UNIVERSITÀ DEGLI STUDI DI BERGAMO | Dipartimento di Scienze Umane e Sociali

UNIVERSITY OF BERGAMO

Department of Human and Social Sciences

Ph.D. Course in "Human Capital Formation and Labour Relations"

Curriculum: Clinical Psychology

SSD: M-PSI/05

XXXII Cycle

Educing Information and Deception Detection: Testing the Efficacy of the Baseline Approach and of Social Influence Tactics in HUMINT Interviewing

Supervisor:

Chiar.ma Prof.ssa Letizia Caso

Coordinator

Chiar.mo Prof. Giuseppe Bertagna

Doctoral Thesis

Nicola PALENA

Student ID: 1044007

ACADEMIC YEAR 2018 / 2019

To Lucrezia. I met you late in my PhD journey, but you made it much better. You are the best thing that has ever happened to me.

To my Father, Mother, Sister and Brother.
To my grandparents and my whole family,
including my two lovely dogs.

To my friends, the true ones

Contents

Acknowledgments

I gratefully thank my supervisor Prof. Letizia Caso, who accompanied and guided me throughout this journey, and patiently endured my doubts and fears. I would also like to thank Prof. Vrij, for being so helpful and a guide for the projects we worked together, and Prof. Granhag, for hosting me at Gothenburg university and for the insights he provided me. A special thanks goes to Professors Braibanti, Morganti and Greco, for also being helpful and supporting and kind to me, even when it came to "chit-chat talking". I have learnt a lot from all of you. Thanks to all my colleagues, for the time spent together, also to discuss about our fears and doubts. Among them, thanks to Lavinia Pontigia, who also helped me with proof-reading. Thanks to Robin Orthey, who I met as a colleague but then became a close friend, and to all the students who helped me in various ways, including but not limited to: Luca Melocchi, Roberto Monticciolo, Viola Beletti, Elena Fratus, Alessia Ambrosini and Camilla Coita. Thanks to Elga Carlessi, the public prosecutor Mr. Mapelli, who we all missed, Dott. Giobbi from the Italian police force, Col. Storoni from the "*Arma dei Carabinier*i", Col. Picciafuochi and Maj. Peluso, from the "*Guardia di Finanza*", and all the other practitioners involved in our shared project: I hope we will continue this collaboration. Finally, thanks to the reviewers who will read this thesis and will provide stimulating insights.

Abstract

English Abstract

The present dissertation focused on two strictly intertwined research topics: Intelligence gathering and deception detection. Such research areas play an important role in today's security settings, especially after 9/11 and other terrorist attacks spread around the globe. The first section (Chapter 1 and Chapter 2) of this work focuses on definitions and background theories of Intelligence gathering and deception detection, as well as on the most effective interviewing techniques available to date (Chapter 3). The second section focuses on empirical studies conducted within this project (Chapter 4). The first experiment concentrates on the role of social influence – operationalized as a form of reciprocity – within Scharff interviews. The experiment was conducted under the assumption that sources can be uncooperative, and there might be the need to persuade them to reveal sensitive information. However, reciprocity did not seem to have a relevant effect on the amount of information disclosed. The third section (from 4.3 to 4.7) of the dissertation concentrates on five experiments that explored the baseline approach. Indeed, after collecting information, the investigators need to assess the veracity of what the interviewee has said. The baseline technique aims at reducing the difficulty of deception detection ascribed to interpersonal differences, and it can be obtained in different ways. The first two experiments on this approach showed that creating a baseline through small-talk questioning is ineffective, while the comparable truth approach seems to be more effective. Indeed, liars showed a greater variation between the different phases of the interview than truth tellers did. The third, fourth, and fifth experiments on the baseline approach focused on the effect it has on the observers. Results showed that the apparent advantage of the comparable truth approach on observers' lie detection accuracy can be explained in terms of a bias shift. That is, rather than improving observers' accuracy when detecting deception, the comparable truth approach apparently reduces their truth bias. Due to the limitation of the baseline approach - partly due to the fact that it makes a comparison of different statements made by the same interviewee during an interview, a *between-statements* comparison – a seventh study explored a novel interviewing approach, named "Theme-Selection Strategy". The approach focuses on the analysis of different topics, or "themes", discussed by

the interviewee within a single interview, a *within-statements* comparison. The results supported the assumption that *within-statements* comparisons are more effective that *between-statements* comparisons, and the hypothesis that the former may be more effective for detecting mixtures of truths and lies than the latter. In conclusion, the research presented in this dissertation leads to two main conclusions. First, there is the need to find effective strategies to persuade sources to disclose relevant information to the investigator, as the use of reciprocity did not seem to lead to satisfactory results. Second, research on deception detection may benefit from moving the attention from *between-subjects* comparisons to *within-subjects* and *within-statements* comparisons.

Abstract Italiano

La presente tesi si è concentrata su due argomenti di ricerca strettamente interconnessi: la raccolta di informazioni e il rilevamento della menzogna. Tali aree di ricerca svolgono un ruolo importante nei contesti di sicurezza nazionale, soprattutto dopo l'11 settembre e gli altri attacchi terroristici avvenuti in tutto il mondo. La prima sezione di questo lavoro (Capitoli 1 e 2) si concentra sulle definizioni e sulle teorie di riferimento che riguardano la raccolta di informazioni (intelligence) ed il rilevamento della menzogna, nonché sulle tecniche di intervista più efficaci ad oggi disponibili (Capitolo 3). La seconda sezione si concentra sugli studi empirici condotti nell'ambito di questo progetto (Capitolo 4). Il primo esperimento presentato si concentra sul ruolo dell'influenza sociale – operazionalizzata secondo il principio della reciprocità – all'interno delle interviste Scharff. L'esperimento è stato condotto tenendo presente che le fonti possono essere poco cooperative e che quindi c'è la necessità di persuaderle a rivelare informazioni sensibili. Tuttavia, la reciprocità non sembra avere avuto un effetto rilevante sulla quantità di informazioni rivelate. La terza sezione (da 4.3 a 4.7) si focalizza su cinque di esperimenti che esplorano l'approccio della baseline. Infatti, dopo aver raccolto le informazioni, gli investigatori devono valutare la veridicità di ciò che la persona intervistata ha dichiarato. La tecnica di baseline mira a ridurre le problematiche che si riscontrano quando si deve valutare la veridicità di quanto dichiarato dalla persona intervistata che sono dovute a differenze interpersonali. Tale tecnica può essere costruita in modi diversi. I primi due esperimenti su questo approccio hanno mostrato che ottenere una baseline attraverso

domande neutre non riguardanti l'evento sotto indagine è inefficace, mentre ottenere una comparable truth baseline può essere più efficace in quanto i partecipanti che avevano mentito hanno mostrato una maggiore variazione tra le diverse fasi dell'intervista rispetto a coloro che avevano detto la verità. Il terzo, quarto e quinto esperimento sulla baseline si sono concentrati sull'effetto che questa ha sull'accuratezza degli osservatori nel discriminare tra verità e bugia e hanno dimostrato che l'apparente vantaggio della comparable truth baseline sull'accuratezza degli osservatori può essere spiegato in termini di un cambiamento di bias. In breve, invece che migliorare l'accuratezza degli osservatori, l'approccio della comparable truth baseline tende a ridurre il truth bias. A causa dei limiti di questo approccio - in parte dovuti al fatto che esso si basa su un confronto tra varie dichiarazioni date da uno stesso soggetto durante un'intervista, un confronto *tra le dichiarazioni* - un settimo studio ha esplorato un nuovo approccio di intervista, chiamato "Theme-Selection Strategy". Tale approccio si concentra sull'analisi di diversi argomenti, o "temi", discussi dall'intervistato all'interno di una singola dichiarazione, un confronto *all'interno delle dichiarazioni.* I risultati sembrano aver sostenuto l'ipotesi che i confronti *all'interno delle dichiarazioni* sono più efficaci rispetto ai confronti *tra le dichiarazioni* e l'ipotesi che il primo tipo di confronto può essere più efficace del secondo per rilevare la menzogna quando la dichiarazione contiene un misto di verità e bugia. In conclusione, la ricerca presentata in questa tesi porta a due conclusioni principali. In primo luogo, vi è la necessità di trovare strategie efficaci per convincere le fonti a rivelare informazioni all'investigatore, in quanto l'uso del principio di reciprocità non ha portato a risultati soddisfacenti. In secondo luogo, la ricerca sul rilevamento della menzogna può trarre vantaggio dallo spostare l'attenzione dai confronti *tra soggetti* ai confronti *fra soggetti* e *all'interno delle dichiarazioni.*

Overview

The need to collect information for security reasons has accompanied the human species for a long time. One of the first secret services can be dated back to A.D. 100, when Domitian instituted the G-4 reform of the Imperial General Staff[1]. Yet, also before that time civilians acted as informants to monitor any potential risk. Romans were not the only ancients who had a secret service: The Spartans had their *Krypteia*, and the Persians also had their intelligence agency (Sinnigen, 1961). The need for intelligence agencies is clear: They aim at protecting their own State, government and people. Also, secret services deal with economic risks and attacks, as well as with cybersecurity and terrorism. The two latter aspects are particularly topical. Indeed, cyberattacks are believed to become a new mean aiming to weaken enemy nations and their economic power. Also, the risk of terrorist attacks increased steeply in recent days, especially after 9/11 (Global Terrorism Index, 2017).

The need to protect the Nation from cyber and economic risks prompted the Italian Intelligence Agency to recruit new IT and cybersecurity professionals, as it was feared that cyberattacks would considerably increase. Indeed, the fear started becoming reality when, in late November 2018, the Anonymous Group hacked several hundred (Italian) institutional and governmental certified emails account (including health services and universities accounts, see Il Sole 24 Ore, 2018; Sistema di Informazione per la Sicurezza della Repubblica, 2018).

Similarly, the increased risk of terrorist attacks in Western cultures made it clear that HUMINT interviewing is very important to mitigate the problem. HUMINT interrogations, indeed, are essential to gather information on terrorist groups and potential attacks in order to prevent fatalities (Bowman, 2010; Brandon, 2011; Granhag, Vrij, & Meissner, 2014; Loftus, 2011).

In recent times, there has been a flourishing collaboration between academics and practitioners to develop effective evidence-based HUMINT interviewing techniques

[1] Due to the vast territory the Roman Empire covered, and to the raise of wars, there was the need to be informed of what was happening all over the empire. Hence, Dominitian created the G-4 reform as a supply section of the Imperial General stuff. With the reform, supply sergeant who had, as their original function that of purchasing and delivering grain to the Romanian troops, actually acted as informants.

(Brandon, 2011; Granhag et al., 2014)- which will be described in more detail in Chapter 1- such as the Scharff technique (Granhag, Kleinman, & Oleszkiewicz, 2016; Oleszkiewicz, Granhag, & Cancino Montecinos, 2014).

Anyway, gathering information is just one of several aspects of any HUMINT interview. Indeed, once practitioners have collected information from sources (or suspects), its veracity needs to be assessed. Here comes into play the literature on deception detection (Granhag, Vrij, & Verschuere, 2015; Vrij, 2008, 2014), which will be dealt with in Chapter 2, and the related interviewing techniques that have been developed (Vrij, 2016, 2018a; Vrij & Fisher, 2016), which will be dealt with in Chapter 3. As explained and discussed in the next chapters, verbal lie detection is more effective than nonverbal lie detection (Vrij, 2018a, 2018b). Nevertheless, evaluating truthfulness through verbal content is only possible when the interviewee is willing to talk. Hence interviewing protocols that make the interviewee talkative are paramount.

