human source may possess firsthand information of an event, people etc. through sight or hearing.\(^{21}\)

**Intelligence interview** is the event during which information from the human source is gathered through dialog. The setting of an intelligence interview may vary depending on whether it is related to military, law enforcement or other situations. Coercive methods are not used during the intelligence interview. The use of highly manipulative approaches, deliberately increased stress and mental pressure are ruled out\(^{22}\). In the intelligence interview the information gathering is mainly based on free expression and active listening.

**Veracity** is used to refer to the information that the human source has given and also to the source him/herself. In this study veracity is mostly used to refer to the source’s truthfulness but also the veracity of the statement is discussed. The **veracity of the source** is high when he/she is telling the truth with the best efforts. When the veracity or truthfulness of the source is assessed, deliberate efforts to be deceitful must be taken into consideration. If the source tells something that he/she truly believes to be true, despite the actual matter of the facts, he/she should not be considered deceitful although the given information is inaccurate\(^{23}\). If, despite the sincerest intentions, the source continuously gives inaccurate information his/her credibility should be questioned. The **veracity or accuracy of information**\(^{24}\) is high when the given information is based on the actual matter of the facts, instead of fabrication. However, inaccuracy of information can be caused by other factors, such as poor or distorted observations, which do not involve deliberate deception\(^{25}\).

**Lie, deception and deceit** are considered as synonyms in the context of this study. Lying is understood to be intentional efforts to communicate misleading messages in order to introduce into receiver’s mind something that the sender does not believe to be true\(^{26}\). Lying can involve staying silent as a sign of not knowing, leaving parts of the truth untold or making up false

---


\(^{23}\) Schum and Morris 2007, 248, 256 and 257.


\(^{25}\) Schum and Morris 2007, 12. Other terms to replace inaccurate could be false, erroneous, incorrect or fallacious.

facts. Also reporting someone else’s experiences as one’s own, embedding own past and unrelated experiences into a new context or altering the order of events can serve deceptive purposes.

27 Neuman and Salinas-Serrano 2006, 150.
2 RESEARCH METHODS

2.1 Meta-analysis

In the beginning of this study a meta-analysis was conducted to build understanding of the existing veracity assessment and lie detection methods. The focus of the literature search was in scientific papers that covered lie detecting, veracity assessment and related communication analysis methods. Google Scholar web search and Science Direct online database searches were used for the initial document collection. It was assessed that these two major date bases cover the most part of published scientific articles and studies. Immediately it became evident that lie detection and human source veracity assessment methods have been actively studied for past years. Science Direct alone returned almost 55 000 hits for lie detection, 1 500 hits for veracity assessment and 150 hits for veracity assessment and lie detection from 2012 to 2015. To narrow down the amount of search results the collection was focused on the latest studies, and the results were filtered according their relevance. As different possible methods for veracity assessment started to unfold additional search criteria such as lie detection methods, human source veracity assessment, spotting deceit, signs of deception, accuracy of lie detection, mechanical lie detection, computer assisted lie detection, nonverbal cues of lying and verbal cues of lying to mention few, was used. References from recent studies were also used as guidelines for the following literature search with narrower focus. In addition, different interrogation and interviewing techniques were searched.

The initial screening of the studies was done by reading the abstracts and after that relevant studies were stored for the following meta-analysis. A broad selection of relevant studies, scientific articles, working papers and other documents was collected. These documents were indexed for their relevancy and the scientific solidness. This was important as the amount of sources started to cumulate and focusing to the most reliable documents was needed.

After the search phase a meta-analysis of the gathered studies and other papers was conducted. The content of the document about a certain method was analyzed in order to find out the scientific principles and the logic of operation. Reliability and validity of the reports were also assessed. Finally the information from all of the papers concerning a certain method was compared with each other and conclusions from the synthesis were drawn. At this stage a general understanding of the methods that were available for further analysis was starting to develop.
From this point on the documentation was started by describing, analyzing and discussing the questioning techniques and veracity assessment methods in detail. Chapters 3 and 4 are the results of this phase. The aim of these chapters was to create an overall view of the questioning techniques and veracity assessment methods in order to answer supporting research questions. The main goal was to present the theoretical background in required detail so that the reader would have an easy access to all relevant information that was needed before and during the analysis of the methods’ applicability.

Picture 2. Description of the research process.

2.2 Multi Criteria Analysis

Multi Criteria Analysis (MCA), utilizing Analytic Hierarchy Process (AHP)\textsuperscript{28}, was used to assess different interviewing and interrogation techniques in Chapter 3 and veracity assessment methods in Chapters 4 and 5. Although the emphasis of the analysis in Chapter 3 was in the qualitative assessment MCA was used to form a second opinion on the different interviewing and interrogation methods.

MCA regarding the applicability of veracity assessment methods was conducted in two phases. The first round of applicability analysis was done in Chapter 4. It was based on the information provided by the scientific studies and other papers. During this preliminary analysis
the traits of the studied veracity assessment methods were described and documented in written form to support quantitative assessment that was done with a three-point scale.29

The selected applicability criteria were:

- **Accuracy**: Does the method work and is it reliable?
- **Ease of Use**: Is it likely that the method is misused or the outputs are misinterpreted?
- **Time Required**: How time consuming is the method?
- **No Need for Special Equipment**: Can the method be applied only by using special equipment?
- **Unobtrusiveness**: Can the method be applied covertly?

The following scale was used for the preliminary analysis:

- 1: Presents noticeable limitations
- 2: Presents some limitations
- 3: Presents no significant limitations

The result of the preliminary analysis was a list of veracity assessment methods that were carried on to the second round of analysis that was conducted in Chapter 5. At this point veracity assessment methods that had evident limitation in their scientific background, in terms of Accuracy, were excluded.

The actual applicability analysis was conducted using an AHP comparison table with relative weighting.30 Every veracity assessment method was rated in terms of each five applicability criteria listed above. The rating was a pairwise comparison answering the following question: Which one of these compared methods is preferred over the other one in terms of current criteria? In addition to choosing the preferred method, a relative weight was given with a nine-point scale, one meaning equal importance and nine meaning extreme importance in favor of the chosen one. After completing the comparison of the veracity assessed methods, in terms of all of the five criteria, a total sum of applicability was calculated.

---

29 The question regarding the preliminary analysis was: “Does the method present limitations in this field of criterion?”
The aim of the above described comparison was not to highlight the ultimately best alternative but to create a relative order of the alternatives in terms of the used criteria. In the scope of this study the result of the comparison would be the assessed applicability of the veracity assessment methods in relation to the others.

Differing from the guidelines of the AHP the result from the pairwise comparison of the main criteria was not included to the final applicability assessment\(^\text{31}\). The pairwise comparison of the main criteria was successful in highlighting the importance of criterion *Accuracy* and emphasized some differences between the other four criteria\(^\text{32}\). However, during the applicability analysis of the veracity assessment methods it was found out that the use of these findings did not add any significant information to the baseline analysis. Since *Accuracy* was analyzed to be five to ten times more significant than other criteria the result would heavily reflect the results of this single criterion.

In order to increase the analytical value of the applicability assessment two scenarios were described to justify the weighting of some applicability criteria. This kind of approach was assessed to be efficient in highlighting the differences or similarities of the studied methods. The interest of the applicability analysis was not only in the theoretical discussion and the requirements of the real life situations should be included to a relevant extent. In addition, the goal of the AHP was to highlight the differences and similarities, not to form a basis for a definitive decision making. Thus a deviation from the basic AHP procedures was considered to be justified.

### 2.3 Field Study

A field study was conducted to get firsthand experience of the applicability of the studied methods. An interview exercise was arranged during which about four hours of audio-video material was recorded in nine different interviews. Afterwards the selected assessment methods were used for the post analysis to assess the interviewees’ veracity. The goal of this phase was to gather additional information of the methods’ applicability to support the findings from the literature research.

Only few selected methods were tested because the latest technical equipment was not available to be included in this study. Also time limitations did not allow a more extensive analysis

---

\(^{31}\) Saaty 2008, 88.
at this point. The firsthand experience from the video recording analysis was included to the second round of applicability assessment.

By doing this the applicability analysis was not only based on literature research and theoretical discussions. This approach gave deeper insight to the actual applicability of the studied methods. Final conclusions were drawn after the second round of analysis. A detailed description of the test setup is included in the Chapter 6.
3 INTERVIEWING AND INTERROGATION TECHNIQUES

3.1 Questioning Techniques Do Matter

The intelligence interview or interrogation is essentially a controlled dialog where the interviewer asks questions and the interviewee answers them. Prerequisite for the information collection is that the interviewee is willing to cooperate and gives relevant answers. The answers can only be as relevant as the questions are. It is reasonable to assume that in some cases the interviewee does not anticipate what kinds of matters the interviewer is exactly interested in. It is relevant what questions are asked and how they are delivered.

Closed questions offer an opportunity to give very short answers and open-ended questions potentially promote longer answers. It has been noted that leading and suggestive questions can produce false or misleading recollection of information. From the information gathering point of view leading or suggestive questions are not preferred.

It becomes more difficult to assess the veracity of the given information if there is not much of it. Common to all veracity assessment methods is that they are based on observations. Observations become statistically more reliable when their number increases. Even the application of the more advanced technical methods requires two-way communication to take place. It has been suggested that such questioning technique that produce longer answers should be preferred. These techniques potentially produce more information and also more opportunities for verbal cues of deception to occur.

Interviewee’s memory plays an important role in this kind of information gathering setup. Human memory does not work like a camera that records and stores every observed event as it originally occurred. Observations are distorted, human brain stores them incompletely, memories decay over time and memorized events or details could get corrupted during the retrieval process. Interviewing technique that makes memory retrieval processes easier, and produces undistorted recollections of past events or other details, should thus be preferred.

33 Fein 2009, 86 and 88.
34 Vrij and Granhag 2012, 115.
36 Fein 2009, 82-84.
Among others, one way of resisting the interrogation or intelligence interview is *not remembering*\(^\text{37}\). As mentioned above, poor memory is not a definite sign of deceit. To make a difference between these two possibilities is not a simple task. An effective questioning technique should also take this factor into account. Questioning technique should challenge the interviewee in a constructive manner if interviewee’s statement indicates inconsistency, is not plausible or is in contradiction with the known facts. The emphasis is on the constructiveness to avoid possible suggestion, diminished willingness to talk and excessively increased stress that could handicap memory. To conclude, questioning techniques that are not leading or suggestive, produce longer and narrative answers, support accurate memory retrieval and challenge inconsistencies in a constructive manner should be preferred.

Some of the commonly known or otherwise promising questioning techniques are described in the following sub chapters. After that they are analyzed according to above mentioned qualities of effective information gathering questioning. It was assessed important to include Accusatorial Questioning technique and the polygraph test in the analysis because they are widely used in the North-American law enforcement and publicly known through the entertainment industry.

A comparison table and further analysis of the questioning techniques are presented in the conclusion of the Chapter 3.

### 3.2 Direct Questions and Repeated Questioning

The Human Intelligence Collector Operations manual suggests direct questions to be used as basic form of questioning. Direct questions are supposed to be simple, easy to understand, begin with interrogative and thus require a narrative answer. This way the HUMINT collector would fulfill his mission, “the rapid collection and dissemination of accurate information”, most efficiently. The manual instructs collector to anticipate the source’s probable answers and to avoid questions that the source would not answer or questions that would provoke negative feedback. Six types of direct questions should be used to make conversation fluent, to make sure relevant topics are discussed, to make the collection objective less obvious and to

\(^{37}\) Fein 2009, 76 and 82.
assess source’s truthfulness. Leading, negative, compound and vague questions should be avoided because they enable the source to answer very shortly or ambiguously.

Repeated and control questions should be used to verify source’s truthfulness if the collector has a reason to suspect that the source is being deceitful. Repeated questions should produce same answers consistently but without the exact wording and detail. That would indicate a planned lie. Correct answers to the control questions are known beforehand. Deceit is confirmed if the source answers them incorrectly. The manual also acknowledges other possible explanations for incorrect answers.

The manual stresses that the collector should never ignore the signs of deceit. To verify that the observations are correct, the collector is advised to collect more information and assess the source’s veracity. Listed additional task can be grouped into four categories: repeated and control questions, addressing the inconsistency directly and ask for explanation, seeking assistance, conducting further analysis and using the polygraph.

Based on the Human Intelligence Collector Operations manual’s descriptions, direct questioning technique seems to work towards the goal – to produce answers that provide information which veracity is tentatively assessed. However, the manual strongly suggests that there is only a limited amount of time and that the collector must not be sidetracked into unimportant discussions or debates. If everything is done by the book, the collector basically leads the conversation, asks a set of questions, listens to the source’s answers, makes additional questions, ends the questioning and leaves to write his report. Naturally this example is over simplified and trained interrogators would not do it in this manner. It must still be noted that too heavy goal orientation could lead to a hasty interview that produces reduced amount of useful information.

Repeated questioning, as a means of verifying the source’s veracity, possesses some potential pitfalls. Granhag and Strömwall (1999) conducted a series of studies where they tested how

---

38 Direct question types are: Initial, topical, non-pertinent, repeat, control and prepared questions.
40 FM 2-22.3 2006, 9-3 and 9-7. Other possible explanations are: The source has misunderstood the question, could make up information to please the interviewer and get promised reward, answers to the best of his knowledge (which is wrong or outdated) or is correct (interviewer has got wrong information).
41 FM 2-22.3 2006, 9-6 - 9-8. Cues of deception are mainly inconsistencies in interviewee’s statement, behaviour or nonverbal indicators.
43 FM 2-22.3 2006, 9-1.
repeated interrogations affected lie-catchers’ deception detection performance and judgment abilities. It was discovered that access to consecutive statements from the same source did not enhance lie-catchers’ performance in comparison to having access to only one statement. In addition, the access to multiple statements enhanced their confidence in the initial veracity assessment. The assumption that inconsistency would indicate deceit was also challenged. The research data showed that deceptive statements were perceived at least equally consistent as the truthful ones and that people tend to disagree whether studied consecutive statements were consistent or not. One explanation for these finding was that liars try to repeat their initial statement to make themselves appear consistent and truth tellers try to reconstruct their actual past experiences.  

As a result of another study, Granhag and Strömwall (2001) found out that a successful veracity assessment is more related to the observer’s deception detection performance than the amount of repeated interrogations. The study also indicated that the ones who conducted the questioning, and thus were in face to face contact with the interviewees, were more credulous than those who were just observing the interviews afterwards. This clearly indicates the possible negative results of truth bias that interviewers must be aware of.  

Asking direct questions is a simple and straightforward way of questioning. At its best direct questioning promote narrative answers with little leading and suggestion. It does not support effective memory retrieval per se but does not work against it either. Repeated and control questions can be used to challenge inconsistencies although it must be emphasized that inconsistencies do not always indicate deceit.

3.3 Accusatorial Questioning (Reid Technique)

In Western countries the accusatorial questioning approach is mainly used by the United States’ and Canada’s law enforcement officials. The goal of this method is to make a presumably guilty suspect to confess by confronting him with undeniable pieces of evidence. Guilt-presumption is based on a preliminary interview, during which the suspect is assessed to be guilty of committing the crime under investigation. The assessment of the suspect’s guiltiness

---

44 Granhag and Strömwall 1999, 163.
45 Granhag and Strömwall 2001, 85.
46 The other general method of questioning suspects is information gathering approach which will be discussed in further detail later in this report.
is based on the observation of the suspect’s nonverbal and verbal cues of deception.\textsuperscript{47} The validity and reliability of the cues provided by nonverbal communication is discussed later in this study.

Although accusatorial questioning technique is used to yield a confession from a crime suspect, an analogy to the context of this study can be found. The kind of offence a human source could be charged with in the intelligence interview context is not telling the truth or withholding relevant information. The confession in this case would be a full and detailed statement about the topic under scrutiny. The use of accusatory questioning technique is potentially tempting in an effort to confront a deceitful human source and therefore it is discussed in further detail to highlight its potential threats against getting reliable and useful information.

During the first phase a non-accusatorial interview is conducted. Based on the information obtained during this initial questioning, a decision is made whether to continue to the next phase or not. If the suspect exhibits traits that indicate him/her being deceitful, and therefore is reasoned to be guilty\textsuperscript{48}, the actual interrogation continues in the second phase. The assessment of the suspect being deceitful is based on a Behavioral Analysis Interview (BAI) during which a Behavioral Symptom Analysis (BSA) is applied\textsuperscript{49}. The analysis relies on nonverbal, linguistic and paralinguistic cues of deception\textsuperscript{50}. BAI was developed by Fred E. Inbau and his associates. BAI utilizes a tailored set of questions to specifically elicit nonverbal responses that are thought to be different in truthful and deceptive interviewees. It is suggested that liars do not feel as comfortable as the truth tellers do and they tend to cross their legs, change frequently their position in a chair, exhibit grooming gestures and avoid eye contact as a result of an investigative interview.\textsuperscript{51}

The accusatory interrogation can be summarized to consist of three basic components: custody and isolation, confrontation and minimization. During custody the suspect’s anxiety and insecurity is intended to rise. During confrontation the pressure is increased and he/she is told to be guilty and presented with evidence, sometimes fabricated, to support this accusation. His/her denials of involvement in the crime are interrupted. Minimization is then used to lower the tension and to offer the suspect a face-saving explanation for the justification of the

\textsuperscript{47} Meissner et al. 2010, 6, 10 and 11.
\textsuperscript{48} To come to conclusion that deceit equals guilty it is assumed to be necessary that the suspect is asked if he/she did commit the crime or not.
\textsuperscript{49} Kageleiry 2007, 28.
\textsuperscript{50} Meissner et al. 2010, 10 and 11. Neuman and Salinas-Serrano 2006, 183 and 184.
\textsuperscript{51} Vrij and Granhag 2012, 112.
crime. At the same time it is implied that there would be less dire consequences if he/she con-
fesses.\textsuperscript{52}

Reid, Royal and Schutt, Aubry and Walters, to mention a few, have all developed their own
versions of accusatory interrogation method. The Reid Technique is probably the most de-
tailed and it advices the interrogator to carefully go through nine steps of interrogation to ob-
tain a confession\textsuperscript{53}. Federal Bureau of Investigation has also adopted a slightly modified ver-
sion of the Reid Technique called Direct Accusation Approach\textsuperscript{54}.

During step 1 the interrogator tells that the suspect is considered to be the one who has com-
mittted the offence. This confrontation is done in a direct and positive manner. Steps 2 to 4
form more or less a monolog where the interrogator presents reasons, excuses and facts which
all support the accusation, cuts off all denials, listens to the suspect’s objections but ignores
them at that time. Now the guilty person is supposed to withdraw from the conversation. Dur-
ing steps 5 and 6 the suspect’s attention is regained so that the interrogator can move on to the
next step. Step 7 is the decisive one during which the alternative question is presented. The al-
ternative question presents a choice between two explanations for his/her offence. The other
one is more attractive and morally acceptable than the other. By accepting either of them the
suspect admits committing the offence. For example the question could be: “Did you plan it in
advance or did it just happen at that heat of the moment?” Last two steps, 8 and 9, include the
elicitation of further details about the offence and finally a written confession. To support the
interrogator’s work, examples of both guilty and innocent suspect’s reactions and overall be-
havior in relation to different steps of the interrogation are given. In general, truthful suspect
are listed to be spontaneous, sincere, helpful, concerned and cooperative, and deceptive ones
the opposite.\textsuperscript{55}

This evidently manipulative\textsuperscript{56} and also slightly coercive nature of the Reid Technique has
raised critique and opposition. It is noteworthy that meta-analysis of existing studies show that
accusatory style interrogation significantly increases the likelihood of false confessions from

\footnotesize{\textsuperscript{52} Meissner et al. 2010, 11.  
\textsuperscript{53} Hirsch 2014, 804 and 805. Reid passed away in 1982 and his collaborator Inbau passed away as
well in 1998. Joseph Buckley and Brian Jayne have continued their work. The latest edition of the Reid
Manual (Fred E. Inbau et al., Criminal Interrogations and Confessions viii, 5th ed. 2011) was published
in 2011. \textsuperscript{54} Neuman and Salinas-Serrano 2006, 187-194 and 204. \textsuperscript{55} Neuman and Salinas-Serrano 2006, 185 and 187-190. To overcome critique against this self-
incriminating alternative question it is suggested that the suspect could always say “neither of them is
true”. \textsuperscript{56} Neuman and Salinas-Serrano 2006, 146.}
innocent suspects\textsuperscript{57}. Although the fifth edition of the Reid Manual has addressed some of the critique concerning the scientific background of the BSA and the relative high probability of false confession, the basic method and its application has stayed the unchanged\textsuperscript{58}.

The use of accusatorial questioning technique to overcome resistance, or to challenge the presentation of false information, is also questionable due to the nature of the context and the main objective – the information collection. What is the value of the interviewee’s confession of being untruthful or having delivered incomplete information? A confession of this kind would not add genuinely relevant information to the overall picture because there is always a slight doubt about the interviewee’s truthfulness. It should also be emphasized that after the confession the interviewee would not transform into being a 100 % reliable and trustworthy source. Reliable and valid veracity assessment methods would still be of the utmost importance during the following information collection sessions.

To conclude, accusatorial questioning approach has very little, if any, qualities that promote information collection. If this technique was to be used in the intelligence interview context the following would most likely occur. The interviewee’s efforts to present his views, opinions and knowledge would be suppressed if they did not support the interviewer’s narrative and the interviewee’s role in it. The interviewer would actively plant his/her idea of truth to the interviewee during the course of the interview. Since the focus of the interview would not be in the information collection, very little attention is paid to the interviewee’s abilities to remember events or other related details correctly. Not only inconsistencies but also all the other pieces of information that did not fit in the interviewer’s pre-assumptions would be forcefully challenged. Based on this assessment, accusatorial questioning does not promote the collection of relevant and reliable intelligence.

3.4 Kinesic Interview and Interrogation

The Kinesic Interview and Interrogation technique is developed by Stan B. Walters and it bears remarkable resemblance to the Reid Technique. Similarly this technique consists of two phases. The first phase is a non-accusative preliminary interview during which information is collected and suspect’s level of deception is assessed. The assessment is based on a process


\textsuperscript{58} Hirsch 2014, 805, 811, 818 and 825.
called Practical Kinesic Analysis Phase (PKAP). PKAP is similar to BSA that is in the core of the Reid Technique’s veracity assessment process.\textsuperscript{59}

The second phase is the actual interrogation. It starts with an “interrogation attack” where the suspect is confronted with the accusation and possibly with supporting evidence. The suspect is then expected to react with specific “ego defense mechanisms” to defend against the accusation. These mechanisms include for example denial, rationalization and minimization. Then the interrogator should ”disarm” these defense mechanisms by confronting the suspect with further questions and evidence. When the time is right, and the suspect starts to yield, the interrogator makes the final push for confession.\textsuperscript{60}

In essence, Walters has developed another accusatorial interrogation technique based on strong guilt-presumption. Like in the Reid Technique, the purpose of the preliminary interview is not to collect information but to evaluate the suspect’s nonverbal and verbal cues of deception. In the PKAP and the BSA the emphasis is on the evaluation of how the suspect answers, not what the answer is.\textsuperscript{61}

Despite of its name, this technique has little to do with interviewing or information collection. In many senses it has the same pitfalls as the Reid Technique, however, less suggestion is present. To conclude, the Kinesic Interview and Interrogation technique has very little qualities that promote information collection.

3.5 Elicitation

The elicitation of information refers to the collection efforts that are somewhat covert and do not immediately give out the topic of the actual interest. The elicitation takes place during a conversation or an interview. Through elicitation it is possible to engage a human source in a manner that he/she reveals valuable information without being aware of it. By definition it differs from the strategic debriefing and interrogation as the latter mentioned techniques do not try to hide the actual interest of the interviewer. It is also implied that during the elicitation the source is engaged in the conversation more or less voluntarily.\textsuperscript{62}

\textsuperscript{59} Neuman and Salinas-Serrano 2006, 183-185.
\textsuperscript{60} Neuman and Salinas-Serrano 2006, 193-194.
\textsuperscript{61} Neuman and Salinas-Serrano 2006, 184 and 186.
\textsuperscript{62} Fein 2006, 2. FM 2-22.3 2006, 9-5.
The Human Intelligence Collector Operations manual emphasizes that “The key to elicitation is the establishment of a rapport between the elicitor and the source, normally based on shared interests.” A good connection between the interviewer and the interviewee should promote mutual trust and uninhibited conversation. The establishment of a free conversational atmosphere might take time and it should not be rushed. Once achieved, a free flowing conversation is kept on going and the topics are gradually moved on towards the actual area of interest. It is advised to use mild flattery or provocation to make the conversation livelier, if needed. It is also important to listen and let the interviewee speak. Non-pertinent questions should be used to conceal the collection objectives.

As the need for a good interpersonal relationship was mentioned before, it is worth noting that master intelligence interviewers are told to be persuasive, flexible and use dynamic approaches. Persuasive approaches may improve interpersonal communication and thus provide more effective elicitation of information.

A lot of room for the execution of elicitation is left as the general guidelines are quite wide. The elicitation of information can be used in different kinds of contexts and real life situations. The elicitation produces a large amount of speech and information. Elicitation is a dialog which potentially covers a wide range of topics. It is assessed that the interviewee is not affected by leading questions or other suggestion, but it cannot be ruled out that the interviewee could pace the interviewer’s opinions. There are only little intentional distractions to interviewee’s memory retrieval processes. Inconsistencies are advised to be challenged to promote the conversation and the exchange of opinions. On the whole, the elicitation promotes information collection by producing open statements that are not forced or otherwise deliberately manipulated.

3.6 Information-gathering Questioning

Information-gathering approach to questioning is used by law enforcement officials mainly in Western Europe, United Kingdom, New Zealand and Australia. In contrast to the accusatorial questioning its aim is to seek information instead of securing a confession. It is based on rapport and active listening. Similar practices can be found in different interviewing protocols.

---

63 FM 2-22.3 2006, 8-22, 8-23 and 9-2.
64 Fein 2009, 27.
65 Meissner et al. 2010, 6.
that foster the use of good questioning techniques like active listening, open ended questions and appropriate nonverbal behavior by the questioner.\textsuperscript{66}

This questioning technique is slightly reminiscent of the elicitation. The main difference is that the purpose of the questioning is openly brought to the interviewee’s attention. The information gathering is mainly based on the interviewee’s uninterrupted answers to open-ended questions. The interviewer’s role is to guide the conversation to cover all of the relevant topics in desired detail. Only after the suspect has been given full opportunity to provide his/her information further questioning is carried on. Inconsistencies and contradictions should be addressed after the suspect’s statement.\textsuperscript{67} As the goal of this interviewing method is \textit{fact finding}, it is emphasized that open-ended questions are used, false evidence is not presented and the investigators are prohibited to deceive the suspects.\textsuperscript{68}

Information-gathering questioning technique promotes long narrative answers and intentionally avoids leading or suggestive questions. Although not specifically mentioned, it is assessed that this technique could support memory retrieval by producing a non-coercive conversation-al atmosphere. False information and inconsistencies are addressed if they occur.

\section*{3.7 Cognitive Interview}

The Cognitive Interview (CI) is a method which purpose is to enhance the interviewee’s ability to remember past events and other related details under investigation. The emphasis is to make memory retrieval easier and more accurate. The basic method of questioning is open ended questions and the overall goal is to collect as accurate and uncorrupted information as possible.\textsuperscript{69}

The CI can be divided into four separate techniques. The first technique is a mental reconstruction of the physical and personal context. During this contextual reinstatement, the interviewee is asked to form an image or impression of the environmental aspects, to describe sounds, smells or other physical aspects and to comment on the feelings and emotional reac-

\textsuperscript{66} Memon and Higham 1999, 185. Structured Interview bears a resemblance to Information-gathering interview.

\textsuperscript{67} Meissner et al, 2010, 11. This technique has been adopted and further developed in the United Kingdom. It is called PEACE-model which stands for Planning and Preparation; Engage and Explain; Obtain an Account; Closure; Evaluation.

\textsuperscript{68} Meissner et al. 2010, 11-12.

\textsuperscript{69} Memon, Meissner and Fraser 2010, 3.
tions that are related to the event. These cues will effectively help memory retrieval. The second technique is to encourage the interviewee to tell everything he/she remembers in full detail without preliminary screening. Every related detail and observation should not be considered indifferent or irrelevant and thus left untold. The third technique is to recall the event from another participant’s point of view. It is assumed that a change in the perspective forces a change in the memory retrieval and enhances recollection. The fourth technique is to recall the event in mixed or reverse order. This, like the change in the point of view, is assumed to change the retrieval process and produce more detailed information.

There are some potential drawbacks included in this technique. The change of the interviewee’s perspective to some others’ could lead to false or fabricated details. In addition, studies have indicated that a repeated imagining of a picture is likely to produce a false memory of visual perception. Reconstructed images may also start to interfere with memory retrieval and the interviewee could confuse these images with the actual events. These examples also highlight the potential danger of the interviewer’s suggestive communication.

An enhanced version of the CI (ECI) has been developed in 1992 by Fisher and Geiselman with the emphasis on the rapport building and effective communication. The interviewee is in the center of the information collection. The interviewer should not interrupt the free flow of information and instead should put all of his/her effort in active listening. First the communication is facilitated with open ended questions about neutral topics. After that the collection is started with context reinstatement and interviewee’s free narrative of events. The interviewer then reminds that it is important to give a full account of events in as much detail as possible. It is also emphasized that the interviewee should not make anything up or guess; instead he/she should simply tell that “I do not know”, if that is the case.

To sum up, four principles can be found in the questioning routine above: neutral interaction for the initial approach, inclusion of the contextual elements, reconstruction of the event in reverse or non-linear order and reiteration of the event from another point of view.

---

70 Memon and Higham 1999, 184. It should be noted that there is a strong resemblance between the contextual reinstatement and Guided Memory Interview (GMI). CI is more comprehensive and it covers GMI’s procedures.
71 Memon and Higham 1999, 178 and 179.
72 Memon and Higham 1999, 179, 181 and 182.
73 Memon, Meissner and Fraser 2010, 5.
The CI technique can be classified as an improved information collection technique. Although it can facilitate additional information to support veracity assessment, the interviewing method itself does not contain any veracity assessment processes\(^7^4\). From the information collection point of view, this technique promotes many good values such as open ended questions, minimal amount of leading and suggestion, long and narrative answers with related contextual information and emphasis on the accurate memory retrieval. There are no direct instructions on how the inconsistencies should be confronted, but the above mentioned specific CI technique number two, tell everything in full detail without preliminary screening, and number four, recall the event in mixed or reverse order, address that issue indirectly. However, it is emphasized that skillful application of the CI requires personal skills and systematic training\(^7^5\).

### 3.8 Strategic Use of Evidence

The Strategic Use of Evidence (SUE) is based on a carefully planned presentation of critical and possibly incriminating information or evidence. The presentation of the facts, that the interviewer knows and has verified to be true, is supposed to produce different verbal responses from truth tellers than liars. It is thought that the guilty suspect will use more avoidance strategies during the discussion compared to the innocent suspect who is thought to be more forthcoming. Avoidance strategies include for example leaving parts or details of the event untold. The truth tellers are generally more likely to tell the truth as it happened.\(^7^6\) Higher correlation between the given statement and the known facts would thus indicate truth and lower correlation would indicate deception.

On tactical level SUE consists of three elements: evidence tactics, question tactics and disclosure tactics. The evidence tactics is mainly used for the preliminary assessment of the evidence at disposal. The question tactics is used to systematically exhaust the suspect’s alternative explanations for the presented incriminating evidence. The disclosure tactics is used to increase the diagnostics value of the evidence by assessing the strength of the source and the degree of the precision to be able to present the evidence in a certain way. Research shows that the step-by-step presentation of stronger and more detailed evidence produces more and

\(^7^4\) The research conducted by Vrij et al. (2008) and Leal et al. (2010) show that increased cognitive load produces more indicators for detecting deception.

\(^7^5\) Memon, Meissner and Fraser 2010, 35 and 36.

\(^7^6\) Vrij and Granhag 2012, 114.
stronger cues to deception than presenting the most incriminating pieces of evidence straight away.\textsuperscript{77}

The effectiveness of SUE relies on two basic assumptions. First, a deceitful interviewee gives more incomplete answers and tries to avoid incriminating topics. Second, the guilty interviewee exhibits more cues\textsuperscript{78} to deception \textit{as the plot thickens} while stronger and more precise incriminating pieces of evidence are presented. If either of these two principals fails to work, the use of SUE would potentially produce incorrect assessment of the interviewee’s veracity.

It is essential to recognize that asking questions presents only half of the technique. The other half includes detecting verbal and nonverbal cues of deception, including the avoidance strategies. Open-ended questions must be used to produce long and comprehensive answers to enable the avoidance strategy assessment. To be able to make correct conclusions, the interviewee is supposed to know the \textit{truth} or \textit{correct answers} beforehand. Once the interviewer is convinced that the correct moment to confront the deceitful interviewee has come, the incriminating evidence is presented step-by-step and the interviewee becomes compelled to confess. In this process some analogy with the Reid Technique can be found. However, SUE is not driven by the guilt-presumption or deliberate manipulation, and mental pressuring is not used.

Since higher correlation between the given statement and the known facts would indicate truth and lower correlation deception, the prerogative for the veracity assessment is a successful collection of reliable background information. The assessment itself is based on logical deduction when the answers are compared with the known truth. It seems that SUE is more suitable for obtaining a confession than collecting information. However, the presence of avoidance strategies regarding the known facts could be a valuable tool in the veracity assessment. In addition, SUE could be effective in overcoming resistance by persuading the interviewee to admit his/her deception.

### 3.9 Polygraph, i.e. Control Question Test and Concealed Information Test

The purpose of the polygraph test is to assess the veracity of the suspect in relation to specific questions. The main goal of the test is to verify if the suspect is being truthful when he/she states to be innocent or claims to have a certain level of knowledge about the topic discussed.

\textsuperscript{77} Vrij and Granhag 2012, 114.
In the Control Question Test (CQT) the physiological responses caused by relevant, and possibly incriminating, questions are compared to the physiological responses caused by control questions. The truthfulness of the interviewee is deducted from the results of this comparison. In the Concealed Information Test (CIT) the physiological responses are thought to be consistently larger after questions regarding critical items than after neutral questions. The interviewee is reasoned to have guilty knowledge, or concealed information, as a result of the consistent variance in measured parameters produced by relevant questions. In both tests closed questions are used and the interviewee answers either yes or no.

The Preliminary Process Theory (PTT) is explained to be the reason behind the changes in physiological reactions during the instrumental assessment process. Stimuli, in this case the questions, initiate cognitive processes which produce variances in physiological responses. The greater the relative significance of the stimuli is, the larger the variance in the monitored parameters. The veracity assessment is based on monitoring and recording of the reactions, as described above.

The research of the polygraph’s theoretical background goes over to the late 19th and early 20th century when the existence of such phenomena as changes in respiratory activity, cardiovascular activity and galvanic skin response were discovered. These are the same parameters that are monitored today during the polygraph test. Nowadays they are called electrodermal skin conductance response (SCR), pause in respiration (RESP), peripheral vasoconstriction in the fingers (PVC) and changes in the heart rate (ECR1 and ECR2). To sum up, a deceitful suspect is expected to slightly sweat, have a pause in breathing, experience changes in blood pressure and heart rate while lying – all involuntarily.

The core of the polygraph testing relies on the assumption that lying causes emotional arousal and anxiousness. The symptoms of the emotional arousal are beyond the control of the suspect and are in this sense incriminating. The chain of deduction becomes very simple: emotional arousal means lying and lying means guilt.

78 The source does not name any specific cues, but it is reasonable to guess that they refer to nonverbal and verbal cues of deception.
79 Palmatier and Rovner 2014, 1, 2 and 6.
80 Vrij and Granhag 2012, 112.
81 Palmatier and Rovner 2014, 1.
82 Palmatier and Rovner 2014, 5 and 7.
83 Palmatier and Rovner 2014, 5.
The polygraph test has also raised a lot of critique. The CQT’s underlying logic is heavily criticized, because truthful suspects are tricked to lie to the control question and thus are expected to get stressed\textsuperscript{84}. There is no guarantee that the relevant questions, essentially accusations for committing a crime, are not stressful for truthful persons because there is always a chance for an error in this type of test\textsuperscript{85}. Stress may also be caused by other factors such as topic related memories\textsuperscript{86}.

The logic of the PPT is not rock solid either. The relative significance of the input stimuli affects the resultant variance of the monitored physiological responses. Since the relative significance plays such an essential role in this process, the effect of a respondent’s core values should be discussed.\textsuperscript{87} Nonetheless, it should not be taken for granted that lying makes everybody uncomfortable\textsuperscript{88}.

Arguments for the effectiveness of the polygraph testing usually consist of the statistical proof based on laboratory studies, field studies and real life experience. As high as 91-98\% accuracy in identifying a guilty suspect and 75-89\% accuracy in identifying an innocent suspect have been reported in laboratory and field studies. Additional studies show that with the correct questioning technique the accuracy can be improved.\textsuperscript{89} On the more skeptical side are studies indicating average accuracy rates of 67-75\%\textsuperscript{90}.

Arguments against the claimed effectiveness of the polygraph testing address the reliability of the test setup in the laboratory studies and the analysis of the results versus the ground truth in real life situations. The benefit of the laboratory studies, where mock crimes are conducted, is the knowledge of the ground truth. However, they lack all the high stakes of real life situations that concern both the guilty and innocent suspects. In real life cases the high correlation between deceptive polygraph test results and the related confession should not be surprising. In this kind of comparison there are no traces left either from the truly innocent suspects that feel compelled to confess or the truly guilty suspects that pass the test for a reason or another.\textsuperscript{91}

\textsuperscript{84} Palmatier and Rovner 2014, 2, 5 and 6. 
\textsuperscript{85} Iacono 2008, 25. 
\textsuperscript{86} Palmatier and Rovner 2014, 5. 
\textsuperscript{87} Palmatier and Rovner 2014, 1 and 5. 
\textsuperscript{88} Vrij and Granhag 2012, 110. Lauerma 2015. 
\textsuperscript{89} Palmatier and Rovner 2014, 4. Iacono 2008, 24. By Palmatier and Rovner the use of non-exclusionary questions is stated to improve the identification of truthful individuals significantly (45\% vs. 91\% accuracy) and the identification of deceptive individuals only marginally (80\% vs 85\% accuracy). Iacono strongly criticizes the average accuracy rates of 98\% reported by American Polygraph Association and argues against the research method as well as the conclusions. 
\textsuperscript{91} Iacono 2008, 25.
From the information collection point of view the Concealed Information Test is slightly more usable than the Control Question Test. Where the CQT is mainly intended to be used for obtaining a confession, during or after the polygraph test, the CIT is used to assess the overall knowledge of the suspect of a certain event or matter. The CQT could be used to assess the veracity of the interviewee by asking questions such as *have you lied to us* or *do you know more than you are telling*. This kind of questioning really does not elicit more information and the known shortcomings of the polygraph test’s reliability should be somehow handled.

The CIT has some potential of eliciting information along with the related veracity assessment. The most significant issue with the CIT is the use of closed questions. The interviewee should know specific details of the discussed matter in order to make correct reaction based assessments. By using the CIT, it would be possible to develop understanding about the topics related to which the interviewee has not been totally truthful. Still this technique is heavily dependent on the soundness of the Preliminary Process Theory and the correct interpretation of the interviewee’s observable stress reactions.

Neither of the questioning methods heavily promotes the desired values of the information gathering. Closed questions produce only *yes* or *no* for an answer. During the CQT the interviewee is fooled to lie to the control questions. This in a way works against the desired total honesty of the interviewee. In these techniques nothing really enhances memory retrieval. Also answers are only limited to the topics that the interviewer has decided to cover. Possible inconsistencies are challenged only afterwards if the deception is detected. The CQT and CIT are essential parts of the technically aided veracity assessment method, the polygraph test, which is further assessed in Chapter 4.

### 3.10 Conclusion

A pairwise comparison was conducted to create a ranking list of the covered interview and interrogation techniques. The goal of the comparison was to highlight the pros and cons of each technique over the other. It is not suggested that the highest rated technique is the ultimately best interviewing and interrogation technique in all situations. This relative comparison highlights their usability in current context in relation to the goal, the collection of usable and reliable information from a human source.
The following criteria were used in the comparison. A preferred technique should 1) Produce longer and narrative answers, 2) Support accurate memory retrieval, 3) Challenge inconsistencies in a constructive manner and 4) Not promote leading or suggestion. An Analytic Hierarchy Process (AHP) comparison table was used for rating. In addition to choosing one technique over another, a relative importance between the compared pairs was rated. A scale from one to nine, one meaning equal importance and nine meaning extreme importance in favor of the chosen one, was used. For example any information gathering oriented technique would be chosen over the polygraph test and they would be at least strongly favored with high importance rating\(^2\). The complete comparison table is presented in Appendix 1.

<table>
<thead>
<tr>
<th>QUESTIONING TECHNIQUE</th>
<th>STRENGTHS</th>
<th>POTENTIAL WEAKNESSES</th>
<th>WEAKNESSES</th>
<th>OTHER OBSERVATIONS</th>
<th>RANKING AFTER AHP (PAIR VICE COMPARISON)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cognitive Interview</td>
<td>-Emphasis for minimal amount of leading or suggestion -Produces long and narrative answers -Emphasis on accurate memory retrieval -Constructive challenging of inconsistencies.</td>
<td>-Contextual reinstatement and another person’s point of view might lead to suggested, potentially false or corrupted recollection.</td>
<td>-</td>
<td>More emphasis on memory retrieval in comparison to Information Gathering Interview. Is potentially less time consuming than</td>
<td>1 of 8 22.2 % weight</td>
</tr>
<tr>
<td>Elicitation</td>
<td>-No leading or suggestion -Source hard to anticipate the real interest of the interviewer -Produces long and narrative answers -No deliberate hampering of memory retrieval, if not improves it either -Constructive challenging of inconsistencies.</td>
<td>-Deliberate concealment of the topic might diminish the time reserved for the collection.</td>
<td>-</td>
<td>Focus is shifted to non-pertinent topics.</td>
<td>2 of 8 21.6 % weight</td>
</tr>
<tr>
<td>Information-gathering Questioning</td>
<td>-No leading or suggestion -Produces long and narrative answers -No deliberate hampering of memory retrieval, an open conversational atmosphere could promote it indirectly -Constructive challenging of inconsistencies.</td>
<td>-</td>
<td>-</td>
<td>Potentially less time consuming than elicitation.</td>
<td>3 of 8 20.6 % weight</td>
</tr>
<tr>
<td>Direct Questions and Repeated Questioning</td>
<td>-Contains little leading or suggestion -Narrative answers are preferred -Somewhat constructive approach to inconsistencies.</td>
<td>-Little support to correct memory retrieval -Limited amount of time could hinder veracity assessment -Control questions potential for veracity assessment is limited.</td>
<td>-</td>
<td>Too strong goal orientation could lead to ignorance of relevant information -Potential interviewer's confirmatory bias in repeated questions.</td>
<td>4 of 8 15.8 % weight</td>
</tr>
</tbody>
</table>

Table 1. Analyzed interviewing and interrogation techniques with highest ranking.

\(^2\) AHP data sheet calculates weights and rank using eigenvector method (EVM).
<table>
<thead>
<tr>
<th>QUESTIONING TECHNIQUE</th>
<th>STRENGTHS</th>
<th>POTENTIAL WEAKNESSES</th>
<th>WEAKNESSES</th>
<th>OTHER OBSERVATIONS</th>
<th>RANKING AFTER AHP (PAIR VICE COMPARISON)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strategic Use of Evidence</td>
<td>- In the beginning emphasis for minimal amount of leading or suggestion</td>
<td>- Some potentially suggestive elements due to assumption of information superiority.</td>
<td>- Questioning and disclosure tactics are leading</td>
<td>- They do not promote information collection per se.</td>
<td>5 of 8</td>
</tr>
<tr>
<td></td>
<td>- Initially produces long and narrative answers</td>
<td></td>
<td></td>
<td></td>
<td>11.1 % weight</td>
</tr>
<tr>
<td></td>
<td>- No special emphasis on accurate memory-retrieval, if not deliberate hampering either</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Carefully constructed ploy of challenging the inconsistencies.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kinesic Interview and Interrogation</td>
<td>-</td>
<td>- Heavy challenging of inconsistencies, though some of them could produce new information.</td>
<td>- Strong leading and suggestion</td>
<td>- Emphasis is on &quot;how things are said&quot; not &quot;what is said&quot;</td>
<td>6 of 8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- Does not make emphasis on correct memory retrieval.</td>
<td></td>
<td>3.2 % weight</td>
</tr>
<tr>
<td>Accusatorial Questioning (Reid)</td>
<td>-</td>
<td>- Heavy challenging of inconsistencies, though some of them could produce new information.</td>
<td>- Strong leading and suggestion</td>
<td>- Promotes short answers, preferably &quot;yes I did it&quot; or &quot;yes I am guilty&quot;</td>
<td>7 of 8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>- Does not make emphasis on correct memory retrieval.</td>
<td>2.6 % weight</td>
</tr>
<tr>
<td>Polygraph, i.e. Control Question Test (CQT) and Concealed Information Test (CIT)</td>
<td>- Enables the challenging of inconsistencies afterwards with additional backup.</td>
<td>- Some leading and suggestion in the form of tricked lying in control questions during the CQT</td>
<td>- Does not promote information gathering.</td>
<td>Method is intended for securing a confession during or after the test.</td>
<td>8 of 8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.7 % weight</td>
</tr>
</tbody>
</table>

Table 2. Analyzed interviewing and interrogation techniques that have some limitations.

The first supporting research question of this study was the following: What kinds of interviewing and interrogation techniques exist and how could they be used in the intelligence interview context? The comparison showed that all of the information gathering oriented methods gained rather high scores. This result was more or less expected on the basis of the ranking criteria. When calculated weights were taken into consideration a clear top three was highlighted. Cognitive Interview, Elicitation and Information-gathering Questioning all promote values that improve the collection in terms of the quantity and quality of the information. This result should not indicate that Direct Questioning is unusable. It is the most straightforward method of the top four and possibly the easiest to apply. Its usability can be improved when potential confirmatory biases from the control and repeated questions are identified and good balance between goal orientation and information collection is found.

According to this analysis, the Cognitive Interview is the most recommended questioning method with a small marginal compared to the two others in top three. The Information-
gathering Questioning and the Elicitation are potentially less suggestive than the Cognitive Interview but on the other hand they lack special methods for enhanced memory retrieval. The Elicitation benefits from the concealment of the actual area of interest but at the same time the focus shifts towards irrelevant topics. It is fair to conclude that all of the top three methods are usable and serve their purpose. Which method to choose is more about the specific needs of the situation than its absolute goodness. In some cases a combination of the three could be the most beneficial.

For the purpose of securing a confession or challenging the source for being deceptive the Strategic Use of Evidence shows the most potential. The foundation of this technique is based on the information superiority and it is used to achieve the goal step by step instead of rushing to the confession with heavy mental pressure. Due to the lesser amount of mental coercion and psychological trickery, the Strategic Use of Evidence is assessed to be more diagnostic and to produce less false confessions compared to Reid Technique or the polygraph. Current results strongly reflect the ranking criteria. If the criteria had been for example the likelihood of securing a true confession, the rating would obviously have been very different. However, in the context of this study, the information collection point of view is the most relevant one.

The analysis of the above mentioned interrogation and interview techniques highlight their difference not only from the information collection’s point of view but also from the veracity assessment’s point of view. The polygraph test is a lie detection test during which questioning is used to evoke physiological responses that indicate deception. The Strategic Use of Evidence is based on an assumption that the content of the truth tellers’ statements differ from the liars’. Guilt-presumption based accusatorial questioning methods are used to secure a confession with moderate mental pressure, and after initial questioning they pay very little attention to actual content of the statement or the veracity assessment. Whatever the embedded veracity assessment or lie detection method in the questioning technique is, it must be based on a theory that has a solid scientific foundation.
4 VERACITY ASSESSMENT METHODS

4.1 Multitude of Approaches to the Truth

In this study, lie detection is heavily in the focus and the topic is also widely covered in scientific literature. Entertainment industry has also benefitted from the glory and mystery that surrounds this topic. It is a double bladed sword since publicity subjects different lie detection methods to scrutiny but at the same time easy shortcuts and cheap gimmicks are sold by eager consultants. However, from the intelligence gathering point of view, the collection of trustworthy data and the understanding of the overall picture is the goal. The veracity assessment of the source and collected intelligence is merely a tool to achieve it.

Traditionally lie detection methods have been based on verbal and nonverbal cues. Studies show that people erroneously have high confidence on their abilities to detect deception and this misperception is widespread among professionals and lay persons. Still these methods are widely used and commonly relied on even though behavioral scientific research shows that they are not direct indicators of lying.

To mention few, hesitation or errors in speech, gaze aversion, smiling, eye blinking, illustrating with arms and shifting positions are not definite signs of deceit. Even flushing, paleness, changes in heart rate and licking of lips have been listed as indicators. All of the above mentioned behavior happens in the everyday life basically because humans are living beings who have thoughts and emotional experiences during their interaction.

---

93 Hazlett 2006, 47. Vrij and Granhag 2012, 111.
94 Neuman and Salinas-Serrano 2006, 197.
Different studies indicate several ways of assessing the veracity of a human source. Before the very first intelligence interview the background information of the event and related persons should be known as comprehensively as possible. The interrogation, and the intelligence interview, should be considered as the final phase of the intelligence gathering from a human source\textsuperscript{95}. As well as in the criminal investigation context, this basic principle applies to intelligence gathering where multiple intelligence disciplines are available.

Depending on the actual situation several intelligence collection methods\textsuperscript{96} can produce complementary pieces of intelligence, from independent sources, that might support veracity assessment. Information superiority is a great benefit. Being familiar with the past events and all of the available background information, also to ones that are related to the source, undoubtedly gives an advantage to the intelligence interviewer. For example, biographic intelligence produces deeper understanding of the interviewee’s background and persona which could increase interviewer’s chances of making correct interpretations and assessments of the source’s motives and behavior\textsuperscript{97}.

Finally the interviewer and the interviewee are engaged in the intelligence interview. The interviewer must be able to lead the dialog, observe and listen to the interviewee, plan ahead,

\textsuperscript{95} Neuman and Salinas-Serrano 2006, 177.
\textsuperscript{96} FM 2-0 2010, 1-22 and 1-23.
and assess the veracity of the source. The assessment of the plausibility and inner logic of the statement is a method that relies mainly on the deduction. A reasonable amount of background information must be available to make the assessment more reliable. Inconsistencies in the timeline and circumstances surrounding the main events could indicate deceit. Interviewee’s knowledge, language and appearance should match his/her story, position and past duties. Also a failure to answer a relatively easy question, that the interviewee is likely to know, might be an indication of a deceit.\textsuperscript{98}

The Human Intelligence Collector Operations manual instructs that repeated and control questions should be used to detect deceit. Behind this logic is an assumption that interviewee is mentally preoccupied and unable to remember his/her previous, presumably false, statements. Inaccuracy in the follow up questions means lying and accuracy means truthfulness. Also other explanations for inaccuracy are given. Misunderstanding the question, making up answers in order to please the interviewer, being truthful according to his/her best knowledge and outdated interviewer’s background knowledge may all produce \textit{a factually incorrect answer}.\textsuperscript{99} However, some research indicates that repeated interrogations may bolster interviewer’s initial perception of the source’s veracity. Repetition increases interviewer’s confidence regardless of the correctness of the assessment and thus possibly produces false confirmatory information.\textsuperscript{100}

In this context, the basis of the information collection is the interview. From this starting point it should not be a surprise that the questioning or the interviewing technique is a critical factor. Reviewed literature covers techniques like direct and repeated questioning, accusatory or quilt assumptive questioning technique, information gathering interview, different types of polygraph tests, the Behavior Analysis Interview and the Strategic Use of Evidence - to mention some. These questioning techniques have been discussed in Chapter 3.

Content analysis focuses on the content of the source’s statement – what is said. The aim of the analysis is to find elements that can be used to indicate the veracity of the statement. Description of feelings, the amount of detail, logical inconsistencies and spontaneous correction has been found to be related to the veracity of statements. The Statement Validity Assessment utilizes the Criteria Based Content Analysis as a tool to evaluate the content of the verbal communication. Reality Monitor can be used to assess statement’s validity in terms of clarity.

\textsuperscript{97} FM 2-22.3 2006, 5-11.
\textsuperscript{98} FM 2-22.3 2006, 9-6 and 9-7.
\textsuperscript{99} FM 2-22.3 2006, 9-3.
realism and other contextual information that support real life experience. These methods were originally developed to systematically judge statement from children that were suspected of being victims of abuse. Additional studies have indicated that these methods can also be used with adults.¹⁰¹

Discourse analysis offers a theory and tools to analyze how the source delivers his/her statements. In this study discourse analysis is approached from the linguistics’, communication’s and cognitive psychology’s point of view.¹⁰² Changes in the communicational pattern and deviations from the baseline can be identified through the analysis. The effectiveness of paralinguistic cues in lie detection has been critically discussed although studies have shown some promising results.¹⁰³ Indicators related to deceit have been found in several studies. Longer latency periods, more pauses, hesitations, speech errors and slower speaking have been found to indicate heightened cognitive load, which are interpreted as an attempt of deceit.¹⁰⁴ It must be emphasized that in this example deceit is a deduction from the observed signs of stress, mental effort of other factors. In essence signs of stress and mental effort could be identified by applying discourse analysis to the interviewee’s statement.

The Human Intelligence Collector Operations manual encourages the collector to pay attention to the source’s body language and physical cues during the conversation. Nonverbal signs such as nervous movement, sweating and inconsistency between verbal and nonverbal messages could indicate deceit.¹⁰⁵ Among the entertainment industry the monitoring of the nonverbal communication is perhaps the most popular method to catch lies. In many television series and silver screen movies scratching, touching different parts of face and gaze aversion is commonly used to underline someone’s untruthfulness.