The thesis

In the present dissertation I will discuss the literature and the experiment we have conducted at the University of Bergamo concerning two main topics: Intelligence collection and deception detection. I decided to focus on these two topics as this is the research area I am very interested in, both under an applied and a theoretical point of view. Past research usually addressed each of these two topics individually. Yet, as noted by Prof. Granhag, (Granhag, in Nahari et al., 2019), there is the urgency to conduct experiments that deal with multiple objectives at the same time. Probably, it would be more fruitful to run experiments which deal with both the *collection* of information and its *veracity assessment*, as in real life the two aspects are tightly connected. Clearly, handling one goal at a time is easier, as the experimenter can exert more control over the factors at play. Nonetheless, in doing so there is the risk that research outputs will not be fed into practitioners' daily work. This makes the need of context-sensitive experiments even more relevant when the goal is to develop evidence-based protocols that aim at being used in real-life settings (Brandon et al., 2019; Caso & Palena, 2018; Fein, Lehner, & Vossekuil, 2006; Granhag et al., 2014;

High Value Detainee Interrogation Group, 2016; Loftus, 2011; Maguire & John, 1995; Swanner, Meissner, Atkinson, & Dianiska, 2016).

In the first part of the dissertation, I will discuss the literature on the collection of information in Intelligence settings (Chapter 1). This will include: i) research on undercover interviewing, that is when the interviewee is not aware that s/he is undergoing an interview; ii) research on overt interviewing, that is when the interviewee is aware of being interviewed, and iii) the role of social influence within investigative interviewing settings.

Undercover interviewing, sometime referred to as *covert* interviewing, occurs for example when undercover police officers become part of a criminal organization with the final goal to dismantle it. This becomes particularly relevant in applied settings since there are countries like Italy where criminal organizations such as Mafia are rooted in the social structure of the country and control criminal activities such as drug and human trafficking, counterfeit goods, environmental crimes, etc[2]. Undercover interviewing will be discussed in section 1.1.

In contrast, overt interviewing is usually more structured than covert interviewing and takes place with the interviewee being aware of being interviewed. One example is the collection of information concerning a potential imminent extremist terrorist attack through HUMINT activities. Recently, a form of HUMINT interviewing, the Scharff technique (Granhag, Kleinman, et al., 2016; Toliver, 1997), received much attention from academics and showed its potential for the elicitation of information. The Scharff technique will be discusses in section 1.2.

Additionally, there is research showing the potential of subtle influencing strategies on suspect's collaboration and information disclosure, which will be dealt with in the last section concerning information elicitation, that is section 1.3.

The following section of the thesis will deal with deception detection (Chapter 2). This section will include deception theories (section 2.1), cues to deception (section

[2] There are specific bodies in Italy dealing with these problems. Concerning the Italian Secret Services, the AISI deals with domestic intelligence (internal security) whereas the AISE deals with foreign intelligence (external security). Concerning police forces, Italy has the DIA (*Direzione Investigativa Antimafia*) which is an interforce body fighting organized crime such as the Mafia.

2.2), and human's lie detection accuracy (section 2.3). I will then focus on interviewing protocols aiming at detecting deception (Chapter 3), which ended up in a publication in a peer-reviewed journal (Caso & Palena, 2018).

Section 4.9 will deal with a 2-days investigative interviewing and lie detection training delivered by Prof. Letizia Caso and myself to Bergamo Police Officers. Here, strengths and limitations of interviewing techniques and the feedback provided by the trainees will be discussed. Indeed, it is important to note that police officers' feedback can inform future research.

The second part of this dissertation will deal with eight studies which have been conducted during this project. Such studies focused on the use of social influence strategies in HUMINT interrogations (section 4.2), strengths and weaknesses of the Baseline approach (sections 4.3 to 4.7), and the Theme-Selection strategy, developed at University of Bergamo (section 4.8).

Lastly, Chapter 5 focuses on general discussions, with an effort to gather all the results we have obtained from our studies and use them as a guidance for future research.

Key terms and definitions

Below is reported the definition of some key terms which the reader will find throughout this dissertation.

Interviewing/Interrogation

De Leo, Scali, and Caso (2005) make a distinction between the terms "Interrogation" and "Interviewing". They see a continuum with the two ends being "Clinical Interview" and "Interrogation", and with the term "Interview" in the middle. The three terms differ under several aspects including structuring, friendly/asymmetric approach, etc. The authors also state that the term "interview" in legal context is more flexible than "clinical interview" and "interrogation", as "interviewing" can be adjusted according to the specific need. Notwithstanding the above, I will use the two terms interchangeably throughout the dissertation as, to date, the two terms are used interchangeably in the psycho-legal literature.

Intelligence collection/Information Gathering/Educing information

With intelligence collection (or information collection or educing information) it is intended any form of activity aiming at obtaining as much information as possible from a source, regardless of the type of the source (human, digital, etc.).

HUMINT

HUMINT stands for "Human Intelligence" and is a term which defines the information that is gathered by means of interpersonal contacts. NATO defines HUMINT as "*a category of intelligence derived from information collected and provided by human sources*" (NATO, 2018). HUMINT includes espionage and counterespionage, diplomatic relations, prisoners of war interrogations etc. HUMINT can be performed both overtly and covertly.

SIGINT

SIGINT (Signal Intelligence) refers to information is collected through electronic sources, such as satellites, ships, etc.

OSINT

OSINT (Open Source Intelligence) refers to information collected via media, news, and any other available public data.

At the very end of the dissertation is the activity report, which lists all the outputs produced during this 3-years PhD programme. Below is a list on how to interpret acronyms put next to each reference that can be found in this dissertation:

- [IP] stands for “In preparation”. Studies with this acronym indicate that we are in the writing process for a specific article.
- [UR] indicates papers that have been submitted and that are under review.
- [PA] indicates published articles.
- [CP] indicates conference presentations.
- [PP] indicates poster presentations.

1 Information gathering in Intelligence settings

A differentiation between HUMINT and law enforcement (or criminal) interviewing is needed (Evans, Meissner, Brandon, Russano, & Kleinman, 2010; Redlich, 2007). Law enforcement interviews usually focus on past events, such as a crime that must be solved. One of the goals within this setting is to gather evidence that can be fed into the prosecutor's file. On the contrary, HUMINT interviewing can also focus on both the present and the future. For example, it takes place when an Intelligence agency strives to collect information to prevent an upcoming terrorist attack (Evans et al., 2010; Hartwig, Meissner, & Semel, 2014). Furthermore, criminal interviewing usually take place in a custodial setting, such as in police stations, whereas HUMINT interviewing can also takes place in unstructured settings, such as on the street, on the phone, via the internet etc. Interestingly, in criminal interrogations the interviewee is aware of being interviewed and of the purpose of the interview itself. In HUMINT settings, on the contrary, there might be cases where the source is unaware of being interviewed, or s/he might be aware that an interview is taking place without knowing the real focus of the encounter. Additionally, criminal interviews can of course be repeated, but usually last shorter than HUMINT interviews. The later can indeed take months of years in some occasions. Of course, with such a long interaction, the relationship between the interviewer(s) and the interviewee(s) become more relevant than in criminal settings (Shumate & Borum, 2006).

Notwithstanding the above-mentioned differences between the two types of interviews, criminal and HUMINT interrogations also share some key aspects. Evans et al. (2010) claim that in both a careful pre-interview preparation takes place. Also, they both aim at collecting information, although the reason for this may differ (incarceration vs. preventing a terrorist attack). Finally, both types of interviews have a post-interview component, where the gathered information is assessed (for example for veracity). Also, Redlich (2007) state that the two types of interview use psychologically-oriented techniques, but they differ in term of purposes (eliciting confessions vs. information-gathering), the degree of interviewers' trainings, and the possibility (or not) to make use of torture.

As previously mentioned, HUMINT interviewing became particularly relevant to date, due to the increase of terrorism globally (Global Terrorism Index, 2017). Unfortunately, there is a long history of unfair treatment of suspects (Senate Select Committee in Intelligence, 2014). Torture is still advocated by some practitioners, although research clearly shows that it is not effective for obtaining reliable information from the source. Vrij, Meissner, et al. (2017) report that Khalid Sheikh Mohammed has been waterboarded at least 183 times for its presumed link to the 9/11 attack. Eventually, he provided unreliable information. The authors also explain why torture, beyond being inhuman, is not effective and rather propose techniques based on psychological research.

Practitioners and academics put their efforts in developing ethic and effective, evidence-based, interviewing protocols for HUMINT interviewing (Brandon, 2014; Fallon, 2014), but there are few examinations of how different techniques perform when the aim is to collect information. Evans, Meissner, et al. (2013) compared the efficacy of several techniques that are employed with sources that are not fully cooperative (which is common in HUMINT settings) and found that the inquisitorial, information-gathering approach works better than the accusatorial approach in that the former permits to obtain more information, and makes the interviewee more talkative, than the latter. Furthermore, Evans et al. (2014) compared several approaches reported in the US Army Field Manual (Department of the Army, 2006) to the direct approach, where the interviewer asks open-ended questions in a business-like manner. The authors found that emotion-based approaches resulted in more information collection regardless of whether they were positively (reduce anxiety and put the interviewee at ease) or negatively (playing on the interviewee's fear) oriented. Yet, playing on negative emotions leaves the interviewee in a stressful condition which is unfair and ineffective when additional interrogations are expected. Indeed, if the interviewee is treated unfairly in the first interviews, then it is likely that s/he will develop a negative attitude toward the interview which will be detrimental for the following interviews.

In conclusion, employing fair interviews protocols, based on (positive) emotional approaches is an effective way to obtain cooperation from the source and, in turn,

information or admissions in HUMINT settings, whereas torture and harsh approaches are unethical and ineffective.

1.1 Undercover information-gathering

Sometimes, especially in intelligence and security settings, interrogators want the sources not to be aware of being interviewed. An example of this situation is when intelligence agencies try to discern if a suspect is part of a terrorist organisation or not. Obviously, a custodial setting could be detrimental both when the source is not a terrorist and when s/he is. In the first case, the source can panic or react with anger and the real terrorist remains free. In the latter case, if the interview is not successful there is the risk that future interrogation of the source will be unsuccessful as well as that investigations will be jeopardised, as the terrorist will develop counter-strategies to pursue his/her objectives and to avoid capture (Jundi, Vrij, Mann, Hillman, & Hope, 2015). Furthermore, when the source does not know what are the interrogator's objectives, s/he may eventually finish by providing actionable intelligence to his/her interrogator (Granhag, Kleinman, et al., 2016). It is more difficult for an interrogee to understand what the information objectives are if s/he is not aware of being interviewed. Furthermore, as Jundi et al. (2015) note, when formally interviewed, the guilty source may find ways to appear as cooperative in order not to look suspicious. On the contrary, when guilty individuals are not aware of being interviewed, they are more likely to be defensive hence less cooperative with the person they talk to. Thus, cooperativeness can be used as a proxy of possible guilt (Jundi et al., 2015). Terrorist activities usually require components of an organisation to perform reconnaissance missions to select targets for potential attacks (Soufan, 2011). Jundi et al. (2015) developed an experiment mirroring such a situation. Truth tellers where asked to take pictures to promote their city as a touristic destination. Liars, on the other hand, where asked to play the part of a mock terrorist whose aim was to plant a decoy device in the tourist location, but to cover-up their stories as they were performing the same activity as truth tellers. A mime artist acted as the undercover interviewer, who engaged the participants in what looked like a casual conversation. As predicted by the authors, truth tellers were more likely to tell the mime artist that they had taken a picture of him, and to show him such a picture when requested so, than liars. Furthermore, in a following formal interview, liars reported more security features

embedded in the pictures they had taken than truth tellers. The authors therefore conclude that interacting with a suspect through undercover interviewing may be effective in discriminating people with innocent plans from people with malintent.