However, studies indicate that only a limited amount of factors can be identified to be related to deceit. On the contrary to common beliefs the aforementioned cues are not reliable signs of deception. Instead liars tend to exhibit less arm, hand, finger, feet and leg movement than truth-tellers¹⁰⁶. Currently three major factors are reported to explain nonverbal signs of deceit:

¹⁰⁰ Granhag and Strömwall 1999, 163.
¹⁰¹ Hazlett 2006, 49.
¹⁰² Schiffrin, Tannen and Hamilton 2001, 1.
¹⁰³ Hazlett 2006, 49.
¹⁰⁴ Vrij et al. 2008, 255.
¹⁰⁵ FM 2-22.3 2006, 9-6 and 9-7.
¹⁰⁶ Neuman and Salinas-Serrano 2006, 197 and 198.
Emotional Arousal, Cognitive Load and Behavioral Control Hypothesis\textsuperscript{107}. Still, emotional arousal, high cognitive load and self-control can be argued to occur also in other occasions.

A plethora of advanced technical aids have been studied during the history of educating information and veracity assessment. Way before computers the first steps towards easy access to the truth were taken when narcotic substances, so called truth serums, were first administered. Narcoanalysis was based on the assumption that intoxication would provide truthful answers because the effect of the drug would make it difficult to lie. Nonetheless truth serums, like sodium amytal and sodium pentothal, do not seem to work that way. Although they typically make subjects more talkative and lower their inhibitions it is not guaranteed that the elicited information is true.\textsuperscript{108}

The polygraph test has been a real workhorse of lie detection over the years. A polygraph is a machine that basically is used to monitor and store the changes in the subject’s heart rate, blood pressure and skin resistance in relation to subject’s answers. A polygrapher, person using the polygraph, looks for a set of changes or lie responses in the measured data and determines when the subject is lying. Despite the polygraph’s known shortcomings, such as producing false positive results from innocent subjects, it is widely used for veracity assessment in law enforcement context.\textsuperscript{109}

In addition to the polygraph test, there are numerous other technically assisted methods for monitoring potential involuntary behavioral or neurobiological results of deceit. They all are more or less based on the hypothesis that a deviation from the subject’s baseline output is an indication of deceit. Almost all of the methods have been proven, or at least claimed, to be useful through field or laboratory tests although some of them have been identified to have significant constraints in their usability. Computer assisted facial action coding system and facial action unit identification, automated eye movement monitoring, computer voice stress analyzer, monitoring of the skin surface temperature, electrogastrogram, electroencephalography, magnetoencephalography, positron emission tomography, functional magnetic resonance imaging and functional near-infrared spectroscopy are among many of these modern day methods.\textsuperscript{110}

\textsuperscript{107} Hart, Fillmore and Griffith 2009, 135.
\textsuperscript{108} Heckman and Happel 2006, 73, 74 and 82. Ford 2006, 168.
\textsuperscript{109} Heckman and Happel 2006, 65 and 66.
\textsuperscript{110} Heckman and Happel 2006, 66-72, 75-78 and 80.
To sum up, it must be noted that most, if not all, of the above mentioned methods are indirect indicators of deceit. Most of the observable phenomena are signs of physiological, psychological or cognitive processes that are rated as abnormal or deviant, and are thus interpreted as signs of deceit. For this kind of deduction to be trustworthy the theories regarding human behavior and their relation to the deception must be based on a rock solid scientific foundation.

4.2 All-Source Intelligence Gathering

One of the most common arguments to legitimize the use of coercive interrogation methods has been the threat of an imminent catastrophe. In this *ticking bomb* scenario there is only little time to stop the catastrophic event from happening and the forced compliance of the *key informant* is going to lead to success. In real life this kind of scenario is very unlikely to happen and there is always more than one source of information available – human source or other. It must also be noted that forced compliance potentially produces false or partially true information, which veracity cannot be reliably assessed\(^{111}\). In addition, unverified single source information is always prone to deliberate manipulation even when coercion was not used\(^{112}\).

Not only could the all-source approach lead investigators, interrogators or intelligence officers to resolve the case through an alternative route, but careful background research also helps the planning of the interview and the veracity assessment on the whole. Both law enforcement and intelligence community seem to be unanimous about the importance of the background work that precedes the actual interrogation or intelligence interview. Facts connected to the event, related persons, their personal history and motivations undisputedly improve the interviewer’s ability to process and analyze any information that is obtained during the intelligence interview\(^{113}\).

Scientific studies concerning lie detection often concentrate on a very narrow field of phenomena that is related to a larger research problem. Only little, if any, background information that could help the observers to spot deceit is usually provided. This is absolutely necessary because scientific studies try to challenge popular beliefs and misconceptions of lie detection methods and their claimed scientific proof. Without this kind of an approach it would be impossible to isolate different factors in order to determine which method is usable and which one is not. Though, in real life this is not the case and investigators or intelligence officers do

---

\(^{111}\) Senate Select Committee on Intelligence 2014, 2, 3 and 13.

\(^{112}\) Borum, Gelles and Kleinman 2009, 114.
more than just passively watch video tapes and try to make razor sharp assessments based solely on that given information\textsuperscript{114}.

From the veracity assessment’s point of view it is essential to make an emphasis on the unique nature of a scientific research setup. In real life there are no deliberate limitations for the collection of the background information. What would be a better foundation for assessing the veracity of a human sources than knowing the right answers, or the most part of them, beforehand? This sounds tempting but when it comes to human sources it is very difficult, if not impossible, to reach a point where the interviewer can claim to know all what the interviewee knows. This kind of thinking can be traitorous and might lead the interviewer to discard relevant pieces of information that the source provides.

To conclude, it is not beneficial to conduct an intelligence interview without the collection of case specific background information. Only in the stories the hero enters a dark room, meets the source for the first time and rapidly elicits crucial information that prevents bad things from happening. On the other hand one should not give up the comprehensive approach and become narrow-minded or biased if the background information seems to be overwhelmingly exhaustive and give all the answers beforehand. However, in every case “Thorough and disciplined all-source analysis reduces the possibility of error, bias, and misinformation through the consideration of multiple sources of information and intelligence.”\textsuperscript{115}

4.3 Plausibility and Inner Logic of the Statement

The Human Intelligence Collector Operations manual instructs to pay attention to the level of the source’s cooperation and potential knowledge during the initial screening. At this stage the emphasis is on the competence assessment. During an interrogation the attention should be paid to the content of the interviewee’s statement. Internal inconsistencies in the events, timeline, surrounding circumstances, demonstrated knowledge in relation to the claimed position or duty, physical appearance and the use of language could indicate deception. Source’s motives for presenting certain information should also be taken into consideration. Only after the first interrogation is finished, it is possible to evaluate the source’s reliability in relation to the additional information from other sources. In this evaluation process the emphasis is on the

\textsuperscript{114} Vrij and Granhag 2012, 111 and 115.
\textsuperscript{115} FM 2-0 2010, 5-2.
confirmatory sources of information, past records of trustworthiness, inner logic and consistency of the information.\textsuperscript{116}

A human source can be analyzed from various perspectives in terms of his/her information value. One of them is competence and credibility. In order to be a competent source one must have made relevant observations or otherwise have access to the information under investigation. Competence should not be confused with credibility. Neither follows from the other. Three attributes of credibility are veracity, objectivity and observational sensitivity. On the contrary to common belief veracity does not refer to what really happened but what the source truly believes to has happened. A person cannot be classified as untruthful if he/she believed his/her statement to be true. Objectivity should be considered carefully because it is related to the observations. If the statement is based on nonbiased observation rather than beliefs or desires the level of objectivity is high. Observational sensitivity is closely connected to objectivity. If the source has based his/her knowledge on good sensory inputs it is reasonable to assume that the event took place as it was observed. During the analysis of a method called the Method for Assessing the Credibility of Evidence (MACE) these three attributes of credibility are assessed with 20 questions.\textsuperscript{117}

Studies show that the description of events in detail is typically more demanding for liars than truth tellers. It is assessed that for liars it could be more difficult to make up a detailed story that appears to be plausible.\textsuperscript{118} Real life law enforcement activities have indicated that under increased cognitive load the suspects gave themselves away with incredible statements filled with inconsistencies.\textsuperscript{119} The amount of detail, logical inconsistencies and spontaneous corrections has been found to vary with the statements’ veracity.\textsuperscript{120}

In the Human Intelligence Collector Operations manual credibility evaluation of the source is mainly based on the information from other independent sources, inner logic of the statement and consistency with other supporting information. In comparison, MACE’s questions fall into seven categories from which the most commonly used are confirmatory or contradictory information from other sources, internal consistency of the statement and assessment on the

\textsuperscript{116} FM 2-22.3 2006, 6-10 - 6-13, 9-6, 9-7, 12-4, B-1 and B-2.  
\textsuperscript{117} Schum and Morris 2007, 254-257, 260-263.  
\textsuperscript{118} Vrij et al. 2008, 255.  
\textsuperscript{119} Vrij et al. 2008, 263.  
\textsuperscript{120} Hart, Fillmore and Griffith 2009, 135.
source’s motives\textsuperscript{121}. Although MACE’s approach is more methodological and comprehensive compared to the simple approach of plausibility and logic, the MACE’s resemblance with the field manual’s directions is clear.

Anyhow, the plausibility and inner logic of the statement might not be enough for a reliable veracity assessment. The analysis of the source’s competence and credibility offers undeniably more depth to overall assessment. A human source, who is assessed to be reliable, should have access to the information, understand what he/she observes, stay objective and comprehensive during the observations and finally deliver the observed information as completely and accurately as possible. MACE’s analytical approach to the truth highlights the complexity of the lie detection and the search for the ground truth. Usually different lie detection methods are mostly concentrated on the veracity, i.e. does the source tell what he/she believes to be true?

Studies show that a well prepared liar is often perceived as consistent as a truth teller\textsuperscript{122}. Internally consistent and plausible statement becomes quickly externally inconsistent as the analysis of interviewee’s competence, motives and other supporting intelligence information is added to the equation. This highlights the importance of the all-source approach and a thorough background research, as it was emphasized in the previous chapter. The more the interviewee knows the easier it is to put new pieces of information into context.

During veracity assessment, that follows the guidelines of the Plausibility and Inner Logic of the Statement, the making of a correct assessment requires training, disciplined application and effort to avoid biases. A true statement must obey the laws of physics concerning time, space and related events. Timeline, actions, actors, motives and other factors can become implausible in fabricated stories. However, studies show that a well prepared lie usually is plausible and logically coherent. The assessment can be done during the collection by the interviewer or supporting personnel. Additional analysis can be conducted during and after the reporting. This method does not require the use of special equipment. The method can be applied covertly if inconsistencies are not over emphasized during the collection.

\textsuperscript{121} See Appendix 2. Seven different categories used in MACE for credibility assessment. Number indicates the amount of questions regarding that category: Other sources (6), internal consistency (4), motives (4), intentional external manipulation (2), value of sensory information (2), nonverbal communication (1) and memory (1).
Table 3. Preliminary analysis of the Plausibility and Inner Logic of the Statement.

<table>
<thead>
<tr>
<th>Accuracy</th>
<th>Ease of Use</th>
<th>Time Required</th>
<th>Need for Special Equipment</th>
<th>Unobtrusiveness</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>

Notes regarding rating:

A true statement must obey the laws of physics concerning time, space and related events. Timeline, actions, actors, motives and other factors can become implausible in fabricated stories. However studies show that a well prepared lie usually is plausible and logically coherent. The making of a correct assessment requires training, disciplined application and effort in avoiding biases. The assessment can be done during the collection by the interviewer or supporting personnel. Additional analysis can be done during the reporting. The method does not require the use of special equipment. The method can be applied covertly if inconsistencies are not over emphasized during the collection.

Rating scale: 1) “Presents noticeable limitations”, 2) “Presents some limitations”, 3) “Presents no significant limitations”.

4.4 Content

Content analysis is a research technique that concentrates on the artifacts of social communication. These artifacts are typically written documents or transcriptions of verbal communication. In essence content analysis is a technique that systematically and objectively identifies different characteristics of messages. Photographs, video recordings or any other items that can be transformed into written format can be subjected to content analysis. Both qualitative and quantitative methods can be applied. While qualitative analysis deals with the duration and frequency of the form, quantitative analysis deals with the forms themselves. Quantitative analysis provides means for identifying, organizing, indexing and retrieving data whereas qualitative analysis enables the researcher to learn about subject’s or author’s view of the social world. Quantitative sampling of the data can be done in several levels alone or simultaneously: with words, phrases, sentences, paragraphs, ideological stances, subject topics or other relevant elements of the data. Seven major elements in written messages, which are usually counted for analysis, are: words or terms, themes, characters (persons), paragraphs, items, concepts and semantics. Different elements of the content can also be sorted, categorized and

122 Granhag and Strömwall 1999, 163. See also Gödert et al. 2005, 236 and Blandón-Gitlin et al. 2009, 912. The ratings of Logical Structure of the Statement indicate no significant difference between the
classified for the purposes of qualitative analysis. Content analysis studies *what is said* and it can be applied on different levels from single words to broad concepts.\textsuperscript{123}

The Human Intelligence Collector Operations manual refers indirectly to content analysis when it suggests paying attention to human source’s use of language, sentence structure and vocabulary. These elements are presented in general level among the discussion of possible signs of deception. Neither specific tools for content analysis nor introduction to its theory is presented in the manual.\textsuperscript{124}

Criteria-Based Content Analysis (CBCA) is a veracity assessment technique that is an integral part of a broader veracity assessment method called the Statement Validity Assessment (SVA). Although the SVA was originally developed by German psychologists to assess the credibility of child witnesses’ testimonies in sexual abuse trials, the CBCA has been found to be suitable for the evaluation of the adults’ statements and is used in criminal investigation\textsuperscript{125}. The Reality Monitoring (RM) is slightly similar to the CBCA. The core theory is that memories of real experiences derive from sensory information. Thus statements of true events should include perceptual information like visual, auditory, spatial and temporal details. Also affective information, what someone felt, should be readily available\textsuperscript{126}. RM falls under two of the CBCA’s criteria. In this sense the CBCA is more comprehensive and covers wider variety of topics than the RM.

The SVA consists of three phases: a structured interview, the CBCA and the application of the Statement Validity Checklist. To minimize possible leading and suggestion, a free recall interviewing style is used. After that additional open ended questions are presented to fill possible gaps of information. During the second phase, the CBCA evaluation criteria are applied to the transcription of the interview. The application of the CBCA provides an estimate of the statement’s veracity. Each of the 18 criteria is viewed against the statement and the existence or absence is evaluated for example with a three or four-point scale\textsuperscript{127}. The presence of a criterion is in favor of the statement’s veracity. The last phase consists of the validity checklist. It guides the rater to assess several aspects that involve interviewee’s psychological characteristics, interviewing procedures, related motivational factors and other facts related to the inves-

\begin{footnotes}
\footnote{Berg 2001, 241-250.}
\footnote{FM 2-22.3 2006, 9-7.}
\footnote{Vrij et al. 2000, 239-240.}
\end{footnotes}
tigation. The overall evaluation of the statement’s veracity is based on all of the three phases.\textsuperscript{128}

The theory behind the CBCA is that, according to Undeutsch’s hypothesis, the representation of self-experienced memories differs from the fabricated ones in their verbal quality and content. 18 evaluation criteria have been generated from this notion\textsuperscript{129}. These criteria can be divided into three categories that are General Characteristics (three criteria), Specific Content of the statement (ten criteria) and Motivational Factors (five criteria)\textsuperscript{130}. During the analysis judges rate the statement according to the CBCA criteria. Each criterion is rated individually and in the end total sum is counted. The main finding from the studies is that true accounts of events score higher than the accounts of events that are considered to be false\textsuperscript{131}.

The classification of the account, as true or lie, is thus based on the case specific comparison between low and high scoring evaluations. Although there have been attempts to describe the minimum amount of existing criteria that validates truth, no definitive threshold is presented that is commonly agreed on or scientifically justified. It has been discussed that some criteria would be more significant than the others. For example some early studies highlighted a possible importance of the criterion 1: Logical Structure and 3: Quantity of Details. It was also emphasized that a certain number of criteria, at least three to five, must be present in a true statement\textsuperscript{132}. More recently some scientists have started to weight the presence of certain criteria although no formal or commonly excepted rules have been created\textsuperscript{133}.

The discriminating power of different criteria is actively studied. Gödert et al. (2005) found in their inter-rater reliability study in a mock-crime scenario that criterion 3: Quantity of Details, 6: Reproduction of Conversations and 9: Superfluous Details were effective in differentiating truthful accounts from fabricated ones. On a four-point scale from zero to three these criteria averaged about one point higher in truth tellers than liars. All the other criteria were found to be inadequate in their discrimination. It should also be noted that all of the assessed accounts

\textsuperscript{127} Three-point scale: 0 = absent, 1 = present and 2 = strongly present. Four-point scale: 0 = not present, 1 = slightly present, 2 = moderately present and 3 = strongly present.
\textsuperscript{128} Ruby and Brigham 1994, 15, 16, 36 and 66.
\textsuperscript{129} Originally there was also 19th criterion called Offense-Specific Elements. Criteria were also arranged in five categories where Peculiarities of the Content was extracted from the Specific Content and criterion 19 formed an independent category (Ruby and Brigham 1994, 18).
\textsuperscript{130} See the list of CBCA criteria in Appendix 3.
\textsuperscript{131} Blandón-Gitlin et al. 2009, 902 and 903.
\textsuperscript{132} Ruby and Brigham 1994, 24-25.
\textsuperscript{133} Blandón-Gitlin et al. 2009, 915.
were consistently considered to be logical (criterion 1).\textsuperscript{134} Vrij et al. (2004) studied the diagnostic value of four selected CBCA criteria as part of a broader study regarding rapid judgment of veracity. Results indicated that 3: Quantity of Details, 4: Contextual Embedding and 6: Reproduction of Conversations were effective\textsuperscript{135}.

Blandón-Gitlin et al. (2009) studied how well the CBCA discriminated between true events and suggested accounts of events that were believed to be true. They had somewhat different findings. In was found out that the CBCA scores discriminated between true, fabricated and suggested true events but a specific set of consistently efficient criteria was not found. The only clear similarity was that both true and fabricated accounts were assessed to be logical.\textsuperscript{136} The results regarding the discriminating power of individual criterion differ from the findings of a broad review from the first 37 published CBCA studies conducted by Vrij in 2005. This review reported that criteria 2: Unstructured Production, 3: Quantity of Details, 4: Contextual Embedding and 6: Reproduction of Conversations were found to be effective in most of the studies\textsuperscript{137}.

There are also some constraints in this method. The assessment should be based only on the content of the statement presented in written format. If additional information is presented, such as audio or video recordings, the objectivity that is built in the method is lost\textsuperscript{138}. Some studies show that familiarity of the event has a larger effect on the CBCA scores than the veracity of the statement\textsuperscript{139}. It has also been discussed that it could be very difficult to differentiate deceitful embedded events\textsuperscript{140}, false memories and suggested events from the truth\textsuperscript{141}. CBCA and other similar perceptually oriented methods also fail to determine the veracity of conceptual presentations, essentially what people truly think and what their real opinions are\textsuperscript{142}.

Intelligence interviewer could benefit from the use of content analysis both in the case of single and multiple human sources. In the case of multiple interviewees it would be possible to study systematically the similarities and differences of their statements that are related to the same event. Also sensitivity of certain topics and other latent meanings could be revealed. It

\textsuperscript{134} Gödert et al. 2005, 236 and 239.
\textsuperscript{135} Vrij et al. 2004, 17.
\textsuperscript{136} Blandón-Gitlin et al. 2009, 903, 911 and 912.
\textsuperscript{137} Vrij 2005, 10 and 11.
\textsuperscript{138} Ruby and Brigham 1994, 48.
\textsuperscript{139} Blandón-Gitlin et al. 2009, 903.
\textsuperscript{140} True events that have been experienced in some other situation and are added to another context.
\textsuperscript{141} Blandón-Gitlin et al. 2009, 904.
\textsuperscript{142} Leal et al. 2010, 323.
can be hypothesized that during an interview, that is based on free telling, preferred and avoided topics could be identified\textsuperscript{143}.

Some of the above mentioned rules also apply with a single interviewee. It can be hypothesized that during a free telling interview evasive answering techniques would produce inconsistency regarding accuracy, the level of detail and the timeline of the events. As the interviewee would be given a chance to speak freely preferred and non-preferred topics could be recognized by a simple word count.

A systematic content analysis could complement other veracity assessments methods like plausibility, consistency and source credibility. Although content analysis would not pinpoint or reveal the actual deception itself it would form a guideline for the following interviews and additional information gathering. The CBCA analyses the statement as a whole. Instead of dividing the statement in different topics, themes, characters, paragraphs and items the whole statement is subjected to the overall analysis regarding the CBCA criteria. Although the CBCA’s validity in discriminating truthful and deceptive statements has been proved, it could be argued that would the method be successful in identifying lies within otherwise true statement.

Valid content analysis requires that the content of the communication is transcribed prior to analysis. The transcription and the analysis must be completed by different persons in order to avoid biases and to maintain required validity. Transcription is a time consuming task and it must be completed by trained personnel. Transcription requires that the statement is either audio or video recorded. Nowadays this kind of equipment is easily available and their presence can be easily hidden if needed.

The CBCA has been scientifically proven to be adequately accurate. The main limitation is that the method is used to analyze the statement as a whole. It is questionable how well lies within mostly true statement can be detected. A successful application requires training. Disciplined application and the use of multiple raters are required to maintain a high level of validity. The application of the method requires a word by word transcription of the statement which is a time consuming task. Statements must be long because the method’s validity increases in relation to the amount of content. The method does not require the use of special equipment and it can be applied covertly.

\textsuperscript{143} Leal et al. 2010, 324.
Method: Criteria Based Content Analysis, CBCA

Theory: The Undeutsch Hypothesis
The representation of self-experienced memories differs from the fabricated ones in their verbal quality and content.

<table>
<thead>
<tr>
<th>Accuracy</th>
<th>Ease of Use</th>
<th>Time Required</th>
<th>Need for Special Equipment</th>
<th>Unobtrusiveness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rating:</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Notes regarding rating:</td>
<td>The method has been proven to be adequately accurate.</td>
<td>A successful application requires training. Disciplined application and the use of multiple raters are required to maintain a high level of validity.</td>
<td>The application requires a word by word transcription of the statement which is a time consuming task. Statements must be long because the method's validity increases in relation to the amount of content.</td>
<td>The method does not require the use of special equipment.</td>
</tr>
<tr>
<td></td>
<td>The main limitation is that the CBCA is used to analyze the statement as a whole. It is questionable how well lies within other vice true statement can be detected.</td>
<td></td>
<td></td>
<td>The method can be applied covertly.</td>
</tr>
</tbody>
</table>

| Rating scale: 1) “Presents noticeable limitations”, 2) “Presents some limitations”, 3) “Presents no significant limitations”. |

Table 4. Preliminary analysis of the Criteria Based Content Analysis.

4.5 Discourse

Speech is an integral part of social interaction. It does not only exchange information but also serves social purposes such as a transmission of emotions, thoughts and attitudes, and an invitation to interaction\(^\text{144}\). Discourse analysis in communication studies how people talk with each other, how they present themselves and persuade other people, just to give a few examples\(^\text{145}\).

Communication is making choices. People plan ahead their actions and messages so that they serve their personal communicational purposes. Communication research seeks to find out why speakers choose a certain way of talking, what purposes it might serve and what are the advantages or disadvantages of choosing a strategy over another\(^\text{146}\). Personal motivations and goals reflect directly to conversational choices.\(^\text{147}\)

In some situations discourse analysis and content analysis are overlapping. For instance differentiating the focus of a sentence from the presupposed information is done by studying

\(^\text{144}\) Harré 2001, 696.
\(^\text{145}\) Tracy 2001, 734 and 735.
\(^\text{146}\) Decisions might serve different identity, task or relationship functions.
\(^\text{147}\) Tracy 2001, 734-736.
what is said rather than how it was said\textsuperscript{148}. There are markers that convey additional information or connotations that reflect speaker’s attitude towards received information\textsuperscript{149}. Strategies to diminish inconsistencies within a statement can also be identified. Separating frames of references, generalization, averaging opposites and modifying terms all serve the same purpose of minimizing contradiction\textsuperscript{150}.

In their individual lives people need to feel themselves of being good, socially proper and stable persons. To achieve this they need to have coherent and acceptable life story which is constantly revised. In story telling people create identities and situate themselves in their social environment through choices made during the narrative. It has been found out that people tend to give away, unselfconsciously, more information than they intended as they narrate about their dangerous or embarrassing experiences. Studies on personal experience narrative (PEN) emphasize that clause can serve one or two functions, referential or evaluative. Referential clauses are related to what the story is about: events, characters and setting. Evaluative clauses have to do with why the listener should be interest in the story: they highlight the point of the story.\textsuperscript{151}

Transcripts of recordings are essential for discourse analysis. They provide a somewhat distilled documentation of the communication that took place in a certain place and time. The level of detail depends on the needs of the research. It is emphasized that transcripts are not unbiased documents of the full data. They are inherently selective and interpretive as the researcher decides what information should be preserved, discarded and what descriptive categories should be used. Transcription captures the flow of discourse events in written format. It should include who said what and to whom, in what manner it was said and under what circumstances. Coding subdomains can be for example pause length, intonation contour and syntactic category. In order to be useful, created categories must be systematically discriminable, exhaustive and usefully contrastive. In the end it is a matter of the researcher’s choice which elements of the information: \textit{words}, \textit{gestures}, \textit{pauses}, \textit{intonation}, \textit{turn taking and nonverbal aspect}, to code.\textsuperscript{152}

In the intelligence interview context deeper thoughts, valuations, motives and speaker’s position related to the setup could be assessed through discourse analysis. Also changes in the

\textsuperscript{148} Ward and Birner 2001, 120.
\textsuperscript{149} Schiffrin 2001, 63-64.
\textsuperscript{150} Norrick 2001, 89 and 90.
\textsuperscript{151} Johnstone 2001, 637, 640 and 641.
speech can be identified, stored systematically and connected to the content. A successful analysis requires a detailed and disciplined transcription of the conversation. Transcription and the actual analysis is a time consuming job that requires trained and skillful professionals.

The Multifactor Model suggests that changes in speaker’s vocal features could be explained by the Arousal Theory, the Cognitive Theory or the Attempted Control Theory. In general, heightened stress, emotional arousal and cognitive load should increase the speaking rate and the amount of hesitation and errors. On the other hand hyper articulation should decrease the amount of these signs as a result of increased control over communication.  

In lie detection research such features as the length of pauses, voice pitch, hesitation and the length of response have been studied as potential indicators of deception. Some of them have shown to be more potential than others. For example studies regarding high stake police interviews with real suspects showed that lying induced increased pauses. Also increased latency period, “ah” speech disturbances and speech rate have been found to correlate with deception both in laboratory studies and in real life situations. In contrast voice pitch has not been found to be a reliable indicator, or at least there have been some challenges to prove it to work in laboratory tests.

There are also studies that show less correlation between lying and speech hesitations, speech rate and speech errors. In a study, concerning the effectiveness of increasing cognitive load to facilitate lie detection, it was found that in the control group above mentioned criteria showed no significant discriminating power. It must be noted that the participants in the control group had received a thorough coaching and very detailed information about the actual event that they lied about. In another research concerning devil’s advocate approach, as an enhancing factor to lie detection, latency period appeared not to be a sufficiently discriminating factor. However, this contradicting result could be argued. The reason for identified inconsistency was truth tellers’ longer latency period when they spoke against their true opinion. Latency period of five seconds is abnormally long in otherwise fluent conversation.

153 Kirchhübel and Howard 2013, 695.
155 Vrij and Granhag 2012, 112.
156 Vrij et al. 2000, 251 and 254. They concluded that suspects seemed to show signs of cognitive load, which would be useful way of detecting deceit in the future.
159 Vrij et al. 2008, 256.
160 Leal et al. 2010, 325.
parison liars showed in average three second latency times in both opinion eliciting questions and devil’s advocate questions. Truth tellers’ average latency time was a bit over one second, which in real life situation is perceived as normal pause compared to three seconds, when they delivered their true opinions.

To sum up, discourse or vocal content analysis offers a way to assess interviewees’ statements in a scientific and systematic manner. A full and in depth discourse analysis could offer deeper insights to speakers’ motivations and values. Discourse analysis also has potential to indicate the areas of discussion where the interviewee is mentally challenged or where heightened control over communication is applied. Communicational strategies related to lying, like avoidance, generalization, averaging opposites and modifying terms, can be identified. Similar to content analysis, discourse analysis requires that the conversation is transcribed first. It is assessed that concentrating on the pauses, latency periods, “ah” speech disturbances and speech rate a lighter version of discourse analysis could be applied online while listening to the conversation as it happens or afterwards from an audio recording. Although that would not be scientifically as valid as the more systematic method, it would offer significant time saving and could be used as a preliminary screening method. Perhaps the most beneficial feature, compared to the content analysis which concentrates on the overall veracity assessment of the statement, is that according to theory, vocal indicators appear almost simultaneously with the deceptive communication.
Method: Discourse Analysis
Theory: Social constructivism and communication theory
The discourse of speech reveals sender’s values, motives, attitudes and possible dissonance related to the statement and lying.

<table>
<thead>
<tr>
<th>Accuracy</th>
<th>Ease of Use</th>
<th>Time Required</th>
<th>Need for Special Equipment</th>
<th>Unobtrusiveness</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

Notes regarding rating:
The method is potential in revealing the interviewee’s deeper values and motivations. Communicational strategies related to lying like avoidance, generalization, averaging opposites and modifying terms can be identified. However, lying is one possible reason for this kind of observable behavior.

A successful application requires extensive training in both transcription and analysis. Disciplined application and the use of multiple raters are required to maintain a high level of validity.

The transcripts of the communication, based on recordings, is essential for the discourse analysis. The application requires a word by word transcription of the statement with necessary coding of the discourse. This is a time consuming and labour intensive task. Statements must be long because the methods validity increases in relation to the amount of content.

The method does not require the use of special equipment. Although a professional transcription software could enhance post analysis and reduce the use of time during the transcription.

Rating scale: 1) “Presents noticeable limitations”, 2) “Presents some limitations”, 3) “Presents no significant limitations”.

Table 5. Preliminary analysis of the Discourse Analysis.

Method: Discourse Analysis, Studied Features
Theory: Multifactor Model: Arousal Theory, Cognitive Theory or Attempted Control Theory
Heightened stress, emotional arousal, cognitive load and attempted behavioral control causes changes in speaking rate, pauses, hesitation and errors.

<table>
<thead>
<tr>
<th>Accuracy</th>
<th>Ease of Use</th>
<th>Time Required</th>
<th>Need for Special Equipment</th>
<th>Unobtrusiveness</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>

Notes regarding rating:
Studies have shown that selected features of speech are adequately accurate.

A successful application requires training in the identification of selected features of speech. Disciplined application and the use of multiple raters are required to maintain a high level of validity.

The assessment can be done initially during the collection by the interviewer or supporting personnel. Additional analysis can be done after the collection which requires the recording of the statement.

The method does not require the use of special equipment.

The method can be applied covertly.

Rating scale: 1) “Presents noticeable limitations”, 2) “Presents some limitations”, 3) “Presents no significant limitations”.

Table 6. Preliminary analysis of the Studied Features of Discourse Analysis.
4.6 Nonverbal Communication

Spotting nonverbal signs of deceit has been a popular and widely used method to detect lies.\footnote{Vrij et al. 2000, 240.} According to common beliefs liars feel themselves uncomfortable and therefore, in an investigative interview, cross their legs, move about in their chairs, show grooming gestures and avoid eye contact.\footnote{Vrij and Granhag 2012, 112.} For example the Behavior Analysis Interview, as a means of detecting deception, relies on a premise that a specific set of questions evoke noticeably different physical responses in liars than truth tellers. The polygraph test is also based on a presumption that lying can be identified by monitoring test person’s breathing and sweating among heart rate and heart pressure.\footnote{Palmatier and Rovner 2014, 5.} It cannot be argued that observations of emotions, body movement, self-manipulation or autonomic physiological responses are false per se, but a straight forward deduction, that lying is the only reason for these kinds of phenomena, is over simplified.\footnote{Ford 2006, 167 and 169. Frank and Svetieva 2012, 132.}

The Human Intelligence Collector Operations manual recognizes nonverbal elements of communication and their possible relation to deception. Sweating and nervous movement might be indicators of deceit, but it is emphasizes that other factors, such as the questioning itself or strong emotional response to the topic, could cause this kind of behavior.\footnote{FM 2-22.3 2006, 9-7.}

Based on the Multifactor Model there are three major theoretical frameworks that explain observable nonverbal behavior.\footnote{See for more detail Zuckerman, DePaulo and Rosenthal 1981, Verbal and nonverbal communication of deception.} According to the Arousal Theory liars are more excited than the truth tellers. This excitement and resulting emotions such as fear, guilt and delight, induce noticeable changes in one’s verbal and nonverbal behavior. The Cognitive Theory is based on observations that lying is cognitively more demanding than telling the truth. Truth is the norm and it is automatically activated. For a deceiver it requires extra effort to suppress the truth and express something that is in contradiction to that. The lie must be plausible and consistent so that the liar does not get caught. In addition to monitoring own behavior, liar must also monitor the receiver to get feedback from his/her success. People have quite stereotypical views about how a sincere person behaves. Good liars try act by these preconceptions and thus must control their behavior accordingly. This leads to the Attempted Control Theory that ex-
plains why liars fail to act naturally and leave a rigid impression while trying to suppress what they think resembles a guilty person’s behavior.167

Although the above described three major theories more or less comprehensively explain what causes people to behave the way they do in certain conditions, these theories do not implicitly tell how people behave when they lie. A meta-analysis in 2003 conducted by DePaulo et al. showed that a great deal of commonly accepted verbal and nonverbal cues to deception were unreliable. Anxiety-based lie detection has continued to dominate nonverbal lie detection and it still relies on some of the above mentioned cues.168

The Orienting Response and the Preliminary Process Theory are understood to be in the core of anxiety and stress based deception detection. They explain why and how a person reacts to a stimulus, which is connected to his/her semantic and autobiographical memories and is highly significant in relation to personal motivations and emotions. The more significant a stimulus is the larger the elicited response becomes. This theory justifies the chain of deduction that connects polygraph questions, or other behavior evoking questions, to successful veracity assessment.169 Paleness, flushing, sweating, changes in breathing, restlessness and expression of deception associated emotions, openly or via micro expressions, have been considered to be signs of stress caused by lying170. There is still room left for misinterpretations. How does the person conducting a lie detection test know that observed emotions or stress are welling from anxiety caused by lying and not some other highly emotional issues that are closely related to the event?

Studies regarding real life high stake police interviews revealed that decreased blinking, hand and finger movement were related to lying171. Police officers who saw the interviews also reported that the suspects seemed to be thinking harder when they lied compared to telling the truth172. Several aspects have been found to potentially contribute to the increase of cognitive load during lying: the formulation of the lie itself, doubts about own credibility which produces self-monitoring and control, monitoring of the interviewer’s reactions, efforts to stick to a chosen role, suppression of the truth and activation of the lie173.

167 Kirchhübel and Howard 2013, 695.
168 Vrij and Granhag 2012, 110, 112.
169 Palmatier and Rowner 2014, 5.
170 Neuman and Salinas-Serrano 2006, 197.
171 Leal et al. (2008) discussed that decrease in blinking rate could be a result of attempted control, because blinking is commonly related to suspicious behavior.
172 Vrij and Granhag 2012, 112.
Research has revealed that people tend to move less when they are thinking hard. This decrease of overall animation is the results of shifting the cognitive focus from body movement to mental work. While thinking hard people also tend to look away, because eye contact is found to distract thinking. Here also resides a chance for deliberate misleading of the receiver. Since it is well-known that gaze aversion makes a suspicious impression, liars may intentionally try to avoid it in order to appear credible.\textsuperscript{174}

In a study concerning the effect of increased cognitive load it was found out that liars exhibited more eye blinking and leg/foot movement. These signs were identified as signs of nervousness rather than cognitive load. In addition, liars showed less hand and finger movement than truth tellers, which has been observed to occur in several other studies. A decrease in hand and finger movement was assessed to be related both to cognitive load and an effort to make a credible impression.\textsuperscript{175}

The changing of eye focus from one place to another is called saccadic eye movement. There are two types of saccadic eye movement. Visual saccadic movement occurs when a person changes the direction of visual attention. Non-visual saccadic movement does not include focusing the vision to some particular object and it seems as if the eyes are wondering around. Non-visual saccades occur when people are engaged in a task that involves information retrieval from the long term memory. Results from a field study show that spontaneous lying requires more long term memory retrieval than telling the truth and thus produces higher average eye movement rate. However, no significant difference in saccadic eye movement rate between planned lie and truth was found. It was discussed that, in the case of a planned lie, perhaps most of the long term memory retrieval task was completed during the rehearsal. Essentially, non-visual saccadic eye movement is related to long term memory retrieval.\textsuperscript{176}

No exclusive and reliable pattern of behavior has been found to be related to lying\textsuperscript{177}. It has been even suggested that intense cognitive effort reduces otherwise visible signs of nervousness\textsuperscript{178}. However, some general guidelines can be identified. Reduced hand/finger movement has been found to be connected to lying\textsuperscript{179}. Reason for that could be increased cognitive load

\textsuperscript{174} Vrij et al. 2008, 256.  
\textsuperscript{175} Vrij et al. 2008, 259.  
\textsuperscript{176} Vrij et al. 2014, 1, 3 and 4.  
\textsuperscript{177} Vrij et al. 2000, 241.  
\textsuperscript{178} Leal et al. 2008, 5.  
caused by lying or attempted self-control. Increased amount of blinking and decreased amount of illustrating hand gestures are additional, but possibly less reliable, nonverbal indicators of lying\textsuperscript{180}. In spite of the doubts regarding blinking as a trustworthy indicator of lying Leal and Vrij (2008) found out that, when lying required cognitive effort, blinking decreased during the deception and increased right after that as a compensatory effect\textsuperscript{181}.

Gaze aversion, self-adapting gestures, smiling and increased leg movement have not been found to correlate reliably with lying\textsuperscript{182}. Some of the poor past deception detection performance via nonverbal cues can be explained by false beliefs about the relationship between lying and nonverbal behavior\textsuperscript{183}. It is also important to notice that, once these signs of deception are known, it is possible to become a more convincing liar by practicing to avoid the most obvious deception related nonverbal behavior\textsuperscript{184}. Another possibility is to deliberately add them when misleading cues are serving the liar’s needs.

False stereotypical beliefs of deceptive nonverbal behavior most likely cause false alarms that reduce the accuracy of lie detection. For that reason cues that have been found to be unreliable like gaze aversion, self-adapting gestures, smiling and increased leg movement should not be included in veracity assessment procedures. Reduced hand/finger movement has been found to be reliable sign of lying. Increased blinking and decreased illustrating hand movement could be another two potential indicators, but their accuracy is questionable.

Reliable and valid observation of nonverbal signs requires scoring of the observed behavior. Video recording is thus necessary in order to avoid possible errors that could occur during online coding. Coding is potentially time consuming task if it is done carefully. The need for special training is assessed to be low. The scorer simply counts the selected phenomenon. The setting of the scoring threshold, i.e. what is considered to be hand/finger movement or illustrating gesture, might be challenging. Also because the method explicitly requires that the changes in the frequency in the observed criteria are assessed, a specific personal baseline must be established.

To sum up, the analysis of nonverbal behavior requires that observed signs are scored, stored and preferably linked to the content of the statement. Scoring is time consuming and is as-

\textsuperscript{181} Leal and Vrij 2008, 187.
\textsuperscript{183} Vrij et al. 2000, 240.
sessed to require only little training. Scoring requires that the statement is video recorded. Nowadays this kind of equipment is easily available and their presence can be easily hidden if needed. To reduce the total time between the statement and results, preliminary coding could be done online. The most significant benefit would be the time saving but it would be achieved at the expense of reliability and would probably require higher training of the scorer.

Traditional signs of deception like gaze aversion, self-adapting gestures, smiling and increased leg movement have not been found to correlate with lying. The detection of these signs causes a high level of false alarms. These will not be subjected to the further analysis because they have evident doubts about their scientific proof.

Research has shown that Studied Features of Nonverbal Communication are accurate for lie detection purposes. These features are reduced hand/finger movement, increased amount of blinking and decreased amount of illustrating hand gestures. A successful detection of any kind of nonverbal communication in tandem with the interview requires training and experience. Recording, disciplined application and the use of supporting personnel are preferred to maintain a high level of validity. The method does not require the use of special equipment. The method can be applied covertly.

<table>
<thead>
<tr>
<th>Method: Nonverbal Communication, Traditional</th>
<th>Theory: Multifactor Model: Arousal Theory, Cognitive Theory or Attempted Control Theory</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>Ease of Use</td>
</tr>
<tr>
<td>Rating: 1</td>
<td>2</td>
</tr>
<tr>
<td>Notes regarding rating:</td>
<td>Traditional signs of deception like gaze aversion, saccadic eye movement, self-adapting gestures, smiling and increased leg movement have not been found to correlate with lying. The detection of these signs causes a high level of false alarms.</td>
</tr>
</tbody>
</table>

Rating scale: 1) "Presents noticeable limitations", 2) "Presents some limitations", 3) "Presents no significant limitations".

Table 7. Preliminary analysis of the Traditional Signs of Nonverbal Communication.

184 Vrij, Mann and Fisher 2006, 329.
### Table 8. Preliminary analysis of the Studied Features of Nonverbal Communication.

<table>
<thead>
<tr>
<th>Accuracy</th>
<th>Ease of Use</th>
<th>Time Required</th>
<th>Need for Special Equipment</th>
<th>Unobtrusiveness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rating: 3</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>

**Notes regarding rating:**
- Studies have shown that selected features of nonverbal behavior are adequately accurate. These are reduced hand/finger movement, increased amount of blinking and decreased amount of illustrating hand gestures.
- A successful detection of any kind of nonverbal communication in tandem with the interview requires training and experience. Recording, disciplined application and the use of supporting personnel are encouraged to maintain a high level of validity.
- The assessment can be done initially during the collection by the interviewer or supporting personnel. Additional analysis can be done after the collection which requires the recording of the statement.
- The method does not require the use of special equipment.
- The method can be applied covertly.

Rating scale: 1) "Presents noticeable limitations", 2) "Presents some limitations", 3) "Presents no significant limitations".

### 4.7 Credibility as a Measure of Veracity

In many cases when people are trying to detect deception, and claim to evaluate the person’s veracity, they actually make judgments of the person’s credibility\(^{185}\). If the judgment is done comprehensively, the sender’s competence is also assessed\(^{186}\). Possibly the most striking example of taking the shortcut, from credible impression to the evaluation of veracity, is the criteria used during the Behavioral Symptom Analysis (BSA). In general the BSA describes truthful people, or suspects in criminal investigation, to be spontaneous, sincere, helpful, concerned and cooperative – and the deceptive ones the opposite\(^{187}\). In 2011 edition of the Reid Manual truthful subjects are still listed to be direct, confident, definitive, spontaneous and open to communication although the scientific proof for this kind of reasoning is non-existing\(^{188}\).

\(^{185}\) George, Tilley and Giordano 2014, 2.
\(^{186}\) Schum and Morris 2007, 247.
\(^{187}\) Neuman and Salinas-Serrano 2006, 185, 187-190.
\(^{188}\) Hirsch 2014, 816-818.
In a series of studies it was found out that people based their deception detection mainly on the demeanor of the sender. When demeanor matched with the veracity, in the cases of sincere truth tellers and insincere liars, the accuracy rate was high. On the other hand when the demeanor and the content of the message mismatched, in the cases of sincere liars and insincere truth tellers, the accuracy rate was low. Because the veracity judgment was mainly based on the experienced credibility the actual truthfulness or deceptiveness had little to do with the results. This supports the previous findings of Bond and DePaulo (2008) that a person’s perceived credibility has a larger impact on the judgment than the actual honesty. If a person is good at evoking a general impression of honesty and truthfulness he/she is most likely perceived as honest and thus successful in delivering any kind of messages. 189

In a study regarding lie detection Leal et al. (2010) made an experiment in which the participants (receivers) were watching video recordings of truthful and deceptive statements. They were given a task to rate senders’ perceived talkativeness, plausibility, immediacy and involvement. These criteria were derived from the meta-analysis of deception detection studies by DePaulo et al. (2003) and Vrij (2008). 190 Even though these criteria bear resemblance to the ones used in the BSA they were not used to rate the behavior of the sender. Instead, the emphasis was on the content of the message, and the plausibility of the statement was rated. In contrast, the BSA is used to evaluate the sender’s credibility regardless of the content.

The Prominence-Interpretation Theory (Fogg, 2003) explains how people make credibility judgments online related to websites. Basically people must first notice the message (Prominence), in this case the information on the site, and after that make a judgment about what they have noticed (Interpretation). Motivation, interest towards to the topic, given task, expertise and individual differences all affect Prominence. Three factors are thought to affect Interpretation. They are presumptions, knowledge and competence in the subject matter of the site, and context or the situation related factors. 191 To sum up, the information is rated credible if the content of the website supports the receiver’s pre-established ideas and values, presents facts that are in concert with the receiver’s knowledge and are presented in un-contradicting manner. An analogy to the interpersonal persuasiveness can be found. Sender’s credibility and persuasiveness is respectively affected by the receiver’s personal preferences, presumptions of competence related cues and the lack of context related contradictions. Perceived authority, familiarity, similarity and personal appeal are confidence building factors in the interpersonal

189 George, Tilley and Giordano 2014, 1-3.
190 Leal et al. 2010, 326.
191 Georg, Tilley and Giordano 2014, 3.
communication, and these factors also lead the receiver to become more trusting towards the sender\textsuperscript{192}.

If needed, senders can learn to become more convincing and practice how to create a credible impression of honesty. Studies have shown that demeanor based credibility does not correlate with honesty. When a person is good at evoking a general impression of honesty and truthfulness he/she is most likely perceived as honest, and thus is successful in delivering any kind of messages.

Above mentioned examples highlight that deception detection, which is based on personal credibility, is prone to deliberate manipulation. One can always become more convincing by exhibiting credibility related cues and by practicing how to make a good impression of one self. The Method for Assessing Credibility of Evidence (MACE) addresses this issue. Schum and Morris (2007) have reported on the development of a systematic approach that is used to analyze the items of information provided by a human source. In their study they utilized different probability methods to assess how much evidence there is about a human source, how completely does it answer questions about the source's competence and credibility, and how strong is the gathered evidence about this particular human source\textsuperscript{193}. One relevant finding was that competence does not mean credibility, although these two separate elements are often confused. In legal context competence denotes the personal qualification of the witness and the credibility his veracity\textsuperscript{194}.

Source’s access to the information and the understanding of the observations has been identified as two major attributes of competence. There are five questions that are used to assess the source’s level of competence: did this source actually make the observation being claimed or have access to the information reported, does this source have an understanding of what was observed or have any knowledge or expertise regarding this observation, is this source generally a capable observer, has this source been consistent in his/her motivation to provide information and has this source been responsive to inquiries? Answers must be based on the existing evidence. The evidence favors, disfavors, does not enable the assessment of competence or is non-existent.\textsuperscript{195}

\textsuperscript{192} Cialdini 2007, 141-147, 216 and 219. Fein 2009, 24 and 27.
\textsuperscript{193} Schum and Morris 2007, 247.
\textsuperscript{194} Schum and Morris 2007, 254-255.
\textsuperscript{195} Schum and Morris 2007, 255, 260 and 261.
The credibility is also systematically assessed according to three main criteria. The credibility consists of veracity, objectivity and observational sensitivity under the conditions of observation\textsuperscript{196}. Based on existing evidence 20 questions are answered with the same set of alternative answers as in the competence assessment. The list of questions can be summarized to consist of seven main categories:

- Internal consistency of the source (in six questions),
- Support from other sources (in five questions),
- Source’s motives to alter the facts (in four questions),
- Intentional external manipulation of the source (in two questions),
- Testimonial value of sensory information (in two questions),
- Ability to remember accurately (in one question) and
- Observations about source’s nonverbal communication (in one question)\textsuperscript{197}.

It is worth noting that three criteria out of seven, internal consistency of the source, ability to remember accurately and nonverbal behavior, are to some extent in the source’s control and thus available for intentional manipulation. A well-rehearsed and convincingly delivered cover story can be internally consistent and to some extent in concert with the information obtained from other sources.

MACE offers a method that is largely based on confirmatory information from other sources and the assessment of the source’s motives or external interferences. The method has got a comprehensive approach to veracity assessment and objectivity is one of its goals. Judgment criteria concerning competence and credibility are analytic and cover relevant aspects of the assessment. There is minimal amount of chances for deliberate manipulation or the exploitation of sender’s perceived honesty and credibility. From this point of view the assessment on the credibility of evidence, as a reliable indicator of veracity, appears promising. It is based on the existing evidence that is collected from multiple sources, disciplined documentation and broad approach from seven different perspectives.

The veracity assessment in MACE is based on the gathered evidence which is subjected to assessment. The analysis requires additional information collection from other sources and trained personnel to complete the assessment. The collection of supporting information can be

\textsuperscript{196} Schum and Morris 2007, 260-261.
\textsuperscript{197} Appendix 2. The total number of questions is 21 because one question is considered to fall into two categories.
time consuming. The credibility assessment of the evidence can be conducted completely only after all the relevant facts are known.

<table>
<thead>
<tr>
<th>Accuracy</th>
<th>Ease of Use</th>
<th>Time Required</th>
<th>Need for Special Equipment</th>
<th>Unobtrusiveness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rating: 1</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Notes regarding rating:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Studies have shown that demeanor based credibility does not correlate with honesty. If a person is good at evoking a general impression of honesty and truthfulness he/she is most likely perceived as honest and thus successful in delivering any kind of messages.</td>
<td>Not assessed because of the method's lack of scientific proof.</td>
<td>Not assessed because of the method's lack of scientific proof.</td>
<td>Not assessed because of the method's lack of scientific proof.</td>
<td>Not assessed because of the method's lack of scientific proof.</td>
</tr>
</tbody>
</table>

Table 9. Preliminary analysis of the Demeanor Based Credibility.
Method: Method for Assessing Credibility of Evidence, MACE

Theory: General deduction, probability modelling, multiple independent sources

In order to be a credible source of information one must have access to information, must be generally a capable observer, must understand the observations, must not have a hidden agenda and be responsive to the inquiries.

<table>
<thead>
<tr>
<th>Accuracy</th>
<th>Ease of Use</th>
<th>Time Required</th>
<th>Need for Special Equipment</th>
<th>Unobtrusiveness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rating: 3</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>

Notes regarding rating:

The method is based on the gathered evidence which is subjected to assessment. Assessment criteria is:
Internal consistency of the source, support from other sources, source’s motives to alter the facts, intentional external manipulation of the source, testimonial value of sensory information, ability to remember accurately and observations about source’s nonverbal communication.

The analysis requires additional information collection from other sources and trained personnel to complete the assessment.

The collection of supporting information can be time consuming. The credibility assessment of the evidence can be conducted completely only after all the relevant facts are known.

The assessment itself does not require the use of special equipment. Supporting software could enhance the analysis.

The method can be applied covertly.


4.8 Advanced Technical Aids

The polygraph is possibly the best known lie detection machine. It has been widely used in the assessment of crime suspects’ veracity. Although the polygraph is just a part of a more comprehensive protocol it could be argued that the actual machine has become the icon of the procedure. Research involving the polygraph test has a long and rich history which stretches to the late 19\textsuperscript{th} and early 20\textsuperscript{th} century. Even today the underlying theory is vividly discussed while debate for and against its validity is continuing. Essentially, the polygraph measures and records different physiological symptoms of stress, that the suspect experiences during the questioning. The polygraph test relies on a premise that lying causes a person to become emotionally aroused and anxious. Suspects are expected to slightly sweat, have pauses in breathing, experience changes in blood pressure and heart rate when they lie. Although different
polygraph protocols have been developed, such as Control Question Test and Concealed Information Test, they all are based on the Arousal Theory. A trained and experienced polygrapher is needed for a successful polygraph test. Despite the very high accuracy scores in laboratory studies and its regular use within North-American law enforcement agencies it is widely argued that a great deal of uncertainty is related to the polygraph test. The polygraph test is solely based on an assumption that during the test lying and nothing else makes every subject anxious.

The polygraph test requires a controlled environment that is relatively calm in order to avoid additional sources of stress that could invalidate the test result. It is not possible to conduct the test covertly because instruments that measure skin conductance, breathing rate, blood pressure and heart rate must be connected to the suspect. The test cannot be conducted in parallel with preliminary hearing or information gathering questioning and thus requires a separate event for it.

The accuracy of the polygraph test is a controversial matter. Several studies have raised doubts about its reliability despite its defender’s claims of 80% or even 98% average accuracy. On the moderate side are studies that have found the average accuracy to be 67-75%, and generally the test produces higher accuracy with liars than truth tellers. This kind of dichotomous choice with the average accuracy of 75-80% might become problematic in the intelligence gathering context.

Statistically one out of four or five sources would receive a false assessment of being deceptive or truthful on the contrary of their real status. It could be argued to what extent this becomes a problem. In the intelligence gathering context the source is always initially considered unreliable until proven otherwise. Anyhow, false negatives would enable one fourth or fifth of the liars to continue feeding lies and fabricated stories as their status is confirmed reliable. Over time a reliable but deceptive source would also start to decay the value of other intelligence that was collected from sincere sources. In return false positives would result in the rejection of one fourth or fifth of the truthful sources. In this context the effect of false positive is less concerning as it decreases the amount of potentially reliable information. In con-

198 Heckman and Happel 2006, 65 and 66. See also Chapter 3.9.
199 Palmatier and Rovner 2014, 5 and 7. Kirchhübel and Howard 2013, 695
201 Vrij and Granhag 2012, 110.
203 A false positive result.
trast, a false negative would be more harmful because deceptive source could earn extra credit over the other ones who perhaps did not go through the lie detection test. If the polygraph was used, the biggest threat would be a total reliance on the results and thus the locking of a source’s status permanently as a liar or truth teller.

Other technical aids for lie detection can be divided into two categories. Those which detect and highlight the physiological stress or anxiety symptoms and those which detect and highlight the results of increased cognitive load. They all are potentially as valid as the underlying Arousal or Cognitive Theory of lie detection, and they all are prone to errors or interference to some extent. Most of them are obtrusive and some are even physically intrusive. Table 11 summarizes a collection of additional technical methods of lie detection. It must be emphasized that all these methods require specially trained users or medical personnel.