Another possible application of undercover interviewing is in airport security settings. After 9/11, there has been a lot of attention to reduce the risk of terrorist attacks at airports and on flights. Billions of dollars have been spent to increase security by developing specific programmes (Ormerod & Dando, 2015; United States Government Accountability Office (US GAO), 2011). The Screening of Passengers by Observation Techniques (SPOT) is one of these, which build on behavioural observation of passengers to detect risks. However, the efficacy of such security programmes has been criticised (United States Government Accountability Office (US GAO), 2011; Weinberger, 2010). One of the reasons for its inefficacy builds on the lack of validation of behavioural cues. Therefore, alternatives have been studied. Ormerod and Dando (2015) developed a real-life, double-blind randomised control trial experiment, which was conducted in real airports. The authors compared the efficacy of two methods: The suspicions sign and a verbal veracity assessment approach. For the first one, behavioural cues of nervousness are considered. For the second method they developed, the focus was on verbal content. Also, their verbal approach alternated open questions to test questions used to assess veracity. The named their approach "Controlled Cognitive Engagement" (CCE), as security officers make use, among the others, of the different levels of cognitive engagement experienced by passengers and control the flow of the questioning in what may appear, for innocent passengers, just a friendly conversation. In order to run their experiment, the authors worked as follows. On the security agents' side, they either asked officers to conduct screening through the suspicious sign approach or the CCE. On the passengers' side, they collected participants acting as mock criminals, who were given false identities, documents, tickets, and cover-up stories. Mock criminals were also given the opportunity to prepare their cover-up story. Real passengers were the innocent passengers. Real and mock passengers both went through security screenings. When comparing officers' detection accuracy of who was lying, the results were striking: Officers in the suspicious sign only reached 3% accuracy, whereas officers in the CCE condition reached 66% accuracy (Ormerod & Dando,

2015), which is impressive considering that the experiment base rate of liars to truth tellers was 1:1000. The CCE also had a different effect on real (innocent) passengers vs. mock criminals. The number of spoken words and the number of informational elements provided to the security officers did not change for real passengers from the beginning to the end of the interview. On the contrary, the depth and quality of answers changed for mock criminals, as they showed a decrease of both indexes when comparing the initial vs. final questioning. The authors read this result in term of cognitive load experienced by truth tellers vs. liars as well as in light of truth tellers perceiving the interview as a friendly conversation vs. mock criminals realising that their honesty was being assessed. Ormerod's and Dando's (2015) study shed light on the advantages of using the CCE rather than the suspicious sign approach. One of the strengths of the CCE is that it produces unpredictable questions, which will eventually catch the liar off-guard. Future efforts should therefore focus on strategic questioning and verbal content rather than on observational techniques.

Another unobtrusive way (perhaps the most known by the general public) to detect deception and collect information is by relying on micro-expressions. Micro-expressions are thought to be fleeting facial expression of emotions, lasting no more than 1/25th of second. These expressions are expected to betray the lie and the emotion a person is trying to suppress (Ekman, 2001). Therefore, the idea is to guide the interrogation based on the micro-expressions the interviewee may show. The approach is based on the idea that a lying interviewee is submersed by negative emotions such as fear, or guilt, or by the positive emotion of fooling the interviewer (also called "duping delight", see Ekman, 2001). Yet, this approach has been strongly criticised. First, it was not possible to replicate Ekman's (2001) results. Second, scientists are not convinced of the data Ekman provided (Weinberger, 2010). Third, the application of micro-expressions as proxy for deception detection is deemed controversial, especially because of their very low incidence (about 2%) in real life scenarios (Porter & ten Brinke, 2008). Fourth, the theory on which this approach builds (the leakage theory, see section 2.1) did not receive strong support. Fifth, it builds on an arousal-based lie detection approach, according to which it is possible to detect lies because they trigger (mostly negative) emotions in the interviewees. However, truth tellers may also experience strong emotions when interviewed, for

example because of the fear of not being believed. A different approach to lie detection, building on the different cognitive processes taking place when telling the truth vs. when lying, received stronger support in the academic literature (Vrij, Fisher, & Blank, 2017). Therefore, it is better to drop the use of micro-expressions when it comes to forensic and intelligence interviewing.

In conclusion, undercover interviewing has potential in intelligence and security settings, but it is not applicable in criminal settings, when formal interviewing takes place. Its strengths are rooted on circumventing (or at least reducing) the efficacy of counter-interrogation strategies that the interviewee may adopt, as well as on the fact that such techniques are sometimes unpredictable, with a clear advantage for the investigator (Ormerod & Dando, 2015). Also, there are some situations where undercover interviewing is essential before any possible formal investigation can take place, such as when undercover policemen infiltrate criminal organisations. Unfortunately, though, undercover interviewing strategies did not receive the same attention in the academic literature as formal interviewing. Hence, future research on this line is needed.

1.2 **The Scharff Technique**

"What did he get out of me? There is no doubt in my mind he did extract something, but I haven't the slightest idea what". (Colonel Zemke, interrogated by Hans Scharff)

The Scharff Technique takes its name from Hans-Joachim Scharff, an WWII interrogator of the German Luftwaffe. The technique came to the attention of academics after the pioneering work by Prof. Granhag's Lab at Gothenburg University. The professor and his team worked to explore what strategies and tactics employed by Scharff were effective in eliciting information from prisoners of war (POWs). As noted elsewhere, there is a lack of empirical research on techniques on gathering human intelligence (Dujmovic, 2005; Varouhakis, 2013). Therefore, the work on the Scharff technique fills this gap and comes as an approach to evidence-based interviewing in HUMINT settings. This is relevant for applied, ethical, and academic reasons. Research, indeed, showed that harsh interrogation with coercion

(regardless of whether it is physical or psychological) is totally ineffective for the elicitation of reliable intelligence (Vrij, Meissner, et al., 2017).

As reported in Granhag, Kleinman, et al. (2016), part of the efficacy of this technique lies on Scharff's understanding of the counter-interrogation strategies employed by POWs (see also Toliver, 1997) which, according to the interrogator, can be described as:

1. I will not tell much
2. I will try to figure out what the interrogator's objectives are, and I will not reveal them
3. It is useless to withhold information that is already known to the interrogator

As we will see later in this paragraph, the Scharff technique hinges in particular on the third strategy. There are also other reasons why Scharff was an effective interrogator. First, he did not use coercive strategies, such as torture. Second, he carefully planned each single interview before it took place. In particular, he collected all the available information he could concerning the specific POW and used it strategically. Third, rather than directly questioning the sources, Scharff initiated his interviews in an informal way, and gave to the source the opportunity to add, confirm, or disconfirm information he (Scharff) presented. Starting from a biographical book on Scharff (Toliver, 1997), Granhag, Kleinman, et al. (2016) concluded that the strategies used by Schaff briefly outlined above can be grouped into the following interrelated tactics:

1. Adopting a friendly approach. As said above, Scharff was never coercive and approached POWs in a friendly manner. He also created rapport by offering his sources some benefits, such as a cigarette or the opportunity to have a walk in a nearby park.
2. Giving the illusion of already knowing it all. This relates to the third strategy used by the POWs reported above. Indeed, by misleading POWs into believing that he already knew a relevant amount of information, Scharff led the interviewee provide information that he (Scharff) did not know.
3. Use of confirmations/disconfirmations. Rather than asking direct questions, Scharff presented specific claims- every now and then- which could be

(dis)confirmed by the source. This approach is twofold. First, by avoiding direct questions, Scharff managed to mask his information objectives. Second, sources were less reluctant to (dis)confirm questions compared to answer specific questions. Likely, this happens because by (dis)confirming a claim verbalised by Scharff, the source avoided the responsibility of having revealed that specific piece of information himself/herself.

4. Do not press for information. Avoiding direct questions also works to reduce the pressure on the source, whose autonomy is encouraged. Autonomy and respect of the source is particularly relevant in interviewing settings (Miller & Rollnick, 2012).
5. Ignore new information. By ignoring any new piece of information revealed by the source, or by downplaying its importance, Scharff managed to mask his information objectives.

Granhag, Kleinman, et al. (2016) developed the above tactics starting from a theoretical background and building on their experimental results. In such experiments, participants usually take part in a study mirroring real-life HUMINT interactions. Interviewees usually play the role of a source holding information about a terrorist act, and interrogators have some intelligence (for example on the attack itself) as well as information gaps. Hence, a situation where both sides (the interviewer and the interviewee) want to obtain something (the interrogator strives for information, whilst the interviewee tries to strike a balance between disclosure and withholding) takes place. When it comes to interviewing the suspects, interviewees are treated friendly, and the interrogator starts the interview by providing some known information, to establish the illusion of knowing it all. After this is done, interviewees are given the opportunity to add information. Then, the interviewer employs the confirmation/disconfirmation tactic. In the first case, s/he presents claims concerning information which are believed to be more probable, which the interviewee can confirm. In the second case, the interviewer presents information that is believed to be less probable, which the interviewee can disconfirm. Once the confirmation/disconfirmation session is over, participants are usually given the possibility to add any additional information, and then the interview ends.

Considering that the Department of the Army (2006) encourages the use of the Direct Approach (DA) as one of the preferred ways to gather intelligence, most of academic work is based on the comparison of the Scharff technique to the DA. The DA is based on a business-like interaction, where direct, open questions are asked to the sources. The review by Granhag, Kleinman, et al. (2016) concludes that the Scharff technique outperforms the DA in several aspects. First, more information is collected with the Scharff technique than with the DA (see for example Oleszkiewicz, 2016; Oleszkiewicz, Granhag, & Cancino Montecinos, 2014). Second, interviewees find it more difficult to understand what are the interviewer's objectives when the Scharff technique is used compared to when the DA is used (Granhag, Montecinos, & Oleszkiewicz, 2015; Granhag, Oleszkiewicz, Strömwall, & Kleinman, 2015; Oleszkiewicz, Granhag, & Kleinman, 2014). Third, sources interviewed with the Scharff technique underestimate, whereas sources interviewed via the DA overestimate, the amount of revealed information (Granhag, Kleinman, et al., 2016; Granhag, Montecinos, et al., 2015). Fourth, when the Scharff technique is employed, sources believe that the interviewer had more knowledge prior to the interview than when the DA is used (Granhag, Kleinman, et al., 2016). Fifth, the Scharff technique showed to be particularly effective also with less cooperative sources (Granhag, Oleszkiewicz, et al., 2015).

Subsequent work on the Scharff technique focused on refining its tactics. A series of experiment showed for example that confirmation tactics are better equipped at masking interviewer's objectives than disconfirmation tactics, as well as that the former increases sources' perception of interviewer knowledge compared to the latter (May, Granhag, & Oleszkiewicz, 2014). The Schaff technique outperforms the DA also when the source underwent repeated interviews (Oleszkiewicz, Granhag, & Kleinman, 2017b), although when interrogating multiple sources the two techniques elicited a similar amount of information (Granhag, Oleszkiewicz, & Kleinman, 2016).

In conclusion, the research outlined above shows the potential of the Scharff technique in HUMINT interviewing. Yet, there is room for improving the technique. Future research should focus on strategies to increase the illusion of knowing it all and on its strategic use. Also, other techniques such as social influence tactics,

which is the focus of one of the experiments presented in this dissertation, could be paired with the approach.