204 A false negative result.
### Table 11. Description of different possible technical aids to lie detection. \*205

<table>
<thead>
<tr>
<th>METHOD*</th>
<th>TECHNOLOGY*</th>
<th>LIE DETECTION THEORY</th>
<th>LIE DETECTION APPLICATION*</th>
<th>OBLTRUSIVENESS*</th>
<th>SCIENTIFIC SUPPORT TO METHOD</th>
<th>OTHER OBSERVATIONS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multi channel polygraph</td>
<td>Factors like heart rate, breathing, sweating and blood pressure are monitored with different sensors.</td>
<td>Arousal Theory</td>
<td>Arousal related to lying can be identified by monitoring changes in heart rate, breathing, skin conductivity and blood pressure.</td>
<td>Cannot be applied covertly.</td>
<td>Accuracy is under scientific discussion.</td>
<td></td>
</tr>
<tr>
<td>Facial Action Coding System, FACS</td>
<td>Visible facial muscle movements are recorded with a video camera and measured with the control computer and image recognition technology.</td>
<td>Arousal Theory</td>
<td>Arousal related to lying can be identified via expressions.</td>
<td>Can be applied covertly.</td>
<td>Life research on automated means of lie detection.</td>
<td>Micro expression detection is a variation of this method.</td>
</tr>
<tr>
<td>(Computer) Voice Stress Analyzer, (C)VSA</td>
<td>Inaudible speech characteristics are monitored, recorded and presented in visual charts.</td>
<td>Arousal Theory</td>
<td>Arousal related to lying can be identified via speech audio.</td>
<td>Can be applied covertly.</td>
<td>Research does not support method's validity.</td>
<td>Deliberately developed alternative to lie detector.</td>
</tr>
<tr>
<td>Skin Surface Temperature, SST</td>
<td>Infrared camera is used to detect and record changes in facial blood circulation.</td>
<td>Arousal Theory</td>
<td>Arousal related to lying can be identified by monitoring blood circulation in the facial area.</td>
<td>Can be applied covertly.</td>
<td>Life research on automated means of lie detection.</td>
<td></td>
</tr>
<tr>
<td>Electrogastrogram, EGG</td>
<td>Variations in stomach muscle activity are measured with Electodes that are attached to the stomach skin.</td>
<td>Arousal Theory</td>
<td>Arousal related to lying causes abnormal stomach muscle activity.</td>
<td>Cannot be applied covertly.</td>
<td>Is prone to error caused by other normal bodily activities.</td>
<td></td>
</tr>
<tr>
<td>Radar Vital Signs Monitoring, RVSM radar</td>
<td>Factors like heart rate, breathing and eye blinking are monitored with extremely high frequency radar.</td>
<td>Arousal Theory</td>
<td>Arousal related to lying can be identified by monitoring changes in heart rate, breathing and eye movement.</td>
<td>Can be applied covertly.</td>
<td>Optional method for polygraph.</td>
<td></td>
</tr>
<tr>
<td>Functional Magnetic Resonance Imaging (fMRI)</td>
<td>Changes in cranial blood circulation, that are related to lying, are measured with cooled superconducting magnets.</td>
<td>Cognitive Theory</td>
<td>Cognitive work related to lying activates specific parts of the brain which is detected with high energy magnetic field.</td>
<td>Cannot be applied covertly.</td>
<td>Is prone to error caused by head movement during the scan.</td>
<td>Superconducting magnets and other equipment are extremely heavy and MRI requires a very controlled environment to be accurate.</td>
</tr>
<tr>
<td>Electronencephalography, EEG</td>
<td>Changes in neural electric activity are monitored with electrodes attached to the face and scalp. Weak electric signals are stored and processed in a computer.</td>
<td>Cognitive Theory</td>
<td>Cognitive work related to lying activates specific parts of the brain which is detected with high sensitivity electronic instrument.</td>
<td>Cannot be applied covertly.</td>
<td>Is prone to error caused by facial muscle activation.</td>
<td>Requires controlled environment.</td>
</tr>
<tr>
<td>Magnencephalography, MEG</td>
<td>Changes in neural electric activity are monitored with high energy magnet placed near head. Weak electric signals are stored and processed in a computer.</td>
<td>Cognitive Theory</td>
<td>Cognitive work related to lying activates specific parts of the brain which is detected with high sensitivity electronic instrument.</td>
<td>Cannot be applied covertly.</td>
<td>Is prone to error caused by movement of the eyes and body.</td>
<td>Equipment is large, heavy and requires shielded room to prevent magnetic disturbance.</td>
</tr>
<tr>
<td>Positron emission tomography (PET)</td>
<td>Radiactive tracer is injected into subjects blood stream. Changes in cranial blood circulation, that are related to lying, are measured with positron scanner that detects changes in tracer activity scanning methods. Equipment is large, heavy and method requires the attendance of special medical personnel.</td>
<td>Cognitive Theory</td>
<td>Cognitive work related to lying activates specific parts of the brain which is detected with positron scanner.</td>
<td>Is physically intrusive because radioactive tracer injection is needed.</td>
<td>Little research on this method's accuracy.</td>
<td>Equipment is large, heavy and method requires the attendance of special medical personnel.</td>
</tr>
<tr>
<td>Functional near-infrared spectroscopy, NIRS</td>
<td>Specific wavelength light is transmitted to the scalp. Changes in cranial blood circulation, which related to lying, are measured with near-infrared camera or spectrometer.</td>
<td>Cognitive Theory</td>
<td>Cognitive work related to lying activates specific parts of the brain and that is detected with a special infrared light and camera.</td>
<td>Cannot be applied covertly.</td>
<td>Is optional to fMRI but is able to monitor only outer layers of the brain.</td>
<td>Is smaller compared to other brain scanning equipment.</td>
</tr>
<tr>
<td>Transcranial magnetic stimulation, TMS</td>
<td>Cranial activity is monitored with the use of high energy magnetic field.</td>
<td>Cognitive Theory</td>
<td>Cognitive work related to lying activates specific parts of the brain and that is detected with an electromagnetic coil on the scalp.</td>
<td>Cannot be applied covertly.</td>
<td>Little research on this method's accuracy.</td>
<td>Is optional to other brain activity scanning methods. Equipment is smaller compared to the other methods. TMS has been reported to cause seizures.</td>
</tr>
</tbody>
</table>
tistics automatically. Data processing is done according to theory that deceptive narratives differ from true ones by their content. Linguistic cues to deception originate from theories like the Reality Monitoring, the Interpersonal Deception Theory, the Information Manipulation Theory and the Self-Presentational Perspective of deception. Findings from a study by Fuller, Biros and Delen (2011) indicate moderate to high potential with 68 to 74 % overall accuracy in real life criminal cases. The Interpersonal Deception Theory and the Self-Presentational Perspective of deception more or less concentrate on the sender’s credibility via nonverbal cues. It can thus be argued that to what extend sender’s pleasantness, positivity, easiness, immediacy and non-deceptive demeanor are transmitted to the text. The connection to the Behavioral Analysis Interview and leakage theory is close. On the other hand, the study showed that the thirteen most significant cues to deception fall into category of content analysis and the Criteria Based Content Analysis.

Most of the technically aided methods are used to enhance human’s ability to detect different signs of deception. These methods are successful in identifying and recording different physiological parameters that are potentially caused by stress, anxiety, cognitive load or other psycho-physiological factors. Despite their potential accuracy of detecting and measuring different physiological artifacts they are only as reliable and valid as the underlying scientific proof that connects deception to the observed physiological phenomena. They also require the use of sophisticated and complex technical equipment that need special operating personnel with extensive professional experience. Most of them are basically improved lie detectors. Thus, even if proven accurate, they are poor in promoting information collection during an intelligence interview. From this point of view, computer assisted content analysis possibly offers more insight to lie detection and could potentially increase the applicability of content analysis.

To sum up, in spite of advertised very high accuracy rates, studies have shown that usually 25% of false negatives and false positives occur. In general the test produces higher accuracy with liar than truth tellers. However, one liar out of four would be assessed to be honest and could thus implant incorrect data to the collection. The test requires trained and experienced personnel to operate the polygraph and to interpret the results. The test setup must be created separately. The polygraph test can be conducted only after the information gathering questioning is finished.

Fuller, Biros and Delen 2011, 8393, 8394 and 8397.
Fuller, Biros and Delen 2011, 8394. George, Tilley and Giordano 2014, 3.
Method: Polygraph
Theory: Arousal Theory
Lying causes a person to become emotionally aroused and anxious. Liars are expected to slightly sweat, have pauses in breathing, experience changes in blood pressure and heart rate when they lie.

<table>
<thead>
<tr>
<th></th>
<th>Accuracy</th>
<th>Ease of Use</th>
<th>Time Required</th>
<th>Need for Special Equipment</th>
<th>Unobtrusiveness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rating:</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Notes regarding rating:</td>
<td>Studies have shown that in average 25% of false negatives and false positives occur. In general the test produces higher accuracy with liar than truth tellers. However one liar out of four would be assessed as honest and could thus implant incorrect data to the collection.</td>
<td>Requires trained and experienced personnel to operate the polygraph and interpret the results.</td>
<td>The test setup must be created separately. Polygraph test can be conducted only after the information gathering questioning.</td>
<td>The method is based on the use of multi-channel polygraph that measures and records physiological symptoms.</td>
<td>Can not be conducted covertly.</td>
</tr>
</tbody>
</table>

Table 12. Preliminary analysis of the Polygraph Test.

4.9 Conclusion

The second and third supporting research questions regarding veracity assessment methods were: What kinds of lie detection and veracity assessment methods exist that are reliable and are based on scientific proof, and what kind of uncertainty and other limitations are included in these methods? Methods presented in Chapter 4 are analyzed here in the light of these questions. After a short summary each of the covered method’s reliability and possible limitations for deception detection are discussed.

Demeanor based credibility assessment, which is thought to have a direct connection the veracity of the sender, is prone to sender’s deliberate manipulation and has a very weak scientific background with low diagnostic value. The same limitations apply also to the other methods that rely exclusively on subject’s nonverbal communication. The Behavior Symptom Analysis, and other similar methods, relies on a very simple and short chain of deduction where there is no room for alternative explanations. Simply stated nervousness, avoidance, uncooperativeness and hesitation mean deception, and in criminal investigation context deception...
tion means guiltiness\textsuperscript{209}. Despite the efforts to connect different behavioral theories\textsuperscript{210} with demeanor based methods, the scientific proof and diagnostic value are still low. However, real life and laboratory studies have found reduced hand/finger movement to be connected to lying\textsuperscript{211}. Increased amount of blinking and decreased amount of illustrating hand gestures are additional but potentially less reliable cues\textsuperscript{212}. Gaze aversion, self-adapting gestures, smiling and increased leg movement have not been found to be reliable signs of deception\textsuperscript{213}.

Evidence based sender credibility assessment offers a systematic way to analyze the information provided by a human source. The Method for Assessing Credibility of Evidence (MACE) recognized the negative effects of observatory biases and demeanor based credibility assessment. The emphasis is on an analytic approach to the source’s competence and credibility. The competence evaluation focuses on the access, understanding, expertise, observatory capabilities, motivations and cooperation. The evaluation of credibility, which consists of veracity, objectivity and observational sensitivity under the conditions of observation, is based on seven categories: internal consistency of the source, support from other sources, source’s motives to alter the facts, intentional external manipulation of the source, testimonial value of sensory information, ability to remember accurately and observations about source’s nonverbal communication\textsuperscript{214}. The evaluation scale is set in a way that it acknowledges the lack of evidence as an approvable alternative. This method’s strength is that it fosters objectivity, comprehensive evaluation and the use of alternative sources to confirm the veracity. The lie detection part in MACE is based on probabilities, plausibility, logic and multiple sources. The emphasis is on the content of the message and not how it is delivered. Although no comprehensive results from field or laboratory studies were presented to back this method’s claimed effectiveness, it is included in further analysis because its scientific background comes from Baconian and Bayesian probability methods\textsuperscript{215}.

Any events that are described by a source must obey the laws of physics. Facts regarding the course of events, related people, surroundings and timeline must be able to exist in real world\textsuperscript{216}. Based on this premise it is assumed that logical inconsistencies or contradictions with the laws of physics indicate that there is something wrong with statement. Lying or fabri-

\textsuperscript{209} Hirsch 2014, 816-818.
\textsuperscript{210} The Orienting Response, the Preliminary Processing Theory, the Arousal and the Cognitive Load hypothesis.
\textsuperscript{214} Appendix 2.
\textsuperscript{215} Schum and Morris 2007, 247 and 266.
cation could be a reason for that. There are also other explanations. Misunderstanding the question or the answer, other errors in communication or negligent note taking can distort the statement and make it unbelievable. It is essential to first note the flaws in plausibility and logic, and then to find out what causes them. If no other explanation exists the probable cause is then deception. Goodness of this method relies on the disciplined application of deduction and bias free search for alternative explanations. When applied correctly with a scientific working approach this kind of deductive method is reliable. The threat in this method is that lying becomes the only possible explanation for all inconsistencies and therefore too hasty conclusions are drawn about source’s veracity.

The content of the speech has been found out to vary in relation to its veracity. Changes in the content can be identified within words, phrases, sentence structures and previously mentioned plausibility or logical consistency. In addition to that, according to Undeutsch’s hypothesis, the representation of self-experienced memories differs from the fabricated ones in their verbal quality and content\textsuperscript{217}. The Statement Validity Assessment (SVA) offers a broader concept according to which testimonial credibility is assessed. In the core of the SVA is Criteria-Based Content Analysis (CBCA). To assess source’s veracity, the CBCA’s 18 evaluation criteria are applied to the transcription of the interview. The presence of a criterion is in favor of the statement’s veracity\textsuperscript{218}. It has been noted that some of the CBCA’s criteria are more effective in their discriminating power. Studies show that in different scenarios Unstructured Production, Quantity of Details, Contextual Embedding and Reproduction of Conversations were found to be the most effective\textsuperscript{219}. All of the statements were also judged to be internally consistent and logical regardless of their veracity\textsuperscript{220}. To sum up, a true statement is expected to include the above mentioned elements along with the Reality Monitoring related affective and perceptual information like visual, auditory, spatial and temporal details\textsuperscript{221}. One major restriction with the above mentioned, content based analyzing methods, is that they assess the statement as a whole. Instead of dividing the statement in different topics, themes, characters, paragraphs and items the whole statement is subjected to the overall analysis regarding the criteria. Although the CBCA’s validity in discriminating between truthful and deceptive statement has been proven it could be argued whether the CBCA is successful in identifying lies

\textsuperscript{216} FM 2-22.3 2006, 9-6 and 9-7.
\textsuperscript{217} Gödert et al. 2005, 225 and 226.
\textsuperscript{221} Vrij, Edward, Roberts, Bull 2000, 239-240.
within otherwise truthful statement. In this kind of scenario the CBCA’s temporal diagnostic value is questionable.

Discourse analysis in communication is used to study how people talk with each other, how they present themselves and persuade other people\(^\text{222}\). The goal is to find out why speakers choose to talk in a certain way and what purposes it might serve\(^\text{223}\). According to the theory personal motivations and goals reflect directly to conversational choices\(^\text{224}\). In the intelligence interview context, deeper thoughts, valuations, motives and speaker’s position can be assessed through the Discourse Analysis. Changes in the speech can be identified, stored systematically and connected to the content. A successful analysis requires detailed and disciplined transcription of the conversation. Transcription and the actual analysis are time consuming tasks which require trained and skillful professionals. The Multifactor Model suggests that changes in speakers’ vocal features could be explained by the Arousal Theory, the Cognitive Theory or the Attempted Control Theory. Heightened stress, emotional arousal and cognitive load should increase the speaking rate and the amount of hesitation and errors. On the other hand hyperarticulation as a result of increased control over communication should decrease the amount of these signs\(^\text{225}\). In lie detection research indicators like increased pauses, increased latency periods, “ah” speech disturbances and slower speech rate have been found to have a reliable connection with lying\(^\text{226}\). Potentially the most beneficial aspect of the Discourse Analysis compared to the content analysis is that, according to the theory, vocal indicators appear almost simultaneously with the deceptive communication. This would significantly increase its diagnostic value over content analysis. It must be noted that most of the above mentioned cues are explained by different aspects of the Multifactor Model. The judgment of the source being deceptive is based on deduction that expressed stress, arousal, cognitive load or self-control means lying. When applied with scientific discipline, this kind of deductive method could be effective in identifying the parts of the statement where sender experiences unusual discomfort or is forced to think harder than usual. The threat is that lying can become the only possible explanation for observed deviant behavior and therefore a false judgment about veracity is made.

\(^{222}\) Tracy 2001, 734 and 735.
\(^{223}\) Decisions might serve different identity, task or relationship functions.
\(^{224}\) Tracy 2001, 736.
\(^{225}\) Kirchhübel and Howard 2013, 695.
\(^{226}\) Vrij et al. 2008, 255.Vrij et al. 2000, 251 and 254. They concluded that suspects seemed to show signs of cognitive load, which would be useful way of detecting deceit in the future.
The polygraph, and most of the other advanced technical deception detection methods, is used to enhance human ability to detect different physiological signs of deception. These machines are undeniably effective in identifying and recording different parameters that are potentially caused by stress, anxiety, cognitive load or other psycho-physiological factors. They require the use of sophisticated and complex technical equipment which need special operating personnel with extensive professional experience. Despite the sensitivity to detect and measure different physiological artifacts, these machines and protocols are only as reliable and valid as the underlying scientific proof that connects observed physiological phenomena to lying. Most of them can be classified as modern day lie detectors. Like with the polygraph, their highest value could be during the last stages of collection when the source’s veracity regarding very specific topics is checked. It must be emphasized that they are not exceptionally good in promoting information gathering during an intelligence interview.

The biggest drawback of the polygraph test, and possibly other lie detection tests with similar methodological background, is the comparative high rate of false assessments. Statistically one out of four or five sources would receive a false assessment of being deceptive or truthful. It can be argued to what extent this would become a problem. In the intelligence gathering context the source is always initially considered unreliable until proven otherwise. False negatives would enable one fourth or fifth of the liars to continue feeding lies and fabricated stories as their status is confirmed reliable. Over time a reliable but actually deceptive source would also start to decay the value of other intelligence that was collected from sincere sources. In return false positives would result in the rejection of one fourth or fifth of the truthful sources. In this context the effect of a false positive is less concerning as it decreases the amount of potentially reliable information. A false negative would be more harmful because a deceptive source could earn an extra credit over the other who did not go through the lie detection test. If the polygraph was used, the biggest threat would be a total reliance on its results and thus the locking of a source’s status permanently as a liar or truth teller. In addition to the fact that there is no guarantee for 100% reliable test result, also the status of the source’s veracity could change at some point during the following collection.

In addition to the methods explained in Chapter 4, some other means to enhance lie detection have been studied. Increasing the interviewee’s cognitive load during the information collection has been found to highlight some of the cues that originate from the cognitive demand of

---

227 A false positive result.
228 A false negative result.
lying. Recalling events in reversed order, giving a Devil’s Advocate opinion and maintaining an eye contact with the interviewer have been found usable ways of increasing cognitive load\textsuperscript{229}. As long as lying is reliably diagnosed to be the reason for observed cognitive load during the interview these methods potentially increase the likelihood of successful lie detection. Also collective interviewing of suspects has been proven to be promising in distinguishing true statements from the fabricated ones because truthful pairs, or groups, of suspects are observed to work more as a team and add more information to each other’s stories\textsuperscript{230}.

To conclude, veracity assessment methods that were found not to have evident doubts about their scientific background were carried to the applicability analysis in Chapter 5. Gaze aversion, self-adapting gestures, smiling and increased leg movement as nonverbal signs of deception and Demeanor Based Credibility as an overall veracity assessment method were excluded.

![Picture 4. Results from the preliminary analysis of the studied veracity assessment methods.](image)

\textsuperscript{230} Vrij et al. 2012, 43.
5 APPLICABILITY OF THE VERACITY ASSESSMENT METHODS

5.1 Operational Environment and Implications to Applicability Criteria

According to the results received from the preliminary analysis, the Plausibility and Inner Logic of the Statement, the Criteria Based Content Analysis (CBCA), the Discourse Analysis, the Studied Features of Discourse\textsuperscript{231}, the Studied Features of Nonverbal Communication\textsuperscript{232}, the Method for Assessing Credibility of Evidence (MACE), and the polygraph test (further referred as the Polygraph)\textsuperscript{233} are all included in the applicability analysis in this chapter. Traditional signs of deception and sender credibility assessment based on demeanor are excluded because they have been found to have evident weaknesses in their scientific foundations. A collection of Advanced Technical Aids were initially analyzed in Chapter 4. The ones that were found to have high potential are represented in the applicability analysis by the Polygraph. Common to all is the need for special equipment, need for trained operating personnel and somewhat questionable scientific proof. In every aspect the Polygraph is not significantly weaker than some other advanced technical method\textsuperscript{234}. Respectively the Discourse Analysis represents a full and detailed content analysis in the following analysis. All of the basic ways of lie detection\textsuperscript{235}, complemented with methods based on multiple sources and general deduction, are thus included in the following analysis.

In general the following rules have been applied to the comparison of applicability criteria. The very basic requirement is that the veracity assessment method must be accurate in the sense that it is based on a real and proven scientific theory and the theory is valid in explaining the observable results of lying. Ease of use is important but it is acknowledged that more or less trained professionals are used to apply the method. It is not considered a deficit if the assessed method is somewhat demanding to apply and training is needed. In some scenarios

\textsuperscript{231} Increased pauses, increased latency periods, increased “ah” speech disturbances and slower speech rate.

\textsuperscript{232} Reduced hand/finger movement, increased amount of blinking and decreased amount of illustrating hand gestures.

\textsuperscript{233} The polygraph test is considered to reflect the traits of other physiological lie detection methods that are based on the Arousal Theory.

\textsuperscript{234} All of the following ratings have been conducted from the polygraph test's point of view. Although the results represent the Polygraph, they also give implications of the applicability of other Advanced Technical Aids.

\textsuperscript{235} Vrij et al. 2000, 239. Three principle ways to catch liars: (1) by observing how they behave, (2) by listening to what they say, and (3) by measuring their physiological responses.
time is of the essence. One method over another should be preferred if it can be applied during or directly after the interview and does not require extensive human resources or time. Further analysis that takes place after the interview and is time consuming can be accepted in certain scenarios. The method should be unobtrusive in order to avoid any interference with the information collection. One method over another should be preferred if it does not require the use of special equipment that cannot be bought from a typical appliance store.

The most important criterion is thus Accuracy and in this sense more accurate methods should always be chosen over others. Evident inaccuracy or major doubts about validity has led to exclusion of the method, as explained earlier in the conclusions of Chapter 3. The need for complicated special equipment can become a limitation in real life situations which makes No Need for Special Equipment the next important criterion. It has been recognized that in some situations time might be a limiting factor. Time Required was selected as the third important criterion. Here in this comparison Ease of Use and Unobtrusiveness are seen as equally important and their relative order of importance is more related to the situation. All the other criteria than Accuracy are here seen more or less case dependent.

It has been widely recognized that human intelligence gathering can be significantly improved with team work and the use of supporting personnel during and after the collection. The rating of Time Required was done with the supposition that a small supporting team was available.

A baseline applicability analysis was conducted without relative weighting between criteria. To highlight pros and cons of the criteria, which are related to possible real life limitations, two simple scenarios were described as a starting point for the second round of applicability analysis. To achieve desired sensitivity relatively high weighting was used.

- Scenario one: Time available for the veracity assessment is limited. This results in a stronger preference of faster methods over the ones that are more time consuming or labor intensive. To highlight this time related demand a relative weight of 10 was given to criterion Time Required.

- Scenario two: There are no time related limitations. Instead there is no access to special equipment and their operating personnel. This resulted in a stronger preference of methods that can be applied without any special equipment or with the help of basic
audio-video recording equipment that is available for everyone. To highlight this limitation a relative weight of 10 was given to criterion *No Need for Special Equipment.*

An Analytic Hierarchy Process (AHP) comparison table was used for rating. Overall comparison was divided into five independent comparisons according to applicability criteria. AHP comparison table connects these five sub comparisons into one assessment with relative weights that are explained within the scenarios.

5.2 Applicability Analysis

5.2.1 Analysis Based on Equally Weighted Criteria

Equally weighted AHP comparison\(^{236}\) revealed clear top two of the methods and three additional ones that have some limitations in their applicability. The Studied Features of Discourse and the Studied Features of Nonverbal Communication achieved the highest ratings and were seen as equally applicable. The Plausibility and Inner Logic of the Statement, MACE and the CBCA formed a group of second best methods. The Discourse Analysis and the Polygraph were assessed to be least applicable.

The Studied Features of Discourse and the Studied Features of Nonverbal Communication both were rated high in terms of all five criteria. They both were rated significantly higher than others within two criteria: *Ease of Use* and *Time Required.* Application of both of these methods requires training but they are more straightforward compared to the others. Both of these assessment methods can be applied during the collection by the interviewer or the supporting team.

The Plausibility and Inner Logic of the Statement, MACE and the CBCA received almost similar ratings. From the *Accuracy’s* point of view MACE is the most comprehensive one of these three. It uses additional information from other sources, assesses statement’s consistency, source’s motivations and other factors. This kind of approach makes MACE more resistant to well prepared, fabricated stories than Plausibility an Inner Logic of the Statement. The most significant limitation of the CBCA is that it assesses the statement as a whole. If existing lies are not covered adequately in the collection or they are otherwise insignificant in the overall context of the statement, it is likely that the CBCA does not indicate their existence.

\(^{236}\) Appendix 5, Table 1.
The most significant limitations of the Discourse Analysis can be found in *Ease of Use* and *Time Required* although it has got high potential in *Accuracy*. A successfully applied discourse analysis requires a full transcription of statement, both content and discourse, and highly trained personnel to do it. The whole process is time consuming and labor intensive.

The Polygraph was not particularly successful in the light of any criteria. Perhaps the biggest drawback, compared to the other methods, is that it must be applied as a secondary means of veracity assessment. The polygraph test must be preceded by traditional interrogation and preliminary veracity assessment until doubts about the source’s deception are raised and the need for the test becomes evident. In addition to time related limitations, the Polygraph’s possible false negative and false positive results can become problematic.

![Graph](image)

*Picture 5. Relative applicability of the studied veracity assessment methods with equally weighted criteria.*
To conclude, both qualitative and quantitative applicability analysis that was based on non-weighted criteria revealed that all of the top five methods have no or only minor limitations in their overall applicability. The Studied Features of Discourse and the Studied Features of Nonverbal Communication are the ones with the least amount of limitations.

5.2.2 Sensitivity Analysis, Scenario One

With a relative weight of 10 placed on *Time Required*, the AHP comparison\(^\text{237}\) revealed the same top two methods as the non-weighted analysis did. The Studied Features of Discourse and the Studied Features of Nonverbal Communication achieved the highest ratings and were seen as equally applicable. The least applicable was the Discourse Analysis.

\(^{237}\) Appendix 5, Table 7.
This evidently abnormal weighting was expected to highlight the far ends of the scale, which was achieved. In this analysis the Polygraph’s relative applicability was higher than in the equally weighted one, and it was seen to be almost as applicable as MACE and the CBCA.