1.3 Rapport and Social Influence

A shared definition of rapport is missing. However, a useful one derives from Tickle-Degnen and Rosenthal (1990), who define rapport as including mutual attention between people (i.e.: being in sync), positivity (i.e.: positive affect between people), and coordination. Positivity, then, can itself be described from different perspectives. In particular, Cuddy, Fiske, and Glick (2008) and Fiske, Cuddy, and Glick (2007) define positivity in terms of social judgment. Abbe and Brandon (2013) instead, underscore the role of warmth and competence[3]. Abbe's team worked on rapport and its role within investigative and intelligence interviewing as they believe that its importance (and effectiveness) has usually been underestimated in the academic literature. According to the authors' view, rapport works in intelligence interviews because it triggers several positive processes: It improves working alliance[4] between the interrogator and the interviewee, it increases information disclosure and it can be a proxy for social influence (Abbe & Brandon, 2013). But the role of rapport is also important to increase cooperation (Bull & Soukara, 2010; Department of the Army, 2006), which in turn increases information elicitation, a key aspect in intelligence interviewing. Summarising, the effect of rapport on educing information can be both direct and mediated, and its importance is recognised by both academics and practitioners (Soufan, 2011).

Notwithstanding the lack of a shared definition of rapport, and perhaps a lack of a solid theory- especially within investigative interviews- experimental results are quite impressive. Alison and colleagues have developed a coding framework focusing on three elements tightly related to rapport, the Observing Rapport-Based Interpersonal Techniques (Alison, Alison, Elntib, & Noone, 2010, 2012). Their tool also builds on Motivational Interviewing (Miller & Rollnick, 2012) and Interpersonal Behaviour Circles (adopted from Interpersonal theories, Leary, 1955), which focuses on the

[3] However, it is important to note that the Fiske et al. papers focus on warmth and competence in social cognition in general, whereas Abbe's and Brandon's focus is in the investigative interviewing setting.
[4] Working alliance can be seen as a situation whereby the interviewer and the interviewee recognise what is the goal of the interview taking place and work together to reach such goal.

interaction between two orthogonal dimensions (challenging vs. cooperative and authoritative vs. passive) when looking at interviewer-interviewee interactions. The output measure is the elicitation of useful information (i.e.: actionable intelligence). In a recent paper, Alison and colleagues observed and coded the behaviours of 58 police interrogators who were interviewing terrorist suspects using the ORBIT tool (Alison, Alison, Noone, Elntib, & Christiansen, 2013) and found that motivational interviewing increased information elicitation through an increased employment of adaptive behaviours shown by the interviewer. The authors' work therefore credited the effectiveness of rapport-based interviews within real life contexts. Kelly, Miller, and Redlich (2016) reached a similar conclusion: They explored the effect of rapport and confrontation on interviewees' cooperation and found that the former increases cooperation, whereas the latter has detrimental effects which can last up to 15 minutes.

Although rapport showed its potential in investigative interviews, it must be borne in mind that it can be established in several ways. Hence, future studies should clearly define how rapport is established and maintained (Alison et al., 2012; Tickle-Degnen & Rosenthal, 1990). According to Goodman-Delahunty and Howes (2016), liking and reciprocity are the most used tactics to develop rapport with the interviewee. Reciprocity is one of the six principles of social influence outlined by Cialdini (2001): Scarcity, Reciprocity, Commitment/Consistency, Social proof, Liking and Authority. Reciprocity- whose efficacy when implemented in Scharff interviewing has been explored in one of the work reported in this dissertation, see section 4.2)- has been defined by Cialdini and Goldstein (2004) as *"[…] the rule that obliges us to repay others for what we have received from them"* (p. 599). In short, this principle works as follows: If we receive something from someone, then we offer them something back. Reciprocity, as rapport, can take several forms. According to Sobel (2005), there are two types of reciprocity: Intrinsic and Instrumental. In the first case, reciprocity is triggered by the perceived intention of the other; Instrumental reciprocity, on the other hand, originates by the perceived future gain we can obtain by reciprocating the action of the other (see also Cabral, Ozbay, & Schotter, 2014).

The positive effect of reciprocity as a form of social influence has been shown in different settings (Cialdini, 2001), such as relationships (Kelln & Ellard, 1999), the amount of tip handed to restaurant employees (Rind & Strohmetz, 1999) and

investigative interviewing (Vrij, Leal, Fisher, et al., 2018). Relevant to this dissertation is the work by Matsumoto and Hwang (2018), who found that a very simple form of reciprocity was effective. The authors found that lying participants who had been offered water revealed more relevant details than liars who were not offered water. Additionally, the authors found that the effect was present across different cultures, which is consistent with Gouldner's (1960) suggestions. Interestingly, the positive effect of reciprocity found by Matsumoto and Hwang (2018) was both direct and mediated by rapport. This is a key aspect in that, as detailed above, the efficacy of social influence tactics in intelligence interviewing is two-fold: direct (it elicits more information) and mediated (it increases rapport which, in turn, elicits more information). Furthermore, its simplicity of use and the fact that it seems not to be culture-dependent, plays positively in intelligence interviewing when considering that terrorism is increasing globally (Global Terrorism Index, 2017; Vrij, Meissner, et al., 2017). In short, rapport and social influence can be used potentially with any source.

Notwithstanding the previous paragraphs on the effectiveness of rapport and social influence when looking ad information elicitation, both academics and practitioners must deal with a very important issue. The first step of an interview is to obtain information from the source. Yet, some of the information elicited might be unreliable or fabricated. This is particularly relevant when considering that we all, as humans, lie (Vrij, 2008). Therefore, once information is gathered its veracity must be assessed. As shown in the following chapters on deception detection, there are several approaches that increase differences in (non)verbal behaviours of liars vs. truth tellers and that permit higher lie detection accuracy rates.

2 Deception detection

Deception detection is the act of uncovering lies. A first step to reach this goal is to define deception. Several definitions have been given. Mitchell (1986) defines deception as *"[…] a false communication that tends to benefit the communicator[…]"*, whereas Burgoon and Buller (1994) state that deception is *"[…] a deliberate act perpetrated by a sender to engender in a receiver beliefs contrary to what the sender believes is true to put the receiver at a disadvantage […]"*. Regardless of the definition we pick, deception involves, among the others, masking true while fabricating false information (or plans, emotions, and opinions). In short, deception entails information management (Hartwig, Granhag, Strömwall, & Doering, 2010; McCornack, Morrison, Paik, Wisner, & Zhu, 2014). However, people also try to give the impression of honesty- both when lying and telling the truth (Hartwig et al., 2010). Hence, one of the drives behind lie detection studies is the search for cues to deception which can derive from impression (mainly nonverbal and paraverbal cues) and information (mainly verbal cues) management. Several theories have been developed as to why and how we lie, which will be dealt with in the next section (2.1). Cues to deception are discussed in section 2.2. Additionally, due to the result that people's lie detection accuracy is usually poor (section 2.3), academics developed several interviewing techniques which aim at magnifying differences between truth telling vs. lying (Chapter 3). As we will see in the following chapters, credibility assessments can be improved with such specific interviewing techniques.

2.1 Deception theories

Perhaps one of the firsts deception theories developed in academia is the Leakage Theory by Ekman and Friesen (1969). This theory predicts that, when lying, people experience high emotional arousal which they try to suppress, often without success. According to Ekman and Friesen (1969), deception thus leaks from various communication channels, but mainly through nonverbal behaviour and facial expressions. The authors also differentiate between clues to deception- indicators of an ongoing lie with no information concerning what is being concealed- and leakage, where the truth "leaks out". Hence, in the latter case the observer has also access to the information being conceived or falsified. As an example of leakage, Ekman (2001)

describes the case of a woman participating in a study who had held the "middle finger" for a long time during an interview, which the author interprets as a leakage of the emotion she felt toward the interviewer or the interviewing setting. It has also been suggested that the emotional arousal is mainly negative, with liars experiencing fear, shame, guilt and disgust, although liars can also feel happy when they perceive they are fooling the receiver (Ekman, 2001; Ekman & Friesen, 1969). According to this theory, the observer should also consider the "sending capacity" of different parts of the body. Indeed, according to Ekman and Friesen (1969), since people receive a great amount of (also neuroanatomical) feedback about their facial expressions, it is unlikely that the truth comes out from this part of the body. It is more likely that it does so through other parts of the body, such as the feet, from which people receive less feedback. The only exception of leakage through the face is the case of fleeting micro-expressions (Ekman, 2001), although the section on undercover interviewing highlighted that these are deemed to be unreliable (Porter & ten Brinke, 2008).

The leakage theory is fascinating, and it has also inspired popular books and TV fictions such as Lie to Me. However, it is deemed to be controversial (Bond, Levine, & Hartwig, 2015) and scientific data has not supported it (Weinberger, 2010). One of the main issues is that it builds on the idea that only liars face such a heightened emotional arousal. Yet, it is well known that truth tellers also do (Vrij, 2008). In short, liars can fear to be caught, truth tellers can be afraid of not being believed.

Another contribution comes from Zuckerman and colleagues who implemented aspects other than emotions, such as arousal, cognitive load and attempted control, into their Four-Factor Theory (Zuckerman, DePaulo, & Rosenthal, 1981). Perhaps the most important contribution of this theory is that, rather than predicting a direct link between deception and (non)verbal behaviours they claim that the link is between deception and the four factors reported above. These, in turn, can be mirrored in different (non)verbal behaviours when comparing truth tellers and liars. For example, people feel emotions such as guilt when lying, which in turn can trigger the expression of negative emotions. This is in part consistent with Ekman and Friesen's (1969) leakage theory. Deception is also expected to increase arousal, which can be expressed through heightened psychophysiological activation (e.g.: increased heart rate) and to put demand on cognitive load (Vrij, 2015a) as, when lying, people face

several tasks at the same time (e.g.: suppressing the truth, creating a credible story, checking if the other person is suspicious, etc.). Lastly, the theory predicts that people attempt to control their behaviour in order not to appear as deceitful, and this can be expressed in a rigid posture and few body movements. Although this theory has provided an important contribution for the theory of deception, it also failed to receive strong support (Bond et al., 2015).

Another relevant theory of deception is the attempted behaviour control which, as already introduced by Zuckerman et al. (1981), predicts that liars try to control their behaviour to appear credible. This makes sense since people know that the others may be assessing his/her honesty while speaking (Buller & Burgoon, 1996; Burgoon & Buller, 1994). In short, when people lie, they try to avoid behaviours believed to be associated with deception (such as lack of eye contact, speech errors, etc.) and try to give an honest impression (Vrij, 2008). However, for several reasons, this is not easy. We are rarely aware of our own behaviour and, even if we do, we may fail to control ourselves. As predicted by the leakage theory outlined above (Ekman & Friesen, 1969) for example, we fail to control our behaviour because of the scarce feedback we receive. Or, we may find it difficult to control psychophysiological reactions (such as heart rate and respiration), regardless of how much effort we put in it[5] (Vrij, 2008). Lastly, the difficulty in controlling our behaviour also relates to practice. We usually give more importance to the words we use rather than nonverbal behaviour or voice[6] (Vrij, 2008). This idea of us finding it more difficult to control our nonverbal behaviour than our words has also been supported in experimental results. Caso, Vrij, Mann, and De Leo (2006) for example, informed their participants about cues to deception and found that they were able to adapt their verbal behaviour but not their nonverbal behaviour. The authors therefore conclude that verbal countermeasures are easier to apply than nonverbal countermeasure, which is consistent with the predictions outlined above.

[5] Measurement of psychophysiological reactions is the base for veracity assessments through the polygraph. However, as this thesis focuses on lie detection without artificial tools, the interested reader is invited to read other work such as Vrij (2008) and Rosenfeld (2018).