5.2.3 Sensitivity Analysis, Scenario Two

With a relative weight of 10 placed on No Need for Special Equipment the AHP\textsuperscript{238} comparison revealed that the applicability of the Plausibility and Inner Logic of the Statement, MACE, the Studied Features of Discourse, the Studied Features of Nonverbal Communication and the CBCA received almost the same relative applicability. This result reflects the Polygraph’s dependence on the lie detecting machine. Rather limited applicability of the Discourse Analysis in this scenario is explained by the assumption during the AHP rating that a successful and efficient discourse analysis requires the use of special computer software for the transcription and coding. If that task could be completed manually with reasonable effort the relative applicability would be rated higher. MACE also performed slightly poorer in this scenario for the same reason than the Discourse Analysis. During the AHP rating it was assessed that the use of special computer software would improve MACE’s applicability. To conclude, this scenario added only minimal amount of information to the equally weighted analysis.

\textsuperscript{238} Appendix 5, Table 8.
5.3 Results from the Applicability Analysis

Based on the description and preliminary analysis conducted in Chapter 4 the selected methods were submitted to applicability analysis. Veracity assessment methods’ different methodological and theoretical approaches to lie detection or veracity assessment made it very difficult to establish definitive ranking from the Accuracy’s point of view. However, during the AHP rating extra credit was given to the comprehensiveness of the Discourse Analysis. It was also recognized that the Plausibility and Inner Logic of the Statement, the CBCA and the Polygraph have some potential shortcomings based on the scientific background, vulnerability to manipulation or diagnostic value.

Overall applicability assessment that was based on five selected criteria produced a ranking list of the studied methods. Additional sensitivity analysis was conducted by giving high weight to one criterion that was described in a simple scenario. Weighting of the Time Required accented the initial result. Weighting of the No Need for Special Equipment did not add any information to the initial analysis, but verified the order of the five most promising ones.
The main research question of this study was: What is the applicability of veracity assessment methods, which are reliable and are based on scientific proof, in terms of the following criteria\textsuperscript{239}?

Based on the analysis it is concluded that, the Studied Features of Discourse and the Studied Features of Nonverbal Communication share first place as the most applicable veracity assessment method. Second place is also shared. The Plausibility and Inner Logic of the Statement, MACE and the CBCA do not have any critical deficits but are assessed to have some limitations compared to the most applicable ones. The Discourse Analysis and the Polygraph were assessed to be the least applicable.

Common to the Plausibility and Inner Logic of the Statement, the Studied Features of Discourse and the Studied Features of Nonverbal Communication is that they can be applied during the collection by the interviewer or other supporting personnel. A successful application evidently requires training and personal abilities for multitasking if the assessment is done alone by the interviewer. If available, supporting personnel can be used for the assessment simultaneously with the collection if access to audial and visual information is established. In addition, audio-video recordings can also be used for a post analysis.

Recording of the statement, at least audio, and transcription is imperative for the application of the CBCA. It is also assessed that reliable and valid application of MACE requires the recording of the statement and transcription.

The most significant limitation of the Discourse Analysis is its labor intensiveness. A successful application requires full transcription of statement, both content and discourse, and highly trained personnel to do it. The Polygraph was not particularly successful in the light of any criteria. In addition to its possible weaknesses regarding false negative results, another significant drawback is that it must be applied as a secondary means of veracity assessment. The polygraph test must be preceded by traditional interrogation and preliminary veracity assessment until the need for the test becomes evident.

\textsuperscript{239} Accuracy, i.e. probability of detecting deception successfully, \textit{Ease of Use}, i.e. easiness to apply the method correctly, \textit{Time Required} to apply the method reliably, \textit{No Need for Special Equipment}, \textit{Unobtrusiveness} of the method.
In the introduction of this study it was hypothesized that no single method is accurate and reliable by itself and higher accuracy can be achieved by using multiple methods side by side. The first part of this hypothesis was confirmed as scientific studies have not yet shown that some particular veracity assessment method is overly accurate and reliable. Since during the veracity assessment content, discourse and nonverbal information can be isolated as individual channels of information it is reasonable to argue that increased accuracy could be achieved by using the most promising methods simultaneously and independently. This particular topic was not addressed directly in this study although Vrij et al (2000 and 2004) have had promising results from this kind of approach.

The second hypothesis was that overall accuracy could be improved with the latest technical aids but their implementation is possibly obtrusive and requires special equipment with oper-
ating personnel. This was also partially confirmed, although the assumption of improved accuracy can be argued. It is clear that the latest technical equipment, which was discussed in Chapter 4 have unique ways of producing otherwise inaccessible information. However, the diagnostic value of this additional information could be argued. These pieces of technical equipment are successful in identifying and recording different physiological parameters that are potentially caused by stress, anxiety, cognitive load or other psycho-physiological factors. Despite their potential accuracy of detecting and measuring different physiological artifacts they are only as reliable and valid as the underlying scientific proof that connects the observed physiological phenomena tightly to deception. Most of the advanced technical methods are obstructive and they must be used by specially trained operating personnel or a physician.
6 FIELD STUDY

6.1 Test Setup

An interview exercise was arranged during which about four hours of audio-video material was recorded in nine different interviews. This exercise was part of a professional skills development course where students were trained to improve their interpersonal communication. The goal of the interviewing exercise was to get warmed up for further activities and get familiar with the effects of increased cognitive load and multitasking. There were 15 participants and no one had previous training for lie detection. The exercise took place in a classroom.

The interviewer’s task was to find out as much as possible about how the interviewee had spent his/her previous weekend. The timeframe was narrowed down to cover only four hours of either Saturday or Sunday. Interviewees had the freedom to select the timeframe, but it had to include some relevant activities and other people’s presence. Interviewing someone about driving alone in a car to the movies and back would not have served the purpose.

The interviewee was instructed to be as normal as possible in his/her communication and not to stall the interview because only a limited amount of time was reserved for the exercise. The interviewee was also reminded that the story must be based on the actual events on that day. The goal was to create a conversation where the most of the content would be true and based on actual events. The interview was divided into two sessions. First session would last for 20 minutes and after a 5 minute break the interview would continue for another 10 minutes.

After it was made sure that participants had understood their basic tasks, it was brought to everyone’s attention that they had additional tasks. The interviewee’s secondary task was to exclude a half an hour long period from his story. If the interviewer referred to that time period the interviewee was supposed to stick to the story and try to avoid giving any information regarding it. The interviewee was also instructed to include one lie to his/her story. Lie did not have to be a major one, but the interviewee was encouraged to make it significant enough so that it would become under discussion during the interview.
The interviewer's secondary task was to detect where the time gap was and what was the topic the interviewee was lying about. The instructor emphasized that the primary task was to find out how the interviewee had spent the weekend and the lie detection part was a secondary task. Rather than trying to spot specific tells, the interviewer was encouraged to pay attention to the overall behavior of the interviewee. No further information or training about verbal or nonverbal methods of lie detection was given.

To support the interviewer in accomplishing the primary task he/she was instructed to construct a timeline based on the interviewee’s activities. He/she was also instructed to use open-ended questions as the basic information gathering method and actively listen what the interviewee was telling. Otherwise the interviewer was free to lead the conversation as he/she wanted.

This kind of interview was considered to be a low stake situation. Since the interviewee was able to choose the time period, he/she could avoid giving too personal or otherwise sensitive information. It was assessed that this kind of approach would lead the interviewee to be as truthful as possible in his/her story, apart from completing the secondary task. The instructor encouraged both parties to do their best, although it was also emphasized that this exercise was conducted only for warming up purposes and they were not expected to be either master liars or master lie detectors.

No material rewards were promised to either party for a successful completion of their tasks. The only reward would be a public acknowledgement of success at the end of the exercise. The main motivation for both parties was assessed to be a successful completion of given tasks which also included beating the fellow student in lie detection. In this kind of a professional skills improvement setup the above mentioned reward was assessed to be high enough to make participants to respect the instructions, try their best and in this way make a personal, and professional, investment in their success.

To be able to accomplish their tasks, the interviewers were instructed to use two specific tools during the interview. First, they were told to use open-ended questions and in this way encourage the interviewees to free telling, and second, they were advised to sketch a timeline regarding different events, their durations and people involved in them.
After the interview was finished, the interviewer was given five minutes to finalize the timeline, to try to find the missing part of the story and to pinpoint what the interviewee lied about. This was immediately followed by a feedback session during which both parties shared their first experiences about the exercise. The interviewer also presented his/her findings, and after that, the interviewee revealed where the missing part was and what the lie was about. The resulting discussion and reflections about the exercise was allowed to continue for another five minutes.

In addition to the video recording, the documentation of the related data was finished. The documentation included the indexing of the video clips, names of the participants, their roles as an interviewer and interviewee, detail about the interviewees’ secondary tasks, detail about the interviewers’ success in their secondary tasks and preliminary notes taken by the head of the exercise. These videos and documents were used afterwards to test selected veracity assessment methods in order to gain additional firsthand information about their applicability.

6.2 Analysis of the Video Recordings

The aim of this phase was to establish deeper understanding about the applicability of the studied veracity assessment methods. The initial plan was to go through all of the nine interviews and apply selected methods on them. During the analysis it was soon noticed that the findings and observations started to repeat themselves. After the analysis of four interviews out of nine this phase was ended because additional information, concerning the applicability of the selected methods, was not cumulating anymore.

The Studied Features of Discourse, the Studied Features of Nonverbal Communication and the Plausibility and Inner Logic of the Statement were included in the analysis. The Plausibility and Inner Logic of the Statement was used during the interview by the interviewers as part of their secondary task. This task was not replicated by the author and the understanding of this method’s applicability was built from the feedback of the interviewing exercise. The Criteria Based Content Analysis (CBCA) was not applied, but after the application of the other methods, the material was briefly viewed through four criteria that were assessed to have a high diagnostic value. The Method for Assessing Credibility of Evidence (MACE) and the Polygraph were not included in this phase because there was not access either to other

---

240 Unstructured Production, Quantity of Details, Contextual Embedding and Reproduction of Conversations.
sources, that could complement the information extracted from the interview, or to the polygraph test. The Discourse Analysis was also excluded due the restrictions in time and availability of trained personnel\(^{241}\).

The analysis of the recordings was started with the Studied Features of Discourse. A spreadsheet with timeline was used to store the findings and to connect the existence of markers with the content. At this stage the content was not transcribed in detail. In order to keep the assessment consistent every piece of recording, it was decided to first analyze the recordings with one method and after that apply another method with the similar procedure. When the Studied Features of Discourse were applied, only the audio track was used to avoid distractions from visual information. Respectively, when the Studied Features of Nonverbal Communication were applied, only the video track was used. Also in the spreadsheet, these two channels were isolated to avoid any interference with previously made notes.

After both methods were applied on the recordings, the notes were merged as one. The rough content of the statement, or the topic of the conversation, was checked and connected with the existing Discourse and Nonverbal Behavior markers. Although the aim of this phase was not to test the lie detection accuracy with scientifically sound procedures, it was interesting find out if markers would occur in parallel with each other and also in parallel with the deception.

### 6.3 Results from the Field Study

Based on the findings from the field study the questioning technique has got a significant influence on the flow of the interview and the information collection. Although the interviewees were encouraged to use Information-gathering questioning, with open ended questions and active listening, the questioning technique naturally tended to shift towards direct questioning. It was quite common that the interviewer did not let the interviewee answer freely or interrupted the interviewee’s silent thinking with a follow on question. This highlights the fact that the use of advanced questioning techniques indeed requires training, and people prefer to use techniques that they feel the most comfortable with.

Results from the analysis of the recordings are presented in two parts. First the findings from the use of the Studied Features of Discourse\(^ {242}\) regarding all of the analyzed recordings are

---

\(^{241}\) Edwards 2001, 336. It is estimated that a word-level transcription takes about ten times the length of the original recording.
summarized. After that, the findings from the use of the Studied Features of Nonverbal Communication\textsuperscript{243} are summarized as well.

Although the audio recordings of the interviews were not transcribed word by word, it took quite a long time to listen to the whole interview and to code the used assessment criteria. In average it took twice the time of the original recording to code the Studied Features of Discourse. The time for coding varied slightly based on the interviewee’s speech rate and the existence of criteria. A speaker with a relatively slow speech rate and few variations in speech was easy to code online and required only little rewinding. However, even the quickest coding session was fifty percent longer that the original recording was.

*Increased pauses* and *increased latency periods* were relatively easy to notice. A baseline of the interviewee’s speech was quickly established and deviations from that were noted mainly based on intuition. These intuitive perceptions were verified by measuring the duration of pauses and latency periods. Abnormal pauses, about 3 seconds and longer, in otherwise fluent conversation were easily noted and coded. *Errors* and “*ah*” *speech disturbances* tend to stand out clearly as they occur, especially when the statement is mainly fluent. On the other hand, if different speech disturbances are common, the detection of the *increase* requires meticulous coding and quantitative analysis. Changes in speech rate are audible to some extent but the threshold for *slower speech* rate is hard to establish solely based on intuition. Reliable detection of slower speech rate requires the counting of delivered words per minute, or other selected time frame, as has been done in the various lie detection studies.

The effect of interviewer’s activity was larger than expected. In one occasion the interviewer was noticeable passive in his interaction. In general he let the interviewee to answer freely but he did not encourage the interviewee to give more detailed answers or to elaborate the subject more. This enabled the interviewee to answers very briefly, and in general level, which reduced the amount of speech, i.e. the chances for the criteria to occur. The amount collected information was reduced, as well. Eventually the interviewer started to use closed questions, which seemed to suit him better. In another session the interviewer did not give the interviewee time to think freely, or to complete his answer, and interrupted him with another question

\textsuperscript{242} Increased pauses, increased latency periods, increased “ah” speech disturbances and slower speech rate.

\textsuperscript{243} Reduced hand/finger movement, increased amount of blinking and decreased amount of illustrating hand gestures.
or complemented the initial one. These kinds of interruptions made it impossible to detect or to measure potentially increased latency periods and pauses.

The coding of the Studied Features of Nonverbal Communication was equally time consuming. It took about twice the length of the recording to code one feature. The number of the coded features multiplies the required time respectively. It was found out that by concentrating on one feature at a time it was easier to detect its existence. During the application of the Studied Features of Nonverbal Communication the intuition seems to play a less significant role than during the detection of the Studied Features of Discourse. One explanation for this might be that there are perhaps less strict conventions connected to the nonverbal communication than to the verbal communication and the flow of a dialog. Thus, it could be argued that the detection of anomalies in nonverbal communication is more based on reasoning, and the quantitative measurement, than intuition.

The initial impressions from the coding of the Studied Features of Nonverbal Communication were twofold. The overall level of interviewee’s animation and the changes of body position were easily detected. In general, the interviewees were very calm, almost rigid, except one who was perhaps overly lively. Clear changes in interviewees’ posture and gesturing could be detected with little effort. On the other hand, if the changes are subtle, the validity of observations becomes quickly questioned. How the changes are detected accurately and how the threshold for a noteworthy deviation is established to ensure the required reliability?

It was found out that the decrease in illustrating gestures was hard to detect when the overall animation of the person was very low. Instead the sudden increase in illustrating gestures was very clearly visible. On the other hand, when the interviewee’s overall behavior was lively, it was easy to detect the sudden decrease. However, the previously mentioned challenge was faced again. What kind of pause or decrease in illustrating gestures should be considered abnormal? It could be argued that there is no absolute value for this kind of assessment, and the threshold for the anomaly should be based on individually created baseline behavior, which makes the analysis even more complicated.

The detection of changes in the hand/finger movement, let alone the detection of the decrease in the hand/finger movement were found to be difficult. It was hard to distinguish between the arm movement, which was related to illustrating gestures, and the actual hand/finger movement. The problem was same with both the controlled and lively interviewees. Again, a relia-
ble detection of reduced hand/finger movement would require very careful coding and the establishment of the individual baseline for the basis of quantitative analysis.

An *increase in blinking rate* was very difficult, or almost impossible, to detect. The video quality of the recording was not the hindering factor. None of the interviewees exhibited the kind of observable behavior that stood out as abnormal. Blinking rate, which was detectable with bare eyes, stayed or at least seemed to stay unchanged during the interviews. Intuition did not seem to offer any aid. These findings highlight the need for computer assistance, if reliable assessment related to the blinking rate is desired.

During the analysis of the Studied Features of Discourse some attention was paid to the actual content of the statement from the CBCA’s point of view. It became immediately evident that, as it was stated in the Chapter 4.4, a reliable application of the CBCA requires a full transcription of the statement. Even when the content was only viewed through the selected criteria, Quantity of Details, Unstructured Production, Contextual Embedding and Reproduction of Conversations, the assessment was more based on an educated guess rather than a scientific assessment. However, Reproduction of Conversation is possibly the easiest to detect because it is usual delivered in the form of a quote. Also Unstructured Production is quite easily detected. Intuition seems to work well, as internally inconsistent or bouncing narration is easily noted.

In addition, preferred topics could be found in all of the interviews. When the interviewees were comfortable with the topic, they were talkative and voluntarily gave detailed information. On the other hand, when the topic was not a preferred one, or perhaps there was not much information available for a reason or another, the interviewer had to make additional questions to encourage the interviewee to continue. The attempt to control one’s behavior is quite easily noted, especially when it significantly differs from their normal behavioral pattern. Of course there are many reasons why someone would like to be very minimal with his/her gesturing, and the attempt to hide possible signs of deception is only one of them.

To sum up, the Studied Features of Discourse were easier to apply than the Studied Features of Nonverbal Communication. Somehow the detection of increased latency periods, increased pauses and speech errors felt intuitive and changes were easily recognized. Clear changes in the level of overall animation were also relatively easy to detect. In contrast, reduced speech rate, increased blinking rate and reduced hand/finger movement were significantly more diffi-
cult to detect with high reliability. Although some changes could be occasionally observed, a reliable veracity assessment using the less intuitive criteria requires careful coding and qualitative analysis. These findings support the idea that, instead of detecting some specific patterns of behavior, the detection of deviation from the baseline could be more beneficial. These findings also support the applicability analysis conducted in Chapter 5.3 in terms of Ease of Use and to some extent Time Required.

Slightly outside the scope of this study, a few facts regarding the lie detection accuracy during the field study are worth mentioning. Five out of nine interviewers managed to find the missing part of the statement based on the use of timeline and logical reasoning. Three were suspicious about some parts of the statement but felt that they did not have enough information for a reliable decision. One interviewer reported being unable to detect the missing part. Two out of nine reported that they successfully had detected the lie. Both successful interviewers explained that they had based their assessment on logical deduction and the lack of consistency and plausibility. Low lie detection accuracy was expected, as this kind of task is notoriously difficult even for the highly experienced professionals. However, based on the fact that the interviewers had not received any training for lie detection or veracity assessment, the outcome of the first task, to find a half an hour gap in the timeline, was surprisingly good and exceeded the expectations. A success rate of closer to zero percent was expected.

Another interesting finding was made during the analysis of the video recordings. If the observed pauses, latency periods, speech errors and nonverbal indicators would have been seen as guidelines for the potential topics of future interest, the collection efforts would have been guided to correct direction in three interviews out of four. In addition, the used method would have produced twice as much false alarms as it offered potentially usable guidelines for future collection. It must be emphasized that these findings have only anecdotal value and this kind of analysis, concerning the accuracy of a lie detection method, does not meet any criteria of a reliable scientific study. However, in future studies this kind of approach should be examined with greater effort.

There are many factors that can hinder the performance of lie detection and veracity assessment. Interviewers’ biases, lack of scientific knowledge related to human behavior, limitations in noticing simultaneous occurring indicators and limitations in remembering all the relevant collected information all have a negative effect on the reliability of the assessment.

---

244 Hart, Fillmore and Griffith 2009, 137.
interviewer, overt or covert, can help to overcome these obstacles. Audio-video recording is a convenient way to store information for future re-examination and enables the investigator or the analyst to gain back some advantage from the sender who is well prepared\textsuperscript{245}.

\textsuperscript{245} Furner and George 2012, 1429-1430.
7 DISCUSSION

7.1 General Discussion

Normally people base their opinions on their own observations, experiences or on the opinions of trusted people. Establishing a scientifically sound view of a phenomenon requires extra effort to avoid presumptions, biases and false believes of everyday thinking. This same rule seems to apply in lie detection and veracity assessment. A vast majority have adopted their views about lie detection from self-observed behavior, mass media, social networks or personal behavior\textsuperscript{246}. Not surprisingly, most of the adopted cues to deception are the ones that do not have scientific support for being reliable signs of deception\textsuperscript{247}. The need for continuous research that cumulates scientific understanding in the field of lie detection and veracity assessment could not have been highlighted better.

The aim of this study was not to find the most accurate lie detection or veracity assessment method or to develop one. The goal was to describe and analyze existing methods so that their applicability in the intelligence interview context could be assessed. In order to achieve this goal Interviewing and Interrogation Techniques were first analyzed to emphasize their potential effects on the intelligence collection and the following veracity assessment. It was discovered that some Interviewing and Interrogation Techniques have embedded lie detection or veracity assessment methods that actually work against the collection task.

In the course of the study it became clear that some of the lie detection and veracity assessment methods had potential pitfalls in their scientific proof. All of the methods had a scientifically defensible theory to support them but in the most argued ones the connection between the theory and the application was weak or otherwise controversial. Evident deficits in the main criterion \textit{Accuracy} led to the exclusion of the method from the applicability analysis. Other used criteria were \textit{Ease of Use, Time Required, No Need for Special Equipment} and \textit{Unobtrusiveness}.

\textsuperscript{246} Hurley, Griffin and Stefanone 2014, 25.  
\textsuperscript{247} Hurley, Griffin and Stefanone 2014, 28.
The Plausibility and Inner Logic of the Statement, the Criteria Based Content Analysis (CBCA), the Discourse Analysis, the Studied Features of Discourse\textsuperscript{248}, the Studied Features of Nonverbal Communication\textsuperscript{249}, the Method for Assessing Credibility of Evidence (MACE) and the Polygraph\textsuperscript{250} were included in the applicability analysis. Traditional signs of deception and sender credibility assessment based on demeanor were excluded because they had evident weaknesses in their scientific foundations. The Polygraph was chosen to represent all the valid Advanced Technical Aids although the Polygraph’s results only offer implications of the applicability of other technical methods.

The analysis revealed that within the context of this research the Studied Features of Discourse and the Studied Features of Nonverbal Communication share first place as the most applicable veracity assessment methods. The Plausibility and Inner Logic of the Statement, MACE and the CBCA are equally applicable with some limitations. The least applicable methods were the Discourse Analysis and the Polygraph.

These findings do not mean that the most applicable methods are the best in every situation and for all purposes. The results do not either suggest that the least applicable methods are useless. All of the above mentioned seven veracity assessment methods have a scientific foundation and some of them have an extensive amount of scientific studies to support that.

Based on this study the foundation for success is laid in the very beginning of the process. Intelligence is gathered during the questioning. Manipulation, accusations and possible threats against the source’s health do not work towards the goal of acquiring reliable and accurate intelligence. Often the answer can only be as relevant as the question is. To prevent the information from corrupting during the collection, questioning techniques that are not leading or suggestive, or produce longer and narrative answers should be preferred. Ultimately they should support accurate memory retrieval and challenge inconsistencies in a constructive manner. From this point of view, information-gathering type questioning techniques should be used.

\textsuperscript{248} Increased pauses, increased latency periods, increased “ah” speech disturbances and slower speech rate.
\textsuperscript{249} Reduced hand/finger movement, increased amount of blinking and decreased amount of illustrating hand gestures.
\textsuperscript{250} The polygraph test is considered to reflect the traits of other physiological lie detection methods that are based on the Arousal Theory.
It was discovered that three channels of information can be subjected to veracity assessment: Content, Discourse and Nonverbal Communication. If the assessment is conducted carefully, these channels can be isolated and analyzed separately. Results from three separate channels can be merged during the follow on analysis. This kind of separated analysis is essential in order to avoid interference or biases caused by information from other channels. Based on this study there is at least one reliable and applicable veracity assessment method for each of the three channels. Content can be subjected to three, discourse to one and nonverbal behavior to one method that gained high or moderate applicability rating in this study.

![Diagram](image)

**Picture 10.** Different channels of information, related theories and veracity assessment methods.

All of the methods require disciplined application and a scientific working approach. There are no quick gains if high accuracy and reliability is desired. The Studied Features of Discourse, the Studied Features of Nonverbal Communication and even the CBCA can be monitored online by the interviewer or supporting personnel. However, a thorough content analysis, discourse analysis and the gathering of supporting all-source intelligence is always time consuming, labor intensive and requires the use of supporting personnel. In all, veracity assessment is not one man’s show where the hero cracks the case based on a single tell tale sign.

Different means of enhancing human’s abilities to detect and store different parameters of human behavior are developing rapidly. High speed video cameras, multispectral imaging and
speech recognition combined with improved signal processing offer indisputable advantages over traditional sensory observations. Not only can they reveal entirely new information but they can also lighten the workload of the analyst and make the analysis quicker and preferably more reliable. However, no matter what the future applications will be, their scientific foundation must be rock solid and based on a confirmed understanding of human behavior.

7.2 Reliability and Validity

In this study the understanding of current lie detection and veracity assessment methods was established through the meta-analysis of existing scientific studies. Although the focus was set to the latest research, the collection of material was not limited to a certain time frame. The emphasis of the collection was on scientific articles, studies and other papers that had been exposed to the scrutiny of the scientific community. After the collection the studied methods were described and analyzed in the light of the assessment criteria to form a starting point for the applicability analysis. Although the analysis of the pros and cons of the different methods was based on the author’s analysis and interpretations, the logic behind the chain of deduction was explained, documented and thus has become traceable.

In this study the applicability analysis was conducted by a single person instead of using a group of experts. Current methods were chosen to overcome time and personnel related limitations. The use of a single person exposes results to the rater’s potential biases and systematic errors. To overcome possible reliability and validity issues stemming from the use of a single rater the Analytic Hierarchy Process was used for the analysis. Another great benefit of the AHP was the documentation that it produced. This documentation was stored to annexes to ensure that the analysis could be reviewed in detail afterwards and repeated, if needed. This way the applicability analysis became transparent which improved the reliability and validity of the results to some extent. In addition to that, the results from the Multi Criteria Analysis were iterated through a sensitivity analysis to reveal possible issues related to rating scale.

Since lie detection and veracity assessment are part of an ever developing field of behavioral science, new methods are being developed and traditional ones are improved continuously. The latest computer assisted methods were not included in this study because there was not as much studied information available from them as there was from the more traditional ones. The inclusion of the latest computer assisted lie detection methods would not have changed
the relative order of the currently studied more traditional methods, but their inclusion would have offered deeper insight for the future possibilities of lie detection.

7.3 Implications for Future Studies

In future studies one aspect of the lie detection and veracity assessment should be addressed with more emphasis. Is the overall veracity assessment of the source, or the content that he/she delivered, the correct direction to head for? In the intelligence interview, or some other human source intelligence gathering context it is not relevant to study whether the source is totally honest or not. Partial dishonesty should be the starting point for the veracity assessment. It should be expected that the source has received training to resist interrogation or is otherwise smart enough to withhold incriminating pieces of information. It should be expected that the source is smart enough not to fabricate an utterly incredible cover story that would immediately give him/her away. He/she is also expected not to lie about matters that can be easily verified from other sources of information. This leads to the hypothesis that most parts of well-prepared lies are true, and the actual task is to find where the missing parts are, what parts are fabricated and what have been slightly manipulated.