[6] As an interesting side note, voice pitch, which we find so difficult to control because it is influenced by the activation of our autonomic system (Frank, Maroulis, & Griffin, 2013), is one of the few cues which showed to be associated with deception/truth telling in DePaulo's and colleagues (2003) meta-analysis on cue to deception, with a Cohen's *d* of .21.

Two last theories are worth mentioning in this context. The cognitive effort approach and the self-presentational theory. The former predicts that lying is, in most situations, more mentally taxing that telling the truth (Vrij, 2008, 2015a). Indeed, liars do not take their credibility for granted whilst truth tellers often do (DePaulo et al., 2003; Gilovich, Savitsky, & Medvec, 1998), must monitor interviewer's reactions and suspiciousness (Buller & Burgoon, 1996; Burgoon & Buller, 1994), must suppress the truth and have to create a credible and plausible story (Vrij, 2008; Vrij, Edward, & Bull, 2001). All this effort, according to the theory, is reflected in cues of cognitive load such as "thinking hard" (Vrij et al., 2001). A recent meta-analysis also supported the cognitive load approach in that it permits higher accuracy rates of observers compared to a standard approach, where the cognitive load on the interviewee is not manipulated by the experimenter (Vrij, Fisher, et al., 2017).

The self-presentational approach predicts similarities, rather than differences, between truth tellers and liars (DePaulo et al., 2003). Experimental finding also support this assumption (Caso, Gnisci, Vrij, & Mann, 2005). Truth tellers also experience strong emotions while interrogated, as well as they try to give an honest impression as liars do, although for obvious different reasons. DePaulo et al. (2003) suggest that the relevant difference between honest and deceptive responses is that truth tellers' claim of honesty, as reported also by Vrij (2008), is legitimate, whereas that of liars is not. Consequently, according to DePaulo et al. (2003), liars' self-presentations are less convincingly embraced than those of truth tellers.

2.2 Cues to deception

The theories of deception reported above are the starting point for the search for cues to deception. In short, it is first needed to develop a theory as to why lying and truth telling is different. Then, building on theory, research can explore if differences in verbal, nonverbal, and paraverbal behaviours between truth telling and lie appear. Interestingly to note, different theories (e.g.: emotional vs. cognitive approach) may predict differences outcomes (e.g.: increase vs. decrease) of the frequency and types of cues to deception. The next sections will deal with research findings concerning nonverbal/paraverbal and verbal cues to deception.

2.2.1 Nonverbal and paraverbal cues

Nonverbal cues to deception are perhaps those that received greater publicity due to their popularity in non-academic books and TV fictions such as Lie to Me. There is also a great deal of attention in seminar and classrooms offers with outstanding titles such as "You will never be lied again", that are rarely (if ever) subjected to scientific test. Unfortunately, as Vrij (2008) notes, there is no Pinocchio's nose (cues directly and reliably always associated with deception). Further, the work by DePaulo and colleagues (2003) underlined that cues to deception are unreliable and also faint. Moreover, as Luke (2019) points out, research findings on cues to deception are not only weak but, probably, also exaggerated or false positives. In short, although there are some cues that appear to be related to deception, the difference between truth tellers and liars in such cues are small or medium in the best case. This also holds for verbal cues, focus of the next session. Table 1[7] shows: the specific nonverbal cue to deception, whether people believe it to be associated with deception or not, and findings of the meta-analysis by DePaulo et al. (2003) about the specific cue.

Table 1. *Subjective and objective nonverbal and paraverbal cues to deception. The sign > indicates an increase during lying, whereas the sign < indicate a decrease. The sign – indicates no relationship with lying.*

Cue	People's beliefs	Actual association with deception
Gaze	<	-
Self-adaptors	>	-
Illustrators	-	<
Hands/fingers hands movements	>	<
Leg and foot	>	-
Eye blink	>	-
Hesitations	>	-
High-pitch voice	>	>
Speech rate	-	-
Latency	-	-

It interesting to note that, according to Table 1, people's belief is coherent with research findings only for one cue: An increase of high-pitch voice while lying. This (misidentifying what cues are related to deception and how) together with the use of

[7] Built starting from Vrij's (2008) rework of DePaulo's et al. meta-analysis (2003).

stereotypes, may partially explain why people (laypeople and professional alike) are poor lie detectors (Bond Jr. & DePaulo, 2006, 2008), focus of section 2.3. The results obtained by the Global Deception Research Team (2006) support the assumption that people hold wrong beliefs about valid cues to deception. On the contrary, however, a recent work by Hartwig and Bond Jr (2011) found that implicit knowledge of cues to deception is actually correct. So, why are people poor lie detectors? There are several possible explanations for the effect. First, cues to deception are faint and unreliable. Second, different interviewing approaches can influence the expression of cues to deception (Gnisci, Caso, & Vrij, 2010; Vrij, Mann, & Fisher, 2006). Third, interpersonal differences influence cues to deception: A person may move *more* while lying, whilst another one may move *less*. On this basis hinges the Baseline approach, which will be dealt with in section 3.4. Fourth, interpersonal difference, the medium of communication (e.g.: written vs. in vivo) and personal characteristics of the sender, all play a joint role that influences observers' accuracy rates (Caso, Maricchiolo, Livi, Vrij, & Palena, 2018). Fifth, different theories can predict different outcomes concerning the same exact cue. For example, the emotional approach would predict that people move *more* whilst lying because the feel nervous; the cognitive effort approach as well as the attempted control approach would predict that people move *less* whilst lying, either as consequence of the former or the latter. Last, a shared and consistent approach to code nonverbal behaviour is missing. Hence, it is difficult to draw conclusions as far as different work code nonverbal behaviour differently. Hands movements for example are expected to decrease (or increase, according to the theory from which we start) when just counting the number of such movements. However, with a more detailed coding, results can change. Caso, Maricchiolo, Bonaiuto, Vrij, and Mann (2006) differentiated between deictic and metaphoric hand gestures and found that, when lying, the former decreased whereas the latter increased.

This all sheds light on the difficulty of deception detection. Verbal behaviour is not exempt from issues, yet it is a better proxy for veracity assessments (Vrij, 2008).

2.2.2 Verbal cues

Verbal cues to deception concern the specific words or types of details (quantitative) used, although it might sometime focus on holistic and qualitative features. Expected

differences between truth telling and lying in verbal behaviour are explicable under different perspectives.

The Undeutsch hypothesis (Steller, 1989; Undeutsch, 1967) predicts that events that have been really experienced are qualitatively different from event that have not (for example imagined events or fantasies). Starting from this hypothesis, a tool, the Statement Validity Assessment (SVA), was developed (see Vrij, 2008; 2015b for an exhaustive description). The SVA includes a checklist, named Criteria Based Content Analysis (CBCA) based on 19 criteria, such as logical structure, amount of detail, contextual embedding, etc. (see for example Vrij, 2015b for the full list). The more these criteria are included in the story, the more credible the story is. CBCA has shown to be useful to discriminate between truthful and deceptive statements, but with a classification accuracy rarely exceeding 70% (Vrij, 2008). Nonetheless, it is one of the most frequently used verbal tool for credibility assessment, and it is also accepted as evidence in trials in some countries such as North American courts and in some European countries (Köhnken, 2004). However, though, the CBCA lacks a shared background theory.

Another verbal assessment tool builds on a solid theoretical background. The Reality Monitoring (RM, see Johnson & Raye, 1981) was not developed with the aim of credibility assessment in mind. Rather, it was created to discriminate between real and imagined events. Nonetheless, perhaps because of its solid theoretical background, it received attention from scholars, and it has been applied to lie detection contexts as well. It is also built on several criteria- eight according to Vrij (2008)- such as clarity and sensory information. Seven out of eight criteria included in the RM are expected to be more present in real than in fabricated stories. The RM obtained accuracy rates similar to the CBCA, but it is much easier to understand and use. Further research showed that if people are informed about CBCA criteria, this tool becomes ineffective as truth tellers and liars appear equally credible; further, the RM can be used as an alternative tool in such situations (Vrij, Akehurst, Soukara, & Bull, 2004). A question remains open as to how resistant to countermeasure RM is.

Very important to note, both tools are influenced by several variables, such as age, status and social skills of the respondent (Vrij, Akehurst, Soukara, & Bull, 2002). Therefore, these tools must be used with caution.

Recently, scholars started to evaluate new types of verbal cues. An example is the use of verifiable details. According to Nahari, Vrij, and Fisher (2014), verifiable details are details that can be checked by the investigators. Their coding system therefore include details that are: i) documented (e.g.: a bank transaction); ii) carried out with some other identifiable person; iii) witnessed by an identifiable person. When comparing the efficacy of the Verifiability Approach (VA) to coding of statements based on RM where details were not coded for their verifiability, the former outperformed the latter. In particular, the VA resulted in classification rate of truth tellers/liars of about 79%, whereas the classification with coding that did not accounted for verifiability of details resulted in a classification rate of about 63%. As the authors themselves underline, the verifiability approach is effective also because it puts the liar in a difficult situation: S/he can either be vague, which would result suspicious, or tell a great amount of details, which are expected not to be verifiable (hence, the lie is at risk of being discovered).

Another recent approach is the use of the ratio Complications/(Complications + Common Knowledge + Self-Handicapping strategies). According to Vrij and colleagues, a complication is "an occurrence that makes a situation more difficult than necessary" (Vrij, Leal, Jupe, & Harvey, 2018, p. 2). Common knowledge is instead a piece of information that is highly stereotypical, whereas self-handicapping strategies occur when the interviewee gives justifications as to why s/he cannot provide information. The authors found that truth tellers reported a higher proportion of complication than liars, showing the potential benefit of using this type of coding.

There are other verbal lie detection tools, but some of them are tightly connected to a specific interviewing strategy, which will be the focus of Chapter 3.

2.3 **Humans' lie detection skills**

Human's lie detection skills are, unfortunately, poor. In one of the first review on laypeople's lie detection abilities, Kraut (1980) reported a range of 45%-60%, with a mean accuracy of 57%. Cheng and Broadhurst (2005) report slightly better results,

with a mean accuracy of 67% for truth detection and of 70% for lie detection. However, when summarising the result of 79 studies on laypeople's lie detection accuracy when evaluating strangers, Vrij (2008) reported a mean accuracy rate of about 54%, with a better performance for truth detection (about 63%) than for lie detection (about 48%). A picture which is not encouraging.

One may wonder whether such a poor figure is due to observers evaluating strangers rather than friends and romantic partners. However, this is not the case. Vrij (2008) reports that even when evaluating close friends and romantic partners, accuracy rates remain in the 49%-59% range. Also, several studies directly compared observers' accuracy rates when evaluating stranger and not strangers, but did not find any significant difference (Anderson, DePaulo, & Ansfield, 2002; Buller, Strzyzewski, & Hunsaker, 1991). Yet, some studies have found a positive effect when the baseline behaviour of the sender was known to the observer (Brandt, Miller, & Hocking, 1980a, 1980b; Caso, Palena, Vrij, & Gnisci, 2019; Feeley, deTurck, & Young, 1995), but this will be dealt with in section 3.4.

Lastly, a meta-analysis reached conclusions similar to Vrij (2008): People are poor lie detectors, with an average truth accuracy of 61%, average lie accuracy of 47% and average total accuracy of 54%, and professional lie catchers (e.g.: police officers) do not perform better than laypeople (Bond Jr. & DePaulo, 2006). Also, judges' lie detection ability (across individuals, i.e.: between judges) does not vary much (Bond Jr. & DePaulo, 2008).

One possible explanation for people's low lie detection accuracy is the lack of knowledge on how to evaluate senders. Therefore, a possible solution may be to train interviewers on how to detect lies. Although there is indeed an improvement of lie detection accuracy after specific trainings, this remains at best "moderate", and the positive effect of training is larger when it focuses on verbal content rather than on a combination of verbal content and nonverbal behaviour or on nonverbal behaviour only (Hauch, Sporer, Michael, & Meissner, 2016).

Trainings, however, may be improved by also teaching the interviewer on how to conduct the interview, rather than just on "cues to deception". Vrij, Leal, Mann, Vernham, and Brankaert (2015) trained 27 police officers on the Cognitive Credibility

Assessment (CCA) and found that total accuracy increased from 59% to 74% with a training that only lasted few hours. Compared to the 54% that would be expected by previous meta-analyses on observers' lie detection accuracy rates, (Bond Jr. & DePaulo, 2006), Vrij's et al. (2015) results are encouraging. Indeed, due to the small, faint and unreliable differences in cues to deception between truth tellers and liars (DePaulo et al., 2003), scholars have suggested that researchers' efforts should focus on developing effective interviewing strategies which can enhance such small differences (Vrij, 2014; Vrij & Granhag, 2012). Such efforts are the focus of the next chapter.

3 Interviewing protocols

This chapter will focus on recent developments within the investigative interviewing research area and will present four techniques[8]

3.1 Strategic Use of Evidence (SUE)

The Strategic Use of Evidence was introduced by Granhag (2010). Its framework is based on two core aspects: A strategical level and a tactical level. The strategic level is focused on the interviewee and his/her counter-interrogation strategies (Granhag & Hartwig, 2015). In particular, it is based on the idea that the interviewer strategizes on the basis of several key aspects, such as:

1. The interviewee's perception of interviewer's knowledge about the crime
2. Counter-interrogation strategies employed by the interviewee to deal with the interviewer's strategies
3. Verbal responses given by the interviewee
4. Changes in perspective-taking processes throughout the interview

The strategical level is crime-independent in that the approach predicts that interviewees will act similarly regardless of the crime committed: The interviewee tries to deny involvement and to give vague answers. Consequently, his/her statements appear as not forthcoming.

The tactic level is more based on the interviewer's side and includes the following aspects:

1. The collection of how much information about the crime before the interview as possible
2. The type of questions asked to the interviewee
3. The presentation of evidence

The main idea behind the approach is to avoid presenting the evidence at the beginning of the interview. Indeed, in doing so, the interviewer gives the interviewee the possibility to create plausible false stories about the evidence. For example, if the

[8] The review of the available literature on interviewing to detect deception led to the publication of a scientific paper (Caso & Palena, 2018)

interviewer says to the interviewee that he has been seen on the location where the crime has been committed, he will very likely find a justification for such a piece of evidence. On the contrary, if the interviewer initially withholds such knowledge, and exhaust all possible justification before presenting evidence to the interviewee, then the interviewee will find it difficult to explain why he just denied to have been at that location. Granhag and Hartwig (2015) report that by using this approach the interviewee shows: i) statement-evidence inconsistencies (that is, an incongruence with what the interviewer presents as evidence later in the interview); ii) within-statement inconsistencies (that is, inconsistencies between information provided during the interview); iii) change of forthcomingness. The first two are interesting for lie detection purposes (denying having been in a specific place when there is hard evidence of that is a clear sign of lying); the latter in also interesting under a theoretical viewpoint. Indeed, what happens with the SUE is that the interviewee, after the interviewer has showed evidence strategically, starts to think that the interviewer holds much more evidence that s/he (the interviewee) believed at the beginning of the interview. As a consequence, the interviewee may end-up providing new details to the interviewer (Tekin, Granhag, Strömwall, & Vrij, 2016). Research shows that training practitioners to the use of the SUE results in high lie detection accuracy rates, with peaks as high as 85% (Hartwig, Granhag, & Luke, 2014; Hartwig, Granhag, Strömwall, & Kronkvist, 2006). Unfortunately, however, the technique is useful only when the investigators do have evidence which has not been divulgated in media.

3.2 **Assessment Criteria Indicative of Deception**

The Assessment Criteria Indicative of Deception (ACID) is a semi-structured interview firstly conceived by K. Colwell, Hiscock, and Memon (2002). It was developed starting from the Cognitive Interview (CI) (Fisher & Geiselman, 1992) and the Reality Monitoring (RM) (Johnson & Raye, 1981). The CI was developed to aid cooperative interviewees to remember details and facts of a past event. Hence, it hinges on research on memory, cognition, and social processes (Fisher & Geiselman, 1992). The CI includes several strategies that help the interviewee to recall past events, such as: recall the event in different chronological orders, reinstate the context of the event (smells, sounds, etc.), recall the event from a different perspective (e.g.: third person), and a "report everything" strategy. The main idea of the ACID is that,

by applying techniques taken from the CI and by asking unexpected questions, a second free recall provided by liars will be different from that provided by truth tellers (K. Colwell, Hiscock-Anisman, & Fede, 2013). Therefore, the ACID starts with rapport building. Then, the interviewee is invited to provide a free recall of the event under investigation, after which mnemonic techniques are applied and forced-choice questions asked. Eventually, the interviewee is invited to provide a new free recall. The core part of the ACID (mnemonics and unexpected questions) is expected to elicit differences between truth tellers and liars, which the authors label Differential Recall Enhancement (DRE) (K. Colwell et al., 2013). Indeed, if the truth teller will benefit from the mnemonic techniques, the liar will not, as the latter will likely try to keep the story simple and to avoid adding new details (Vrij, 2008). Research showed that by applying the ACID, truth tellers add more details in the second free recall than liars, speak more words, and show a lower type-token ratio[9] (K. Colwell et al., 2013; K. Colwell, Hiscock-Anisman, Memon, Rachel, & Colwell, 2007; K. Colwell, Hiscock-Anisman, Memon, Taylor, & Prewett, 2007; K. Colwell et al., 2002). Furthermore, the technique showed to be effective with different languages and with interviewees speaking in a non-native language (K. Colwell et al., 2013). Additionally, the ACID showed to increase observers' accuracy up to 70% even with a training as short as one lasting two-hours (L. H. Colwell et al., 2012). Research therefore showed the potential of the ACID technique for investigative interviewing. The issue here is the technique only works with cooperative suspects.

3.3 **Cognitive Credibility Assessment**

The Cognitive Credibility Assessment was developed starting from the idea that lying can be more mentally taxing than truth telling (Christ, Van Essen, Watson, Brubaker, & McDermott, 2009) as reported in section 2.1. Research showed that the cognitive approach is indeed effective for discriminating truth tellers from liars (for a complete overview, see Vrij, 2015a). A recent meta-analysis showed that the cognitive approach brings to higher accuracy rates than standard approaches (Vrij, Fisher, et al., 2017). The CCA is still under development, but a series of tactics and protocols have

[9] The type-token ration is a measure of the complexity of the statement provided ad is obtained by dividing the occurrence of unique words (words that are not repeated) by the total number of words.

been developed, such as: i) imposing cognitive load; ii) encouraging the interviewee to tell more; iii) asking unexpected questions.

In the imposing cognitive load technique, the interviewer makes the interview cognitively demanding. Although both truth tellers and liars are expected to suffer from this enhanced cognitive strain, liars are expected to be more subjected to it. Indeed, part of liars' cognitive resources are allocated to a series of tasks, such as controlling the interviewer's reactions, suppressing the truth, creating a credible story, avoid inconsistencies with previously said details, etc. (Vrij, 2015a). Consequently, liars will have less resources left over than truth tellers. Cognitive load can be increased in several ways. For example, the interviewer can require the interviewee to perform a second task while answering to questions (Debey, Verschuere, & Crombez, 2012). The expected result is that liars will face a decrease of performance in one of the two tasks, usually the one the interviewee believes to be the less relevant (Caso & Vrij, 2009). An example on a secondary task request tactic is to ask to the interviewee to perform the Corsi's tapping task during while undergoing an interview (Caso, Morganti, Palena, & Vrij, under review). Another request which increases cognitive load is to ask to recall the event in a reverse-chronological order (Evans, Michael, Meissner, & Brandon, 2013; Vrij et al., 2015). This latter approach makes the interview more cognitively demanding (especially for liars) on one hand, but it also aids memory recall (for truth tellers) on the other, as seen in the section dealing with the ACID technique.

Encouraging the interviewee to tell more builds on two main ideas. First, interviewee usually do not successfully grasp how much detail is expected from them. Second, liars tend to use, among the others, an "avoid and keep it simple" strategy (Granhag, 2010; Leins, Fisher, & Ross, 2013; Vrij, 2008). Consequently, by encouraging to tell more, the benefit can be two-fold: Truth tellers are encouraged to report details that they ignored, which results in more information elicitation which can in turn be beneficial for case construction. Liars, on the other hand, are forced to add details because in not doing so they may appear suspicious. Such details, then, can betray liars' lies. Again, there are several tactics to encourage the interviewee to tell more, mainly divided in nonverbal and verbal approaches. For the nonverbal approach, the interviewer should mimic interviewees' behaviour, which results in truth tellers being

more detailed, and reporting more accurate pieces of information, than liars (Shaw et al., 2015). Another nonverbal approach is to have an interviewer being supportive to the interviewee. By simply having a second interviewer who nodded and smiled at the interviewee while a first interviewer was questioning mock suspects, Mann et al. (2013) found that truth tellers reported more details than liars, while this effect was absent when the second interviewer acted neutrally. As far as for the verbal strategy, the interviewer can provide a model statement to the interviewee. A model statement is an example of a detailed answer, of an event unrelated to the one under investigation, that can be presented to the interviewee (for a recent overview of the model statement technique, see Vrij, Leal, & Fisher, 2018). The model statement can be provided both as an audio or as a text (Leal, Vrij, Warmelink, Vernham, & Fisher, 2015). One of the theoretical reasons behind the model statement is the principle of social comparison (Festinger, 1954), which results in an increase of interviewee's expectation of how much detail is needed (Leal et al., 2015; Vrij, Leal, & Fisher, 2018; Vrij, Leal, et al., 2017). Research on the use of the model statement shows that it consistently elicits more detail in both truth tellers and liars (Ewens et al., 2016; Kleinberg, Toolen, Vrij, Arntz, & Verschuere, 2018; Leal, Vrij, Deeb, & Jupe, 2018; Leal et al., 2015; Vernham, Vrij, & Leal, 2018; Vrij, Leal, & Fisher, 2018; Vrij, Leal, et al., 2017), as well as that it can be helpful for lie detection purposes when specific details are analysed. Vrij, Leal, Jupe, et al. (2018), found that liars reported a lower proportion of complication than truth tellers, but only in the post-model statement recall. Similarly, Leal et al. (2018) found that after the model statement was presented, liars reported more peripheral details than truth tellers. Considering its utility for both information elicitation and lie detection purposes, the model statement is receiving positive feedback from practitioners as well (Vrij, Leal, & Fisher, 2018).