In human intelligence the success of veracity assessment is not measured with a dichotomous choice between a truthful or deceitful source. Success is measured with the correct assessment on the source’s veracity regarding the specific areas of intelligence that the source has provided. Thus the assessment is always in close connection to the content. In this context it is essential to assess which parts of the statement can be considered reliable and which parts of the statement are doubtful. The adopted veracity assessment methods must produce insight into whether some of the following scenarios take place: the source tries to control the delivery of content or personal behavior, leaves facts untold, uses avoidance strategies, embeds unrelated true facts, ads fabricated pieces of information, speculates, remembers facts incorrectly, tells what he/she really knows or presents a well prepared cover story with no connection to true events. In the light of these scenarios, the results of this study are promising. Analyzed veracity assessment methods offer tools to address the requirements either individually or in a combination of multiple methods.

Since most of the lie detection studies are concentrated around the last scenario, where roughly half of the assessed people are liars and present a well prepared cover story and the other half is totally truthful, it is proposed that lie detection and veracity assessment methods are
tested against partially truthful human sources. It is suggested that this kind of test setup would highlight new challenges and opportunities for the use of existing, more widely studied methods, as well as for the modern ones that are still under development.
REFERENCES


Blandón-Gitlin, Iris, Pezdek, Kathy, Lindsay, D. Stephen and Hagen, Lisa. CRITERIA-BASED CONTENT ANALYSIS OF TRUE AND SUGGESTED ACCOUNTS OF EVENTS. *Applied Cognitive Psychology*, no. 23 (2009): 901-907. doi: 10.1002/acp.1504


Lauerma, Hannu; Chief Physician of the Psychiatric Hospital for Prisoners, Research Professor in the National Institute for Health and Welfare. Interview 2015-06-22.


Senate Select Committee on Intelligence. Committee Study of the Central Intelligence Agency’s Detention and Interrogation Program, Findings and Conclusions. United States Senate, United States of America 2014.


LIST OF APPENDIXES

APPENDIX 1: AHP Comparison of the Interviewing and Interrogation Techniques = 5 pages

APPENDIX 2: Analysis of the Method for Assessing the Credibility of Evidence (MACE) = 5 pages

APPENDIX 3: The List of the Content Based Credibility Assessment (CBCA) criteria = 1 page

APPENDIX 4: AHP Comparison of the Applicability Criteria = 1 page

APPENDIX 5: AHP Comparison of the Veracity Assessment Methods = 8 pages
### AHP Analytic Hierarchy Process (EVM multiple inputs)


**Only input data in the light green fields and worksheets!**

- **n=** Number of criteria (3 to 10)  
  - **Scale:** 5  
  - **Linear**
- **N=** Number of Participants (1 to 20)  
  - **a:** 0,1  
  - **Consensus:** 89.5%
- **p=** selected Participant (0=consol.)  
  - **2**  
  - **7**  
  - **Consolidated**

**Objective**: To compare different questioning techniques with each other in terms of their good information collection values: Little leading or suggestion, longer and narrative

**Author**  
- **Date**: March 2015  
  - **Thresh:** 1E-07  
  - **Iterations:** 3  
  - **EVM check:** 4.4E-08

### Table 1. Total rating of the interviewing and interrogation techniques.

<table>
<thead>
<tr>
<th>Table</th>
<th>Criterion</th>
<th>Comment</th>
<th>Weights</th>
<th>Rk</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Direct Q &amp; Repeated Qns</td>
<td></td>
<td>15,8%</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>Accusatorial Qn (Reid)</td>
<td></td>
<td>2,8%</td>
<td>7</td>
</tr>
<tr>
<td>3</td>
<td>Kinesic I &amp; I</td>
<td></td>
<td>3,2%</td>
<td>6</td>
</tr>
<tr>
<td>4</td>
<td>Elicitation</td>
<td></td>
<td>21,6%</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>Inform. Gathering Qn</td>
<td></td>
<td>20,6%</td>
<td>3</td>
</tr>
<tr>
<td>6</td>
<td>Cognitive I</td>
<td></td>
<td>22,2%</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>Strategic U of E</td>
<td></td>
<td>11,1%</td>
<td>5</td>
</tr>
<tr>
<td>8</td>
<td>Polygraph CQT &amp; CIT</td>
<td></td>
<td>2,7%</td>
<td>8</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td>0,0%</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td>0,0%</td>
<td></td>
</tr>
</tbody>
</table>

**Result**  
- **Eigenvalue**  
  - **lambda:** 8,048
- **Consistency Ratio**  
  - **CR:** 0.37
  - **GCI:** 0.02
  - **OR:** 0.5%

**Matrix**

<table>
<thead>
<tr>
<th></th>
<th>Direct Q &amp; Repeated Qns</th>
<th>Accusatorial Qn (Reid)</th>
<th>Kinesic I &amp; I</th>
<th>Elicitation</th>
<th>Inform. Gathering</th>
<th>Cognitive I</th>
<th>Strategic U of E</th>
<th>Polygraph CQT &amp; CIT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>8</td>
<td>4/3</td>
<td>3/5</td>
<td>3/4</td>
<td>1/2</td>
<td>1/2</td>
<td>1/2</td>
</tr>
<tr>
<td>2</td>
<td>1/8</td>
<td>1</td>
<td>1/1</td>
<td>1/1</td>
<td>1/1</td>
<td>1/1</td>
<td>1/1</td>
<td>1/1</td>
</tr>
<tr>
<td>3</td>
<td>1/5</td>
<td>1/1/9</td>
<td>1/6</td>
<td>1/1/7</td>
<td>1/1/7</td>
<td>1/1</td>
<td>1/1</td>
<td>1/1</td>
</tr>
<tr>
<td>4</td>
<td>1/2/3</td>
<td>7/1/3</td>
<td>5/5/6</td>
<td>1/1/1</td>
<td>1/1/1</td>
<td>1/1</td>
<td>1/1</td>
<td>1/1</td>
</tr>
<tr>
<td>5</td>
<td>1/3/8</td>
<td>7/1/3</td>
<td>6/5/9</td>
<td>1/1/1</td>
<td>1/1/1</td>
<td>1/1</td>
<td>1/1</td>
<td>1/1</td>
</tr>
<tr>
<td>6</td>
<td>1/7/9</td>
<td>7/1/7</td>
<td>6/5/9</td>
<td>1/1/1</td>
<td>1/1/1</td>
<td>1/1</td>
<td>1/1</td>
<td>1/1</td>
</tr>
<tr>
<td>7</td>
<td>4/5</td>
<td>3/3/4</td>
<td>3/1/2</td>
<td>1/2/1</td>
<td>1/2/1</td>
<td>1/1/1</td>
<td>1/1/1</td>
<td>1/1/1</td>
</tr>
<tr>
<td>8</td>
<td>1/7</td>
<td>1/1</td>
<td>1/8/1/1</td>
<td>1/1/1/1</td>
<td>1/1/1</td>
<td>1/1/1</td>
<td>1/1/1</td>
<td>1/1/1</td>
</tr>
<tr>
<td>9</td>
<td>1/7</td>
<td>1/1</td>
<td>1/8/1/1</td>
<td>1/1/1/1</td>
<td>1/1/1</td>
<td>1/1/1</td>
<td>1/1/1</td>
<td>1/1/1</td>
</tr>
<tr>
<td>10</td>
<td>1/7</td>
<td>1/1</td>
<td>1/8/1/1</td>
<td>1/1/1/1</td>
<td>1/1/1</td>
<td>1/1/1</td>
<td>1/1/1</td>
<td>1/1/1</td>
</tr>
</tbody>
</table>

**Normalized principal Eigenvector**

- **Direct Q & Repeated Qns**: 15.76%
- **Accusatorial Qn (Reid)**: 2.75%
- **Kinesic I & I**: 3.18%
- **Elicitation**: 21.63%
- **Inform. Gathering**: 20.62%
- **Cognitive I**: 22.21%
- **Strategic U of E**: 11.10%
- **Polygraph CQT & CIT**: 2.74%

Table 1. Total rating of the interviewing and interrogation techniques.
### Objective
To compare different questioning techniques with each other in terms of their good information collection values: Little leading or suggestion, support for accurate memory retrieval and constructive challenging of inconsistencies.

Only input data in the light green fields!

Please compare the importance of the elements in relation to the objective and fill in the table: Which element of each pair is more important, A or B, and how much more on a scale 1-9 as given below.

Once completed, you might adjust highlighted comparisons 1 to 3 to improve consistency.

<table>
<thead>
<tr>
<th>n</th>
<th>Criteria</th>
<th>Comment</th>
<th>RGMM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Direct Q &amp; Repeated Qns</td>
<td></td>
<td>16%</td>
</tr>
<tr>
<td>2</td>
<td>Accusatorial Qn (Reid)</td>
<td></td>
<td>2%</td>
</tr>
<tr>
<td>3</td>
<td>Kinesic I &amp; I</td>
<td></td>
<td>4%</td>
</tr>
<tr>
<td>4</td>
<td>Elicitation</td>
<td></td>
<td>26%</td>
</tr>
<tr>
<td>5</td>
<td>Inform. Gathering Qn</td>
<td></td>
<td>22%</td>
</tr>
<tr>
<td>6</td>
<td>Cognitive I</td>
<td></td>
<td>17%</td>
</tr>
<tr>
<td>7</td>
<td>Strategic U of E</td>
<td></td>
<td>9%</td>
</tr>
<tr>
<td>8</td>
<td>Polygraph CQT &amp; CIT</td>
<td></td>
<td>4%</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td>for 9&amp;10 unprotect the input sheets and expand the question section (“+” in row 66)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Name</th>
<th>Weight</th>
<th>Date</th>
<th>Consistency Ratio</th>
<th>Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>B</td>
<td>A or B</td>
<td>(1-9)</td>
<td>CR</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>Direct Q &amp; Repeated Qns</td>
<td>Accusatorial Qn (Reid)</td>
<td>A</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>Accusatorial Qn (Reid)</td>
<td>Kinesic I &amp; I</td>
<td>A</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>Kinesic I &amp; I</td>
<td>Elicitation</td>
<td>B</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>Elicitation</td>
<td>Inform. Gathering Qn</td>
<td>B</td>
</tr>
<tr>
<td>1</td>
<td>6</td>
<td>Inform. Gathering Qn</td>
<td>Cognitive I</td>
<td>B</td>
</tr>
<tr>
<td>1</td>
<td>7</td>
<td>Cognitive I</td>
<td>Strategic U of E</td>
<td>A</td>
</tr>
<tr>
<td>1</td>
<td>8</td>
<td>Strategic U of E</td>
<td>Polygraph CQT &amp; CIT</td>
<td>A</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>Accusatorial Qn (Reid)</td>
<td>Kinesic I &amp; I</td>
<td>B</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>Kinesic I &amp; I</td>
<td>Elicitation</td>
<td>B</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>Elicitation</td>
<td>Inform. Gathering Qn</td>
<td>B</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>Inform. Gathering Qn</td>
<td>Cognitive I</td>
<td>B</td>
</tr>
<tr>
<td>2</td>
<td>7</td>
<td>Cognitive I</td>
<td>Strategic U of E</td>
<td>B</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>Strategic U of E</td>
<td>Polygraph CQT &amp; CIT</td>
<td>B</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>Kinesic I &amp; I</td>
<td>Elicitation</td>
<td>B</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>Elicitation</td>
<td>Inform. Gathering Qn</td>
<td>B</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
<td>Inform. Gathering Qn</td>
<td>Cognitive I</td>
<td>B</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>Cognitive I</td>
<td>Strategic U of E</td>
<td>B</td>
</tr>
<tr>
<td>3</td>
<td>8</td>
<td>Strategic U of E</td>
<td>Polygraph CQT &amp; CIT</td>
<td>B</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>Elicitation</td>
<td>Inform. Gathering Qn</td>
<td>A</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>Inform. Gathering Qn</td>
<td>Cognitive I</td>
<td>A</td>
</tr>
<tr>
<td>4</td>
<td>7</td>
<td>Cognitive I</td>
<td>Strategic U of E</td>
<td>A</td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>Strategic U of E</td>
<td>Polygraph CQT &amp; CIT</td>
<td>A</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>Inform. Gathering Qn</td>
<td>Cognitive I</td>
<td>A</td>
</tr>
<tr>
<td>5</td>
<td>7</td>
<td>Cognitive I</td>
<td>Strategic U of E</td>
<td>A</td>
</tr>
<tr>
<td>5</td>
<td>8</td>
<td>Strategic U of E</td>
<td>Polygraph CQT &amp; CIT</td>
<td>A</td>
</tr>
<tr>
<td>6</td>
<td>7</td>
<td>Strategic U of E</td>
<td>Polygraph CQT &amp; CIT</td>
<td>A</td>
</tr>
<tr>
<td>6</td>
<td>8</td>
<td>Polygraph CQT &amp; CIT</td>
<td>Polygraph CQT &amp; CIT</td>
<td>A</td>
</tr>
<tr>
<td>7</td>
<td>8</td>
<td>Polygraph CQT &amp; CIT</td>
<td>Polygraph CQT &amp; CIT</td>
<td>A</td>
</tr>
</tbody>
</table>

Table 2. Comparison table of criterion Little leading and suggestion.
**AHP Analytic Hierarchy Process**

**Objective:** To compare different questioning techniques with each other in terms of their good information collection values: Longer and narrative answers, support for accurate memory retrieval and constructive challenging of inconsistencies.

*Only input data in the light green fields!*

Please compare the importance of the elements in relation to the objective and fill in the table: Which element of each pair is more important, A or B, and how much more on a scale 1-9 as given below.

Once completed, you might adjust highlighted comparisons 1 to 3 to improve consistency.

<table>
<thead>
<tr>
<th>n</th>
<th>Criteria</th>
<th>Comment</th>
<th>RSMM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Direct Q &amp; Repeated Qns</td>
<td></td>
<td>13 %</td>
</tr>
<tr>
<td>2</td>
<td>Accusatorial Qn (Reid)</td>
<td></td>
<td>3 %</td>
</tr>
<tr>
<td>3</td>
<td>Kinesic I &amp; I</td>
<td></td>
<td>3 %</td>
</tr>
<tr>
<td>4</td>
<td>Elicitation</td>
<td></td>
<td>26 %</td>
</tr>
<tr>
<td>5</td>
<td>Inform. Gathering Qn</td>
<td></td>
<td>23 %</td>
</tr>
<tr>
<td>6</td>
<td>Cognitive I</td>
<td></td>
<td>23 %</td>
</tr>
<tr>
<td>7</td>
<td>Strategic U of E</td>
<td></td>
<td>7 %</td>
</tr>
<tr>
<td>8</td>
<td>Polygraph CQT &amp; CIT</td>
<td></td>
<td>2 %</td>
</tr>
</tbody>
</table>

For 9 & 10 unprotected the input sheets and expand the question section (+" in row 66).

<table>
<thead>
<tr>
<th>Criteria</th>
<th>more important ?</th>
<th>Scale (1-9)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct Q &amp; Repeated Qns</td>
<td>A 9</td>
<td></td>
</tr>
<tr>
<td>Accusatorial Qn (Reid)</td>
<td>A 7</td>
<td></td>
</tr>
<tr>
<td>Kinesic I &amp; I</td>
<td>B 6</td>
<td></td>
</tr>
<tr>
<td>Elicitation</td>
<td>B 9</td>
<td></td>
</tr>
<tr>
<td>Inform. Gathering Qn</td>
<td>B 9</td>
<td></td>
</tr>
<tr>
<td>Strategic U of E</td>
<td>A 5</td>
<td></td>
</tr>
<tr>
<td>Polygraph CQT &amp; CIT</td>
<td>A 9</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Criteria</th>
<th>more important ?</th>
<th>Scale (1-9)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct Q &amp; Repeated Qns</td>
<td>A 9</td>
<td></td>
</tr>
<tr>
<td>Accusatorial Qn (Reid)</td>
<td>A 7</td>
<td></td>
</tr>
<tr>
<td>Kinesic I &amp; I</td>
<td>B 6</td>
<td></td>
</tr>
<tr>
<td>Elicitation</td>
<td>B 9</td>
<td></td>
</tr>
<tr>
<td>Inform. Gathering Qn</td>
<td>B 9</td>
<td></td>
</tr>
<tr>
<td>Strategic U of E</td>
<td>A 5</td>
<td></td>
</tr>
<tr>
<td>Polygraph CQT &amp; CIT</td>
<td>A 9</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Criteria</th>
<th>more important ?</th>
<th>Scale (1-9)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct Q &amp; Repeated Qns</td>
<td>A 9</td>
<td></td>
</tr>
<tr>
<td>Accusatorial Qn (Reid)</td>
<td>A 7</td>
<td></td>
</tr>
<tr>
<td>Kinesic I &amp; I</td>
<td>B 6</td>
<td></td>
</tr>
<tr>
<td>Elicitation</td>
<td>B 9</td>
<td></td>
</tr>
<tr>
<td>Inform. Gathering Qn</td>
<td>B 9</td>
<td></td>
</tr>
<tr>
<td>Strategic U of E</td>
<td>A 5</td>
<td></td>
</tr>
<tr>
<td>Polygraph CQT &amp; CIT</td>
<td>A 9</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Criteria</th>
<th>more important ?</th>
<th>Scale (1-9)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct Q &amp; Repeated Qns</td>
<td>A 9</td>
<td></td>
</tr>
<tr>
<td>Accusatorial Qn (Reid)</td>
<td>A 7</td>
<td></td>
</tr>
<tr>
<td>Kinesic I &amp; I</td>
<td>B 6</td>
<td></td>
</tr>
<tr>
<td>Elicitation</td>
<td>B 9</td>
<td></td>
</tr>
<tr>
<td>Inform. Gathering Qn</td>
<td>B 9</td>
<td></td>
</tr>
<tr>
<td>Strategic U of E</td>
<td>A 5</td>
<td></td>
</tr>
<tr>
<td>Polygraph CQT &amp; CIT</td>
<td>A 9</td>
<td></td>
</tr>
</tbody>
</table>

Table 3. Comparison table of criterion *Longer and narrative answers.*
Table 4. Comparison table of criterion Support for accurate memory retrieval.

<table>
<thead>
<tr>
<th>n</th>
<th>Criteria</th>
<th>Comment</th>
<th>RGMM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Direct Q &amp; Repeated Qns</td>
<td></td>
<td>15 %</td>
</tr>
<tr>
<td>2</td>
<td>Accusatorial Qn (Reid)</td>
<td></td>
<td>3 %</td>
</tr>
<tr>
<td>3</td>
<td>Kinesic I &amp; I</td>
<td></td>
<td>3 %</td>
</tr>
<tr>
<td>4</td>
<td>Elicitation</td>
<td></td>
<td>14 %</td>
</tr>
<tr>
<td>5</td>
<td>Inform. Gathering Qn</td>
<td></td>
<td>15 %</td>
</tr>
<tr>
<td>6</td>
<td>Cognitive I</td>
<td></td>
<td>21 %</td>
</tr>
<tr>
<td>7</td>
<td>Strategic U of E</td>
<td></td>
<td>27 %</td>
</tr>
<tr>
<td>8</td>
<td>Polygraph CQT &amp; CIT</td>
<td></td>
<td>2 %</td>
</tr>
</tbody>
</table>

Please compare the importance of the elements in relation to the objective and fill in the table: Which element of each pair is more important, A or B, and how much more on a scale 1-9 as given below. Once completed, you might adjust highlighted comparisons 1 to 3 to improve consistency.

<table>
<thead>
<tr>
<th>i</th>
<th>j</th>
<th>Criteria</th>
<th>Comment</th>
<th>A or B</th>
<th>Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>Direct Q &amp; Repeated Qns</td>
<td>Accusatorial Qn (Reid)</td>
<td>A 8</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>Direct Q &amp; Repeated Qns</td>
<td>Kinesic I &amp; I</td>
<td>A 7</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>Direct Q &amp; Repeated Qns</td>
<td>Elicitation</td>
<td>A 2</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>Direct Q &amp; Repeated Qns</td>
<td>Inform. Gathering Qn</td>
<td>B 2</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>6</td>
<td>Direct Q &amp; Repeated Qns</td>
<td>Cognitive I</td>
<td>B 3</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>7</td>
<td>Direct Q &amp; Repeated Qns</td>
<td>Strategic U of E</td>
<td>B 5</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>8</td>
<td>Direct Q &amp; Repeated Qns</td>
<td>Polygraph CQT &amp; CIT</td>
<td>A 9</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>Accusatorial Qn (Reid)</td>
<td>Kinesic I &amp; I</td>
<td>A 1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>Accusatorial Qn (Reid)</td>
<td>Elicitation</td>
<td>B 7</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>Accusatorial Qn (Reid)</td>
<td>Inform. Gathering Qn</td>
<td>B 7</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>Accusatorial Qn (Reid)</td>
<td>Cognitive I</td>
<td>B 8</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>7</td>
<td>Accusatorial Qn (Reid)</td>
<td>Strategic U of E</td>
<td>B 9</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>Accusatorial Qn (Reid)</td>
<td>Polygraph CQT &amp; CIT</td>
<td>A 1</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>Kinesic I &amp; I</td>
<td>Elicitation</td>
<td>B 7</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>Kinesic I &amp; I</td>
<td>Inform. Gathering Qn</td>
<td>B 7</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>6</td>
<td>Kinesic I &amp; I</td>
<td>Cognitive I</td>
<td>B 8</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>Kinesic I &amp; I</td>
<td>Strategic U of E</td>
<td>B 9</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>8</td>
<td>Kinesic I &amp; I</td>
<td>Polygraph CQT &amp; CIT</td>
<td>A 1</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>Elicitation</td>
<td>Inform. Gathering Qn</td>
<td>B 1</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>Elicitation</td>
<td>Cognitive I</td>
<td>B 3</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>7</td>
<td>Elicitation</td>
<td>Strategic U of E</td>
<td>B 4</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>Elicitation</td>
<td>Polygraph CQT &amp; CIT</td>
<td>A 9</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>Inform. Gathering Qn</td>
<td>Cognitive I</td>
<td>B 3</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>7</td>
<td>Inform. Gathering Qn</td>
<td>Strategic U of E</td>
<td>B 4</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>8</td>
<td>Inform. Gathering Qn</td>
<td>Polygraph CQT &amp; CIT</td>
<td>A 9</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>7</td>
<td>Cognitive I</td>
<td>Strategic U of E</td>
<td>B 2</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>8</td>
<td>Cognitive I</td>
<td>Polygraph CQT &amp; CIT</td>
<td>A 9</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>8</td>
<td>Strategic U of E</td>
<td>Polygraph CQT &amp; CIT</td>
<td>A 9</td>
<td></td>
</tr>
</tbody>
</table>

AHP Analytic Hierarchy Process

Objective: To compare different questioning techniques with each other in terms of their good information collection values: Little ... longer and narrative answers, support for accurate memory retrieval and constructive challenging of inconsistencies.

Only input data in the light green fields!

α: 0.1
CR: 1 %

for 9&10 unprotect the input sheets and expand the question section ("+" in row 66)
**AHP Analytic Hierarchy Process**

**Objective:** To compare different questioning techniques with each other in terms of their good information collection values: Little construction and narrative answers, support for accurate memory retrieval and constructive challenging of inconsistencies.

Please compare the importance of the elements in relation to the objective and fill in the table: Which element of each pair is more important, A or B, and how much more on a scale 1-9 as given below.

Once completed, you might adjust highlighted comparisons 1 to 3 to improve consistency.

For 9&10 unprotected the input sheets and expand the question section ("+" in row 66)

<table>
<thead>
<tr>
<th>n</th>
<th>Criteria</th>
<th>Comment</th>
<th>RGMM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Direct Q &amp; Repeated Qns</td>
<td></td>
<td>16%</td>
</tr>
<tr>
<td>2</td>
<td>Accusatorial Qn (Reid)</td>
<td></td>
<td>3%</td>
</tr>
<tr>
<td>3</td>
<td>Kinesic I &amp; I</td>
<td></td>
<td>3%</td>
</tr>
<tr>
<td>4</td>
<td>Elicitation</td>
<td></td>
<td>20%</td>
</tr>
<tr>
<td>5</td>
<td>Inform. Gathering Qn</td>
<td></td>
<td>21%</td>
</tr>
<tr>
<td>6</td>
<td>Cognitive I</td>
<td></td>
<td>27%</td>
</tr>
<tr>
<td>7</td>
<td>Strategic U of E</td>
<td></td>
<td>8%</td>
</tr>
<tr>
<td>8</td>
<td>Polygraph CQT &amp; CIT</td>
<td></td>
<td>3%</td>
</tr>
</tbody>
</table>

Support for accurate m

<table>
<thead>
<tr>
<th>Name</th>
<th>Weight</th>
<th>Date</th>
<th>Consistency</th>
<th>Ratio</th>
<th>Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Criteria</th>
<th>more important</th>
<th>Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A or B (1-9)</td>
<td></td>
</tr>
<tr>
<td>1 2 Direct Q &amp; Repeated Qns</td>
<td>A 9</td>
<td></td>
</tr>
<tr>
<td>1 3 Accusatorial Qn (Reid)</td>
<td>B 1</td>
<td></td>
</tr>
<tr>
<td>1 4 Kinesic I &amp; I</td>
<td>A 8</td>
<td></td>
</tr>
<tr>
<td>1 5 Elicitation</td>
<td>B 1</td>
<td></td>
</tr>
<tr>
<td>1 6 Inform. Gathering Qn</td>
<td>A 8</td>
<td></td>
</tr>
<tr>
<td>1 7 Cognitive I</td>
<td>B 5</td>
<td></td>
</tr>
<tr>
<td>1 8 Strategic U of E</td>
<td>A 3</td>
<td></td>
</tr>
<tr>
<td>1 9 Polygraph CQT &amp; CIT</td>
<td>B 1</td>
<td></td>
</tr>
<tr>
<td>2 3 Accusatorial Qn (Reid)</td>
<td>A 9</td>
<td></td>
</tr>
<tr>
<td>2 4 Kinesic I &amp; I</td>
<td>B 9</td>
<td></td>
</tr>
<tr>
<td>2 5 Elicitation</td>
<td>B 9</td>
<td></td>
</tr>
<tr>
<td>2 6 Inform. Gathering Qn</td>
<td>B 9</td>
<td></td>
</tr>
<tr>
<td>2 7 Cognitive I</td>
<td>B 9</td>
<td></td>
</tr>
<tr>
<td>2 8 Strategic U of E</td>
<td>B 5</td>
<td></td>
</tr>
<tr>
<td>2 9 Polygraph CQT &amp; CIT</td>
<td>B 1</td>
<td></td>
</tr>
<tr>
<td>3 4 Kinesic I &amp; I</td>
<td>B 8</td>
<td></td>
</tr>
<tr>
<td>3 5 Elicitation</td>
<td>B 9</td>
<td></td>
</tr>
<tr>
<td>3 6 Inform. Gathering Qn</td>
<td>B 9</td>
<td></td>
</tr>
<tr>
<td>3 7 Cognitive I</td>
<td>B 9</td>
<td></td>
</tr>
<tr>
<td>3 8 Strategic U of E</td>
<td>B 5</td>
<td></td>
</tr>
<tr>
<td>3 9 Polygraph CQT &amp; CIT</td>
<td>B 2</td>
<td></td>
</tr>
<tr>
<td>4 5 Elicitation</td>
<td>A 1</td>
<td></td>
</tr>
<tr>
<td>4 6 Inform. Gathering Qn</td>
<td>B 2</td>
<td></td>
</tr>
<tr>
<td>4 7 Cognitive I</td>
<td>A 5</td>
<td></td>
</tr>
<tr>
<td>4 8 Polygraph CQT &amp; CIT</td>
<td>A 9</td>
<td></td>
</tr>
<tr>
<td>5 6 Inform. Gathering Qn</td>
<td>B 2</td>
<td></td>
</tr>
<tr>
<td>5 7 Cognitive I</td>
<td>A 5</td>
<td></td>
</tr>
<tr>
<td>5 8 Polygraph CQT &amp; CIT</td>
<td>A 9</td>
<td></td>
</tr>
<tr>
<td>6 7 Strategic U of E</td>
<td>A 7</td>
<td></td>
</tr>
<tr>
<td>6 8 Polygraph CQT &amp; CIT</td>
<td>A 9</td>
<td></td>
</tr>
<tr>
<td>7 8 Strategic U of E</td>
<td>A 7</td>
<td></td>
</tr>
</tbody>
</table>

Table 5. Comparison table of criterion Constructive challenging of inconsistencies.
ANALYSIS OF THE METHOD FOR ASSESSING THE CREDIBILITY OF EVIDENCE (MACE)

In the following tables different assessment criteria of MACE are analyzed from the deliberate manipulation’s and deception’s point of view. In each table Questions are described as they were in the original study of Schum and Morris (2007). They are followed by the analysis of source’s chances for intentional manipulation of MACE’s results.

Regarding Competence questions source’s ability to manipulate the assessment, other sources of information are analyzed with additional conclusions from these findings.

Questions concerning Credibility were analyzed and classified by their aim and after that Evaluation Categories were created as a synthesis. It was found out that Credibility evaluation falls under seven categories:

- Support from other sources (in 6 questions),
- Internal consistency of the source (4),
- Source’s motives to manipulate the facts (4),
- Intentional external manipulation of the source (2),
- Value of source’s sensory information (2),
- Source’s nonverbal communication (1) and
- Source’s ability to remember correctly (1).

Concerning twelve of these twenty questions the source is able to manipulate the outcome of the assessment by lying and creating a plausible cover story with other people involved in the course of events. If accomplices are not available only six answers to these twenty questions are subjected to source’s manipulation.

According the analysis questions used in MACE seem to build a strong foundation for a comprehensive, objective and evidence based credibility assessment.
<table>
<thead>
<tr>
<th>QUESTION (Schum and Morris 2007, 23-24)</th>
<th>CHANCE FOR SOURCE'S INTENTIONAL MANIPULATION</th>
<th>SOURCE'S ABILITY TO MANIPULATE OTHER SOURCES OF INFORMATION</th>
<th>CONCLUSIONS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Did this source actually make the observation being claimed or have access to the information reported?</td>
<td>Some chances concerning the story about claimed observation or the access to the information.</td>
<td>Cooperation with other human sources is possible if setup enables it.</td>
<td>Other human sources make manipulation possible. Hard to manipulate sources that are not related to people.</td>
</tr>
<tr>
<td>2. Does this source have an understanding of what was observed or have any knowledge or expertise regarding this observation?</td>
<td>May claim to have adequate understanding or expertise.</td>
<td>Backup for the claims from other human sources is possible if setup enables it.</td>
<td>Other human sources make manipulation possible. Hard to manipulate sources that are not related to people.</td>
</tr>
<tr>
<td>3. Is this source generally a capable observer?</td>
<td>May try to be convincing about his observing abilities.</td>
<td>Poor, at least to the extent that other person's opinion would win the investigators' observations or other evidence.</td>
<td>Low chances for successful manipulation.</td>
</tr>
<tr>
<td>4. Has this source been consistent in his/her motivation to provide us with information?</td>
<td>May create a cover story with plausible explanation for motives.</td>
<td>Cooperation with other human sources is possible if setup enables it.</td>
<td>Other human sources make manipulation possible. Hard to manipulate sources that are not related to people.</td>
</tr>
<tr>
<td>5. Has this source been responsive to inquiries we have made of him/her?</td>
<td>May control his/her own behavior and actively seek cooperation.</td>
<td>Question is related only to this particular source.</td>
<td>No chances for additional manipulation.</td>
</tr>
</tbody>
</table>

Table 1. Analysis of MACE's Competence questions.
<table>
<thead>
<tr>
<th>QUESTION FOR VERACITY (Schum and Morris 2007, 24-25)</th>
<th>EVALUATION CATEGORY</th>
<th>SOURCE’S ABILITY TO MANIPULATE THE ASSESSMENT</th>
<th>CONCLUSIONS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Has this source told us anything that is inconsistent with what he/she has just reported to us?</td>
<td>Internal consistency of the source.</td>
<td>May rehearse a cover story.</td>
<td>Well rehearsed cover story can be consistent.</td>
</tr>
<tr>
<td>2. Is this source subject to any outside influences?</td>
<td>Intentional external manipulation of the source.</td>
<td>May claim to be free of influence and give appealing reasons for that.</td>
<td>Source’s statements should not have influence on the evidence from other sources.</td>
</tr>
<tr>
<td>3. Could this source have been exploited in any way in his/her reporting to us?</td>
<td>Intentional external manipulation of the source.</td>
<td>May claim to be free of exploitation and give appealing reasons for that.</td>
<td>Source’s statements should not have influence on the evidence from other sources.</td>
</tr>
<tr>
<td>4. Is there any other evidence from other sources that contradicts or conflicts with what this source has just reported?</td>
<td>Support from other sources.</td>
<td>Cooperation with other human sources is possible if setup enables it.</td>
<td>Other human sources make manipulation possible. Hard to manipulate sources that are not related to people.</td>
</tr>
<tr>
<td>5. Is there any evidence from other sources that corroborates or confirms what this source has just told us?</td>
<td>Support from other sources.</td>
<td>Cooperation with other human sources is possible if setup enables it.</td>
<td>Other human sources make manipulation possible. Hard to manipulate sources that are not related to people.</td>
</tr>
<tr>
<td>6. What evidence do we have about this source’s character and honesty?</td>
<td>Internal consistency of the source and support from other sources.</td>
<td>May rehearse a cover story. Cooperation with other human sources is possible if setup enables it.</td>
<td>Well rehearsed cover story can be consistent. Other human sources make manipulation possible. Hard to manipulate sources that are not related to people.</td>
</tr>
<tr>
<td>7. What does this source’s reporting track record show about his/her honesty in reporting to us?</td>
<td>Internal consistency of the source.</td>
<td>May rehearse a cover story.</td>
<td>Well rehearsed cover story can be consistent.</td>
</tr>
<tr>
<td>8. Is there evidence that this source tailored this report in a way that he/she believes will capture our attention?</td>
<td>Source’s motives to alter the facts.</td>
<td>May try to exhibit sincerity.</td>
<td>Other human sources make manipulation possible. Hard to manipulate sources that are not related to people.</td>
</tr>
<tr>
<td>9. Are there collateral details in this report that reflect the possibility of this source’s dishonesty?</td>
<td>Internal consistency of the source.</td>
<td>May rehearse a cover story.</td>
<td>Well rehearsed cover story can be consistent.</td>
</tr>
<tr>
<td>10. Demeanor and bearing of this source during his/her interview?</td>
<td>Observations about source’s nonverbal communication.</td>
<td>May rehearse confidence evoking and credible behavior.</td>
<td>Relies solely on easily manipulated single source.</td>
</tr>
</tbody>
</table>

Table 2. Analysis of MACE’s Credibility questions regarding Veracity.
<table>
<thead>
<tr>
<th>QUESTION FOR OBJECTIVITY (Schum and Morris 2007, 26)</th>
<th>EVALUATION CATEGORY</th>
<th>SOURCE’S ABILITY TO MANIPULATE THE ASSESSMENT</th>
<th>CONCLUSIONS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Is there evidence about what this source expected to observe during his/her reported observation?</td>
<td>Source’s motives (biases) to alter the facts.</td>
<td>May claim to be free of biases and give appealing reasons for that.</td>
<td>Source’s statements should not have influence on the evidence from other sources.</td>
</tr>
<tr>
<td>2. Is there evidence about what this source wished to observe during his/her reported observation?</td>
<td>Source’s motives (desires) to alter the facts.</td>
<td>May claim to be free of interfering motives and give appealing reasons for that.</td>
<td>Source’s statements should not have influence on the evidence from other sources.</td>
</tr>
<tr>
<td>3. Was this source concerned about the consequences of what he/she believed during his/her observation?</td>
<td>Source’s motives (concerns) to alter the facts.</td>
<td>May claim to be free of interfering motives and give appealing reasons for that.</td>
<td>Source’s statements should not have influence on the evidence from other sources.</td>
</tr>
<tr>
<td>4. Is there any evidence concerning possible defects in this source’s memory? Also, how long ago did this source’s observation take place?</td>
<td>Ability to remember accurately.</td>
<td>May deny having any defects in memory or may praise his/her good memory.</td>
<td>Source’s statements should not have influence on the evidence from other sources.</td>
</tr>
<tr>
<td>5. Is there any other evidence from other sources that contradicts or conflicts with what this source has just reported?</td>
<td>Support from other sources.</td>
<td>Cooperation with other human sources is possible if setup enables it.</td>
<td>Other human sources make manipulation possible. Hard to manipulate sources that are not related to people.</td>
</tr>
</tbody>
</table>

Table 3. Analysis of MACE’s Credibility questions regarding Objectivity.
<table>
<thead>
<tr>
<th>Question</th>
<th>Evaluation Category</th>
<th>Source's Ability to Manipulate the Assessment</th>
<th>Conclusions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. The source's sensory capacity at the time of the observation?</td>
<td>Testimonial value of sensory information, ability to observe.</td>
<td>May claim to have unobstructed sensory capacities.</td>
<td>Source's statements should not have influence on the evidence from other sources.</td>
</tr>
<tr>
<td>2. The conditions under which the observation took place?</td>
<td>Testimonial value of sensory information, ability to make correct observations.</td>
<td>May claim to have unobstructed sensory capacities.</td>
<td>Source's statements should not have influence on the evidence from other sources.</td>
</tr>
<tr>
<td>3. The source's track record of accuracy in his/her reports?</td>
<td>Internal consistency of the source.</td>
<td>May rehearse a cover story.</td>
<td>Well rehearsed cover story may be consistent and accurate.</td>
</tr>
<tr>
<td>4. Is there any other evidence from other sources that contradicts or conflicts with what this source has just reported?</td>
<td>Support from other sources.</td>
<td>Cooperation with other human sources is possible if setup enables it.</td>
<td>Other human sources make manipulation possible. Hard to manipulate sources that are not related to people.</td>
</tr>
<tr>
<td>5. Are there collateral details in this report that reflect the possibility of this source's inaccuracy?</td>
<td>Internal consistency of the source.</td>
<td>May rehearse a cover story.</td>
<td>Well rehearsed cover story may be consistent and accurate.</td>
</tr>
</tbody>
</table>

Table 4. Analysis of MACE's Credibility questions regarding Observational Sensitivity.
THE LIST OF THE CONTENT BASED CREDIBILITY ASSESSMENT (CBCA) CRITERIA

1. Logical structure
2. Unstructured production
3. Quantity of details
4. Contextual embedding
5. Description of interaction
6. Reproduction of conversation
7. Unexpected complications
8. Unusual details
9. Superfluous details
10. Accurately reported details misunderstood
11. Related external associations
12. Subjective mental state
13. Perpetrator’s mental state
14. Spontaneous corrections
15. Lack of memory
16. Self-doubts on the memories or knowledge
17. Self-deprecation
18. Pardoning the perpetrator
19. Details characteristic of the offense

AHP COMPARISON OF THE APPLICABILITY CRITERIA

Table 1. Comparison of the veracity assessment applicability criteria.
# AHP COMPARISON OF THE VERACITY ASSESSMENT METHODS

The Analytic Hierarchy Process (AHP) is a multi-criteria decision making technique that allows for the comparison of different methods against each other based on a set of criteria. In this section, we will compare the veracity assessment methods using AHP. The methods include Plausibility & Inner Logic, CBCA, Discourse Analysis, Selected feat. Discourse Analysis, Selected feat. Non-verbal Communication, MACE, and Polygraph. The comparison is done using a pairwise comparison matrix with five criteria: 1: Accuracy, 2: Ease of Use, 3: Time Required, 4: No Need for Special Equipment and 5: Is Unobtrusive.

### Objective
Rate methods against each other with according criteria: 1: Accuracy, 2: Ease of Use, 3: Time Required, 4: No Need for Special Equipment and 5: Is Unobtrusive.

### Author
MUO

### Date
March 2015

### Table 1

<table>
<thead>
<tr>
<th>Table</th>
<th>Criterion</th>
<th>Comment</th>
<th>Weights</th>
<th>Rk</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Plausibility &amp; Inner Logic</td>
<td></td>
<td>16.6%</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>CBCA</td>
<td></td>
<td>12.9%</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>Discourse Analysis</td>
<td></td>
<td>9.0%</td>
<td>6</td>
</tr>
<tr>
<td>4</td>
<td>Selected feat. Discourse Analysis</td>
<td></td>
<td>21.9%</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>Selected feat. Non-verbal Communication</td>
<td></td>
<td>20.3%</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>MACE</td>
<td></td>
<td>14.3%</td>
<td>4</td>
</tr>
<tr>
<td>7</td>
<td>Polygraph</td>
<td></td>
<td>4.8%</td>
<td>7</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td>0.0%</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td>0.0%</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td>0.0%</td>
<td></td>
</tr>
</tbody>
</table>

### Result

<table>
<thead>
<tr>
<th>Weight</th>
<th>Rk</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>8</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
</tr>
</tbody>
</table>

### Eigenvalue

| Consistency Ratio | 0.37 | GCI  | 0.01 | CR  | 0.2% |

### Matrix

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Plausibility &amp; Inner Logic</th>
<th>CBCA</th>
<th>Discourse Analysis</th>
<th>Selected feat. Discourse Analysis</th>
<th>Selected feat. Non-verbal Communication</th>
<th>MACE</th>
<th>Polygraph</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5/7</td>
<td>-</td>
<td>1/2</td>
<td>1/2</td>
<td>1/5</td>
<td>1/4</td>
</tr>
<tr>
<td>2</td>
<td>1/2</td>
<td>1/2</td>
<td>-</td>
<td>2/3</td>
<td>3/8</td>
<td>2/3</td>
</tr>
<tr>
<td>3</td>
<td>5/7</td>
<td>1/2</td>
<td>1/2</td>
<td>-</td>
<td>1/5</td>
<td>1/4</td>
</tr>
<tr>
<td>4</td>
<td>1/2</td>
<td>1/2</td>
<td>2/3</td>
<td>5/6</td>
<td>1/5</td>
<td>1/4</td>
</tr>
<tr>
<td>5</td>
<td>1/2</td>
<td>1/2</td>
<td>2/3</td>
<td>2/3</td>
<td>1/2</td>
<td>1/4</td>
</tr>
<tr>
<td>6</td>
<td>1/2</td>
<td>1/2</td>
<td>2/3</td>
<td>2/3</td>
<td>1/2</td>
<td>1/4</td>
</tr>
<tr>
<td>7</td>
<td>1/2</td>
<td>1/2</td>
<td>2/3</td>
<td>2/3</td>
<td>1/2</td>
<td>1/4</td>
</tr>
<tr>
<td>8</td>
<td>1/2</td>
<td>1/2</td>
<td>2/3</td>
<td>2/3</td>
<td>1/2</td>
<td>1/4</td>
</tr>
<tr>
<td>9</td>
<td>1/2</td>
<td>1/2</td>
<td>2/3</td>
<td>2/3</td>
<td>1/2</td>
<td>1/4</td>
</tr>
<tr>
<td>10</td>
<td>1/2</td>
<td>1/2</td>
<td>2/3</td>
<td>2/3</td>
<td>1/2</td>
<td>1/4</td>
</tr>
</tbody>
</table>

### Normalized Principal Eigenvector

<table>
<thead>
<tr>
<th>Plausibility &amp; Inner Logic</th>
<th>CBCA</th>
<th>Discourse Analysis</th>
<th>Selected feat. Discourse Analysis</th>
<th>Selected feat. Non-verbal Communication</th>
<th>MACE</th>
<th>Polygraph</th>
</tr>
</thead>
<tbody>
<tr>
<td>16.80%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12.88%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9.00%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21.86%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20.35%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14.31%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.80%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.00%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.00%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.00%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 1. Total applicability, equally weighted criteria.
<table>
<thead>
<tr>
<th>n</th>
<th>Criteria</th>
<th>Comment</th>
<th>RGMM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Plausibility &amp; Inner Logic</td>
<td></td>
<td>11%</td>
</tr>
<tr>
<td>2</td>
<td>CBCA</td>
<td></td>
<td>9%</td>
</tr>
<tr>
<td>3</td>
<td>Discourse Analysis</td>
<td></td>
<td>28%</td>
</tr>
<tr>
<td>4</td>
<td>Studied feat. Discourse</td>
<td></td>
<td>15%</td>
</tr>
<tr>
<td>5</td>
<td>Studied feat. Non-verbal</td>
<td></td>
<td>13%</td>
</tr>
<tr>
<td>6</td>
<td>MACE</td>
<td></td>
<td>19%</td>
</tr>
<tr>
<td>7</td>
<td>Polygraph</td>
<td></td>
<td>5%</td>
</tr>
</tbody>
</table>

for 9&10 unprotect the input sheets and expand the question section ("+" in row 66)

<table>
<thead>
<tr>
<th>Name</th>
<th>Weight</th>
<th>Date</th>
<th>Accuracy</th>
<th>Consistency Ratio</th>
<th>Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td></td>
<td></td>
<td></td>
<td>0,1</td>
<td>5</td>
</tr>
<tr>
<td>B</td>
<td></td>
<td></td>
<td></td>
<td>1%</td>
<td>5</td>
</tr>
</tbody>
</table>

Table 2. Comparison table of criterion *Accuracy*, equally weighted.
<table>
<thead>
<tr>
<th>Criteria</th>
<th>Comment</th>
<th>RGMM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plausibility &amp; Inner Logic</td>
<td></td>
<td>16%</td>
</tr>
<tr>
<td>CBCA</td>
<td></td>
<td>13%</td>
</tr>
<tr>
<td>Discourse Analysis</td>
<td></td>
<td>4%</td>
</tr>
<tr>
<td>Studied feat. Discourse</td>
<td></td>
<td>29%</td>
</tr>
<tr>
<td>Studied feat. Non-verbal Communication</td>
<td></td>
<td>23%</td>
</tr>
<tr>
<td>MACE</td>
<td></td>
<td>11%</td>
</tr>
<tr>
<td>Polygraph</td>
<td></td>
<td>4%</td>
</tr>
</tbody>
</table>

for 9&10 unprotect the input sheets and expand the question section (“+” in row 66)

<table>
<thead>
<tr>
<th>Criteria</th>
<th>more important?</th>
<th>Scale (1-9)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ease of Use</td>
<td>A: 0,1</td>
<td>CR: 1%</td>
</tr>
</tbody>
</table>

Table 3. Comparison table of criterion Ease of Use, equally weighted.
<table>
<thead>
<tr>
<th>Criteria</th>
<th>Comment</th>
<th>RGMM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Plausibility &amp; Inner Logic</td>
<td>17%</td>
</tr>
<tr>
<td>2</td>
<td>CBCA</td>
<td>8%</td>
</tr>
<tr>
<td>3</td>
<td>Discourse Analysis</td>
<td>3%</td>
</tr>
<tr>
<td>4</td>
<td>Studied feat. Discourse</td>
<td>27%</td>
</tr>
<tr>
<td>5</td>
<td>Studied feat. Non-verbal Communication</td>
<td>27%</td>
</tr>
<tr>
<td>6</td>
<td>MACE</td>
<td>8%</td>
</tr>
<tr>
<td>7</td>
<td>Polygraph</td>
<td>10%</td>
</tr>
</tbody>
</table>

Table 4. Comparison table of criterion *Time Required*, equally weighted.
Table 5. Comparison table of criterion *No Need for Special Equipment*, equally weighted.
Table 6. Comparison table of criterion *Unobtrusiveness*, equally weighted.
AHP Analytic Hierarchy Process (EVM multiple inputs)

Only input data in the light green fields and worksheets!

n= 7 Number of criteria (3 to 10) Scale: 5
N= 5 Number of Participants (1 to 20) \( \alpha : 0.1 \) Consensus: 82.2%
p= 0 selected Participant (0=consol.) 2 7 Consolidated

Objective
Rate methods against each other with according criteria: 1: Accuracy, 2: Ease of Use, 3: Time Required, 4: No Need for Special Equipment and 5: Is Unobtrusive

Author MUO
Date March 2015
Thresh: 1E-07 Iterations: 4 EVM check: 5,5E-09

Table
<table>
<thead>
<tr>
<th>Criterion</th>
<th>Comment</th>
<th>Weights</th>
<th>Rk</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Plausibility &amp; Inner Logic</td>
<td></td>
<td>17.7%</td>
<td>3</td>
</tr>
<tr>
<td>2 CBCA</td>
<td></td>
<td>9.6%</td>
<td>5</td>
</tr>
<tr>
<td>3 Discourse Analysis</td>
<td></td>
<td>4.4%</td>
<td>7</td>
</tr>
<tr>
<td>4 Selected feat. Discourse</td>
<td></td>
<td>25.5%</td>
<td>1</td>
</tr>
<tr>
<td>5 Selected feat. Non-verbal Communication</td>
<td></td>
<td>24.9%</td>
<td>2</td>
</tr>
<tr>
<td>6 MACE</td>
<td></td>
<td>10.0%</td>
<td>4</td>
</tr>
<tr>
<td>7 Polygraph</td>
<td></td>
<td>7.8%</td>
<td>6</td>
</tr>
<tr>
<td>8</td>
<td>for 9&amp;10 unprotect the input sheets and expand the question section (&quot;+&quot; in row 66)</td>
<td>0.0%</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td>0.0%</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td>0.0%</td>
<td></td>
</tr>
</tbody>
</table>

Result
Eigenvalue
lambda: 7.079
Consistency Ratio 0.37 GCI: 0.04 CR: 1.0%

Matrix

<table>
<thead>
<tr>
<th>Plausibility &amp; Inner Logic</th>
<th>CBCA</th>
<th>Discourse Analysis</th>
<th>Selected feat. Discourse</th>
<th>Selected feat. Non-verbal Communication</th>
<th>MACE</th>
<th>Polygraph</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1/3</td>
<td>1/2/3</td>
<td>1/3/3</td>
<td>2/3</td>
<td>5/7</td>
<td>1/7</td>
</tr>
<tr>
<td>2</td>
<td>3/5</td>
<td>2/4/5</td>
<td>1/3</td>
<td>1/3</td>
<td>1</td>
<td>1/1/9</td>
</tr>
<tr>
<td>3</td>
<td>1/5</td>
<td>1/3</td>
<td>1/5</td>
<td>1/5</td>
<td>1/2</td>
<td>5/7</td>
</tr>
<tr>
<td>4</td>
<td>1/1/2</td>
<td>2 8/9</td>
<td>5 4/9</td>
<td>1</td>
<td>2/5/6</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>1/2/5</td>
<td>2 6/7</td>
<td>5 4/9</td>
<td>1</td>
<td>2/5/6</td>
<td>2 5/6</td>
</tr>
<tr>
<td>6</td>
<td>7/8</td>
<td>1</td>
<td>2</td>
<td>1/3</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>1/3</td>
<td>1/2/5</td>
<td>1/3</td>
<td>1/3</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

normalized principal Eigenvector

<table>
<thead>
<tr>
<th>Plausibility &amp; Inner Logic</th>
<th>CBCA</th>
<th>Discourse Analysis</th>
<th>Selected feat. Discourse</th>
<th>Selected feat. Non-verbal Communication</th>
<th>MACE</th>
<th>Polygraph</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>17.71 %</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>9.63 %</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>4.38 %</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>25.54 %</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>24.93 %</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>10.00 %</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>7.81 %</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0.00 %</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>0.00 %</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.00 %</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 7. Total applicability, Time Required weighted 10 times over others.
AHP Analytic Hierarchy Process (EVM multiple inputs)  
Free web based AHP software on:  http://bpmsg.com

Only input data in the light green fields and worksheets!

n= 7  Number of criteria (3 to 10)  Scale: 5  Linear
N= 5  Number of Participants (1 to 20)  \( \alpha \): 0,1  Consensus: 82,2%
p= 0  selected Participant (0=consol.)  2  7  Consolidated

Objective
Rate methods agains each other with according criteria: 1: Accuracy, 2: Ease of Use, 3: Time Required, 4: No Need for Special Equipment and 5: Is Unobtrusive

Author  MUO
Date  March 2015  Thresh: 1E-07  Iterations: 4  EVM check: 7,1E-09

Table
<table>
<thead>
<tr>
<th>Criterion</th>
<th>Comment</th>
<th>Weights</th>
<th>Rk</th>
</tr>
</thead>
<tbody>
<tr>
<td>1  Plausibility &amp; Inner Logic</td>
<td></td>
<td>18,1%</td>
<td>3</td>
</tr>
<tr>
<td>2  CBCA</td>
<td></td>
<td>15,6%</td>
<td>4</td>
</tr>
<tr>
<td>3  Discourse Analysis</td>
<td></td>
<td>8,0%</td>
<td>6</td>
</tr>
<tr>
<td>4  Selected feat. Discourse</td>
<td></td>
<td>19,2%</td>
<td>1</td>
</tr>
<tr>
<td>5  Selected feat. Non-Verbal Communication</td>
<td></td>
<td>18,7%</td>
<td>2</td>
</tr>
<tr>
<td>6  MACE</td>
<td></td>
<td>15,5%</td>
<td>5</td>
</tr>
<tr>
<td>7  Polygraph</td>
<td></td>
<td>4,9%</td>
<td>7</td>
</tr>
<tr>
<td>8  Discourse Analysis</td>
<td></td>
<td>0,0%</td>
<td>7</td>
</tr>
<tr>
<td>9  Selected feat. Non-Verbal Communication</td>
<td></td>
<td>0,0%</td>
<td>7</td>
</tr>
<tr>
<td>10 MACE</td>
<td></td>
<td>0,0%</td>
<td>7</td>
</tr>
</tbody>
</table>

Result
Eigenvalue  lambda: 7,018
Consistency Ratio  0,37  GCI: 0,01  CR: 0,2%

<table>
<thead>
<tr>
<th>Matrix</th>
<th>normalized principal Eigenvector</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plausibility &amp; Inner Logic</td>
<td></td>
</tr>
<tr>
<td>1 1</td>
<td>(18,12 %)</td>
</tr>
<tr>
<td>2 7/9</td>
<td></td>
</tr>
<tr>
<td>3 4/9</td>
<td></td>
</tr>
<tr>
<td>4 1/8</td>
<td></td>
</tr>
<tr>
<td>5 1 1/6</td>
<td></td>
</tr>
<tr>
<td>6 7/8</td>
<td></td>
</tr>
<tr>
<td>7 1/4</td>
<td></td>
</tr>
<tr>
<td>8 2/7</td>
<td></td>
</tr>
<tr>
<td>9 5/8</td>
<td></td>
</tr>
<tr>
<td>10 1/4</td>
<td></td>
</tr>
</tbody>
</table>

Table 8. Total applicability, No Need for Special Equipment weighted 10 times over others.