The asking unexpected questions tactic builds on the research finding that liars prepare themselves for questions they expect to be asked (Hartwig, Anders Granhag, & Strömwall, 2007). Yet, the strategy only works when the lying interviewees correctly predicts what question will be posed to them. Therefore, this tactic builds on the idea that interviewee should be asked questions that they do not expect, together with questions they expect. Alternating expected and unexpected questions makes possible to compare interviewee's answers to the two types of questions. Truth

tellers should not show a big difference, as they will recover from memory answers for both types of questions. Liars, on the contrary, are affected by such a request. Knieps, Granhag, and Vrij (2013) for example found liars to be *more* detailed than truth tellers when answering expected questions, but *less* detailed than truth tellers when answering unexpected questions. The unexpectedness of questions can relate to their content (such as spatial details, or processess vs. outcomes, Mac Giolla, Granhag, & Liu-Jönsson, 2013; Vrij et al., 2009) or to the format. For example, the investigator can ask the interviewee to draw a sketch of places where the crime took place. This works in two ways. First, it encourages the interviewee to tell more. Second, it is an unexpected request. Indeed, research shows that this is the case, and that liars show less overlap between their verbal answers and sketches than truth tellers (Leins, Fisher, & Vrij, 2012; Leins, Fisher, Vrij, Leal, & Mann, 2011).

In conclusion, the CCA appears to be a powerful tool, and some of its tactics are already used (or thought to be ready to be used) in real life scenarios (Vrij & Fisher, 2016).

3.4 **The Baseline approach**

Background

Research showed that there is no Pinocchio's nose (Vrij, 2008): Different people can show different cues to deception. That is, someone may move their hands *more* while lying, whereas someone else may do just the opposite. This may partially explain why cues to deception are faint and unreliable (DePaulo et al., 2003). Indeed, if lying is not consistently associated with a specific change (direction) of a specific cue (e.g.: hands movements always *decrease* while lying) then it is not surprising that research failed to find reliable cues to deception. There are several reasons that can explain such inconsistency. Cues to deception may change according to context, emotional tone, level of cognitive engagement, personality factors, communication styles, etc. (Caso et al., 2018; Vrij, 2016).

At a first sight this is clearly inconvenient. Yet, the factors reported above can be turned in favour of the investigator for lie detection purposes. Take communicative styles as an example: Each person may show a specific communicative pattern when lying. This is not to say that each person *always* behaves in a specific way when lying.

Rather, it is to say that a trend may appear. This rationale brought to research on the efficacy of the baseline approach. In short, it is based on the idea that if an observer has a "reference line" (in terms of verbal and nonverbal behaviour) of sender's truthful behaviour (message), then deviations from such a reference will be used as an indication of lying.

A first line of research on the baseline approach framed the "baseline" as "familiarity" with the sender and explored whether this increased observer's accuracy rates (Brandt et al., 1980a, 1980b). For example, Feeley et al. (1995) provided their observers with zero, one, two, three, or four exposures to senders' baseline behaviour and found a positive linear relationship between baseline familiarity and observer's accuracy. This is an interesting result showing that observer's accuracy can increase when they are provided with senders' truthful behaviour. Yet, it is difficult to see how this can help investigative interviews. Indeed, the interviewer and the interviewee are often strangers and there is no possibility to become "familiar" with the suspect.

Baselining in investigative interviewing: The Small Talk approach

Since it is not common that an interrogator is familiar with the interrogee, there is the need to find alternative ways to obtain baseline messages. Research on the baseline approach in investigative interview builds on the assumption that the investigator should create a baseline at the beginning of the interview through specific questioning. That is, the interrogator should start the interview by asking non-threatening questions that are unrelated to the topic under investigation and from which s/he (the investigator) expects truthful answers. Suspect's answers to such questions should then be used as a baseline. The initial questioning here is two-fold: Creating a baseline, on the one side, and building rapport with the interviewee, on the other (Frank, Yarbrough, & Ekman, 2006; Inbau, Reid, Buckley, & Jayne, 2013). Again, any deviation from the baseline should be interpreted of a sign of lying. There is a clear problem with this approach: Non-threatening questions are very different from the topic under investigation. They entail different stakes, arousal, context, emotional responses, and cognitive load (Ewens, Vrij, Jang, & Jo, 2014; Palena, Caso, Vrij, & Orthey, 2018a; Vrij, 2016). Further, people change their behaviour and speech when speaking about different topics (Kleinke, 1986), over the course of an interview

(Buller & Burgoon, 1996; Stiff, Corman, Krizek, & Snider, 1994), when questioned with different styles (Caso, Maricchiolo, et al., 2006; Vrij, 2006) or by different people (Mann et al., 2013; Vrij & Winkel, 1991). Also, and perhaps even more relevant to this context, is the fact that people communicate differently when they speak about the topic they value vs. a topic they found less relevant (Davis & Hadiks, 1995). Consequently, the risk, when creating a baseline by asking non-threatening questions at the beginning of an interview (Small Talk baseline), is that *both* truth tellers and liars will different behaviours and speech patterns when answering baseline questions vs. investigative questions.

Ewens et al. (2014) empirically tested the efficacy of Small Talk baselining. They interviewed 243 participants, all answering honestly to the baseline question (focusing on the content of the consent form signed to take part in the experiment). Concerning the investigative questioning, truth tellers had to describe their job honestly (half of the sample), whereas liars had to describe a job they pretended to have (remaining half of the sample). The authors found that both truth tellers' and liars' answer to the baseline question differed from answers to target questioning, in that all participants changed their level of cognitive load (thinking hard) and self-monitoring. Ewens' et al. (2014) results hence showed that the Small Talk baseline approach is not an effective way to elicit cues to deception.

Baselining in investigative interviewing: The Comparable Truth approach

Ewens' et al. (2014) results were not promising for the baseline approach. Yet, the fact that the Small Talk baseline is not an effective lie detection approach does not mean that the baseline approach is not effective at all. The problem is in the way the Small Talk baseline is built. Obtaining the baseline in a different way can make it more effective. Caso and Vrij (2009) and Vrij (2016) suggest that a different form of baseline may benefit lie detection. That is, the investigator should create a Comparable Truth Baseline, that must be *"similar in content, context, stakes, and cognitive and emotional involvement to investigative questions"* (Vrij, 2016, p. 1114). Indeed, if these factors are kept constant throughout the whole interview (baseline vs. investigative questioning), behavioural and speech deviations will be more likely due to lying.

Our research at University of Bergamo, which is part of this dissertation, experimentally explored the efficacy of such type of baselining and compared it to that of the small talk approach both for the elicitation of cues to deception and the effects on observers' accuracy (section from 4.3 to 4.7).

4 The empirical contribution

4.1 General and specific aims

After 9/11, there has been a great deal of attention within the academic community to the development of ethically valid interviewing techniques aimed at eliciting information and detecting deception (Department of the Army, 2006; High Value Detainee Interrogation Group, 2016; Vrij, Fisher, et al., 2017). Therefore, the scientific literature now offers an interesting amount of information that can be shared with, and applied by, practitioners in real-life situations. This dissertation had three main goals.

First, to increase information elicitation through the combination of the Scharff Technique (Granhag, Kleinman, et al., 2016; Toliver, 1997) and social influence strategies (Cialdini, 2001) (section 4.2). Indeed, when considering real life applications, the first goal of intelligence agencies (as well as that of the police for the criminal justice system) is to obtain information from the source. Yet, although interviewees very rarely remain completely silent, they may not be fully cooperative (Granhag, Kleinman, et al., 2016; Soufan, 2011; Toliver, 1997). Hence, noncoercive techniques can be employed to boost interviewees' cooperation and willingness to disclose information to the interviewer. There are several approaches that can be used for this goal, yet they are rarely used together. As said above, Study I is an endeavour to explore whether combining the Scharff technique and the use of persuasion enhance information disclosure.

Once information is gathered, there is the need to assess its veracity. Indeed, especially in these contexts, there is a high risk that sources lie to pursue their own (as well as their organisations) goals. Information-gathering hence becomes the first, essential, step to assess the veracity of what the interviewee says. The second goal of this dissertation was thus to explore the efficacy of a specific interviewing technique for lie detection purposes, the Baseline technique. We tested its performance for eliciting cues to deception (Studies II-III, sections 4.3 and 4.4) and improving observers' lie detection accuracy (Studies IV-VI, sections 4.5, 4.6, and 4.7). The technique, however, does not deal with the issue of embedded lies and mixtures of truths and lies (Study VII, section 4.8). It is clear from personal experience and

research outputs that people opt for the strategy to tell a mixture of truths and lies to appear more credible, but most of the research focuses on outright fabrications (Leins et al., 2013). Therefore, we also developed a novel interviewing approach, the Theme-Selection Strategy, to fill this important gap.

The third and last goal was to explore practitioners' point of view on the utility of interviewing protocols that have been developed in academia (section 4.9). Indeed, once research shows their potential, it is paramount to explore whether they have practical value. Yet, there is evidence showing that it can be difficult for practitioners to adopt such protocols in their professional activities (Vrij et al., 2015). Study VIII thus explored Bergamo police officers' point of view in this regard.

It is important to note here that, for most of the studies reported in this chapter, it was not possible to conduct a priori power analysis effectively because of the novelty of the studies themselves. Hence, we planned to have at least 30 participants per condition whenever possible, as it has already been suggested that this number should bring to a power of 80%, given medium to large effect sizes (Cohen, 1988; VanVoorhis & Morgan, 2007). Consequently, there is the risk that the studies reported here were only able to detect effect sizes in this range, which in turn increases the risk of type II errors and high uncertainty of the effect sizes obtained. Results, also those that have been published, should hence be interpreted with caution.

4.2 Study I – Reciprocity: Does it help in Scharff interviews?[10]

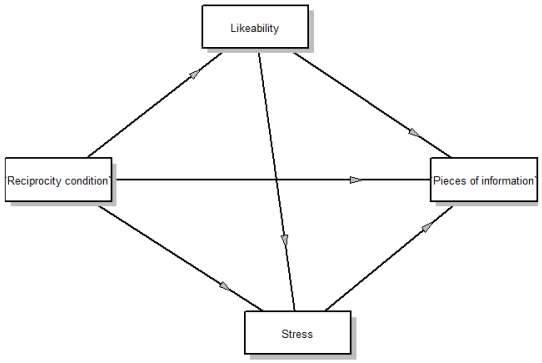
4.2.1 Background

Study one focused on improving information disclosure in Scharff interviewing through reciprocation. On the one hand, studies on the Scharff technique showed that sources reveal new information to obtain a gain (Granhag, Kleinman, et al., 2016). Yet, sources still reveal few new pieces of information (Oleszkiewicz, Granhag, & Cancino Montecinos, 2014). On the other, studies on reciprocation show that this principle is an effective form of social influence (Cialdini, 2001). Additionally, research shows that reciprocating through sharing can amplify experiences (Boothby,

[10] The paper relating to this study is in preparation for submission as:
Palena, N., Caso, L., Granhag, P. A., Orthey, R., Monticciolo, R., & Vrij, A. Reciprocity: Does it help in Scharff interviews? [IP2]. See appendix 1.

Clark, & Bargh, 2014). The main goal of this study was to combine the two and explore whether applying the social principle within Scharff interviewing increases information disclosure. We hypothesised that offering water to the interviewee would increase the number of new pieces of information revealed[11], particularly when the water was shared (the interviewer drunk water together with the interviewee) (Hypothesis 1). We also hypothesized that this effect was mediated by interviewer's likeability and participant's stress (Hypothesis 2). That is, we hypothesized that reciprocity would increase interviewer's likeability which, in turn, would increase information disclosure. Also, we predicted that higher likeability would reduce the stress experienced by the interviewee which, in turn, would increase information disclosure (Figure 2).

Figure 1. *Mediation path diagram. Source: Palena et al. (in preparation).*



4.2.2 Method

Sixty university students took part in the experiment (55 females and 5 males). Participants played the role of a source belonging to a criminal organisation. As in previous work on the Scharff technique, participant were asked to strike a balance between not providing too much nor too little information, and were told that according to their performance they would obtain a penalty discount (role-playing)

[11] The initial focus was to use the proportion of true new pieces of information revealed (true new pieces of information/total number of information revealed). Yet, preliminary analyses showed that the incidence of lying was extremely low. Therefore, the analyses focused on the raw amount of new pieces of information revealed.

and between 1 and 3 points for an University examination (all participants received 2 points, eventually). One-third of the sample went straight into the interview as soon as they memorised the information concerning the criminal organisation. One-third was invited to have a glass of water before the beginning of the interview. One-third was offered water by the interviewer who always drunk with them. The interview then took place, its structure was identical for all conditions. The interviewer started by presenting already known pieces of information. Afterwards, the interviewer gave to the participant the chance to add any new piece of information they wished. Once the participant had finished talking, the interviewer implemented the confirmation/disconfirmation tactic. Then, the interviewer asked to the participants if they had anything to add. Participants were left free to lie if the wished to. When the interview was over, participants filled a questionnaire. The questions concerned how much motivated to succeed in the experiment the participant felt; how likeable the interviewer was; how pleasant the interviewing room was; how stressful the interview was; all on a likert scale from 1 (not at all) to 7 (very much). Further, the interviewee was asked how much information they believed the interviewer already knew prior to the interview, on a likert scale ranging from 1 (none) to 7 (all); also, the interviewee was asked to tick each of the 29 pieces of information that they believed the interviewer knew prior to the interview. Finally, they were asked to rate the importance of each piece of information, were 1 = little important, 2 = moderately important, and 3 = very important. Disclosure was coded via a checklist. Coders noted what specific piece of information had been revealed and if it was truthful or falsified. Information concerning the confirmation tactic were only counted when clearly affirmed. Inter-rater reliability was high ($ICC_{(2,2)} = .91$)[12].

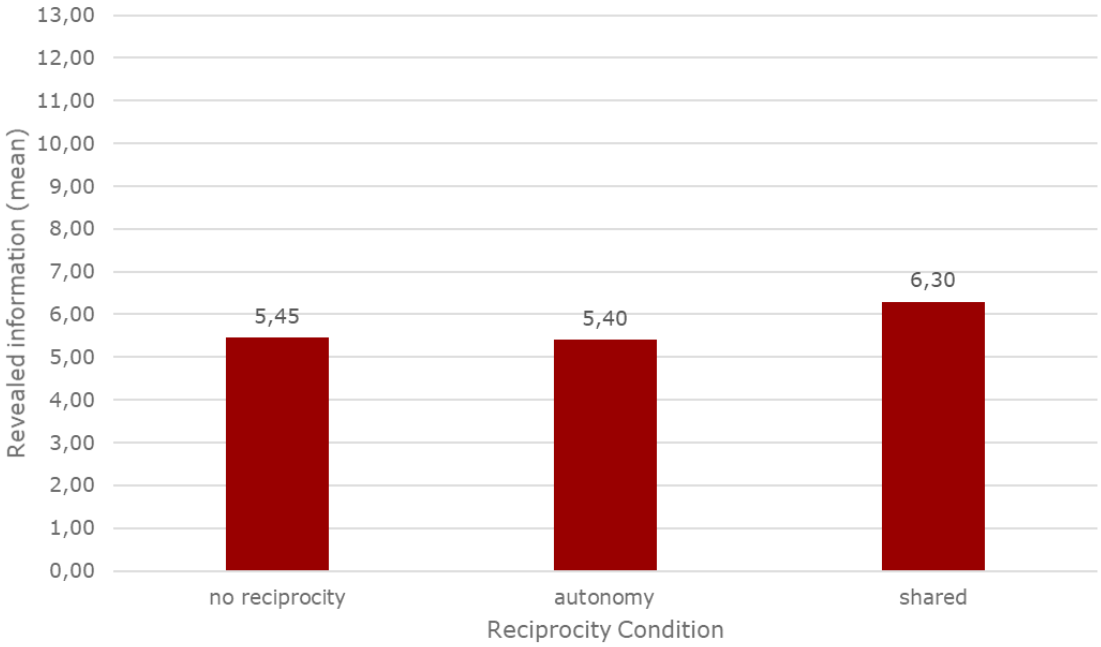
4.2.3 Results and discussion

Interviewees in the sharing condition revealed more new pieces of information than interviewees in the two other conditions, yet not significantly so (Figure 2). Therefore, Hypothesis 1 was not supported. Hypothesis 2 was not supported neither as the mediated effect (through interviewer's likeability and interviewee's stress) of reciprocity on information disclosure was not significant. The same held true for

[12] Both coders coded 50% of the transcripts. Only one coder coded the remaining 50%.

pieces of information that were *believed* to be known to the interviewer by the interviewee. Further exploration showed that interviewees revealed about 50% of information perceived to be known to the interviewer but only about 30% of information perceived to be unknown to the interviewer.

Figure 2. *Means and standard deviations of the amount of information revealed in each reciprocity condition. Source: Palena et al. (in preparation).*



Results seem to show that reciprocity did not increase information disclosure in Scharff interviewing. One possible explanation for this result is due to the Scharff technique itself. Indeed, offering a penalty discount/additional points is itself a form of reciprocation. This may have overshadowed any possible effect of offering water to the interviewee. Future studies should explore if this is the case. Another possible explanation is based on the type of reciprocity, which can be both intrinsic or instrumental (Sobel, 2005). In the first case, reciprocity takes place because the receiver responds to sender's perceived intention. That is, if the receiver perceives that the sender is kind, s/he may reciprocate him/her with being kind back. In the second case, a more "selfish" type of reciprocity takes place, thus the receiver reciprocates the sender because s/he believes this to be a benefit for himself/herself. Offering water is likely more an intrinsic form of reciprocity, which may be effective

in personal relationships but not in bargains such as those taking place in HUMINT interviewing. Indeed, sources strategize on what information to disclose according to what they believe is best for themselves (Neequaye & Luke, 2018). An additional explanation relates to the statistical power. The small group may have led to an underpowered analysis, which was only able to detect large effect sizes. Yet, a Bayesian Analysis with "Reciprocity" as the only factor and the number of information revealed as the dependent variable showed a BF_{01} of 3.33, a moderate support for the null hypothesis (Wagenmakers et al., 2018). Hence, future studies should try to disentangle this issue and explore what is more likely between a lack of an effect (i.e.: reciprocity does not help in Scharff interviews) and the presence of a small effect.

The fact that interviewees revealed more information considered as known than as unknown to the interviewer fits with the extant literature (Granhag & Luke, 2018) and is consistent with common-sense. By disclosing known information, the source appears cooperative whilst providing little (if any) actionable intelligence.

This study seems to show that implementing reciprocation into Scharff interviews does not benefit information disclosure. Future studies should explore the effect of reciprocity in a different manner. Intrinsic vs. instrumental reciprocity should be analysed. Further, future studies should explore whether the effect of reciprocity increases information disclosure when applied within-subjects rather than between-subjects. Last, the mediated effect of reciprocity on information disclosure should focus on rapport as a mediator rather than on interviewer's likeability. Indeed, having a kind interviewer may result in a ceiling effect of likeability, with the consequence of an impossibility to gauge any mediated effect. Rapport may better represent a mediated effect of reciprocity on information disclosure.

4.3 Study II – Efficacy of the baseline technique when assessing credibility[13]

4.3.1 Background

It has been suggested that cues to deception may be enhanced when a truthful baseline behaviour of the sender is available. Creating such a baseline through Small Talk proved ineffective (Ewens et al., 2014; Palena et al., 2018a; Palena, Caso, Vrij, & Orthey, 2018b). The present study focused on the efficacy of a different type of baseline, the Comparable Truth Baseline. This baseline is defined as *"similar in content, context, stakes, and cognitive and emotional involvement to investigative questions"* (Vrij, 2016, p. 1114). The rationale here is that if such factors are kept constant, deviation from the baseline in senders' answers to questions relating to the event under investigation should appear when lying. Indeed, it is difficult to find other reasons for behavioural and speech differences if the two events (baseline vs. investigative section) are similar and their relative memories can be recalled with the same level of cognitive engagement. In case the Comparable Truth Baseline is effective, thus, only liars should behave/speak differently when answering baseline vs. investigative questions. Hence, we hypothesize an interaction effect between veracity (truth tellers vs. liars) and time (baseline answer vs. target answer), both for impressionistic cues based on likert scales (Hypothesis 1), such as "vagueness", and for count variables, such as the frequency of specific details (Hypothesis 2).

4.3.2 Method

Thirty-seven (24 females and 13 males) took part in the experiment. All participants were offered two additional points for a university examination for their participation in the study. Participants had to commit a mock crime mirroring an undercover mission. They had to access to a notebook with a password they received from the experimenter. Once logged in, the participants read a word document instructing them to locate a CD-ROM from a backpack. Once played, the CD-ROM showed a man

[13] The paper relating to this study has been published as:
Palena, N., Caso, L., Carlotto, G., De Mizio, L., & Marciali, M. (2017). Efficacia della tecnica di baselining nella valutazione della credibilità (Efficacy of the baseline technique when assessing credibility). *Giornale italiano di psicologia, 44*(4), 905-916. doi:10.1421/88773. [PA1]. See appendix 2.

talking in front of a camera, who instructed the participant to look for a key hidden in the same backpack from where the participants collected the CD-ROM. The key served to open a safe deposit box which included written instructions (writing an email to a specific email address) and to wait for a person to come. When this person arrived (a confederate), the participants received a newspaper with additional instructions hidden in it. The instructions informed the participants that they had to: i) go to another room; ii) find a USB stick hidden on a coat-hook and take it; iii) look for a specific book in a wardrobe located in the room; iv) switch the USB drive they collected from the coat-hook with another USB drive hidden inside the book; v) leave the newspaper next to the book and go back to the first room and wait for the experimenter to come. The meeting with the confederate served as a landmark: Everything happened before the meeting was part of the baseline section of the interview, whereas everything happened after that was part of the investigative section of the interview. Truth tellers were asked to answer truthfully to all questions. Liars were asked to be truthful for everything except anything happened after the meeting with the confederate. They were also asked to create a credible and detailed cover-up story for this. The interview started with the interviewer introducing himself and welcoming the participants. Then two open-ended questions were asked, one concerning the baseline section and one concerning the target section of the interview[14]. The interview then ended.

All interviews were transcribed and coded for:

a) Cognitive load
b) Body rigidity
c) Detail
d) Logical structure of the statement
e) Unstructured production

[14] Once the participant answered the baseline question, the interviewer left the room with an excuse. The experimenter then entered the room and gave veracity instructions to the participant. Then the interviewer came back and asked the target question. Although this is a limitation, it was essential to proceed in this manner. Indeed, this was the first experimental study on the Comparable Truth baseline, and experimental control over the variables was needed. There was indeed the risk that lying interviewees would lie during the baseline section to appear credible, but this would have vanished the efforts aiming at exploring whether the Comparable Truth baseline is effective when a truthful comparable baseline is indeed available.

f) Vagueness

on a likert scale from one (totally absent) to four (very much), and for:

a) Visual details
b) Auditory details
c) Spatial details
d) Temporal details
e) Action details

Two coders coded 25% of the transcripts, and inter-rater reliability was good (ranging from ICC = .86 to ICC = .96). The first coder coded 100% of the transcripts.

4.3.3 Results and discussion

For the variables coded on a likert scale, the Time by Veracity interaction appeared only for vagueness. Follow-up analyses showed that the difference between the baseline and the target answer for the level of vagueness was significant only for lying participants. Liars appeared as vaguer when answering the target question then when answering the baseline question, in partial support of Hypothesis 1. Additionally, the interaction appeared for the variables spatial and visual details, in support of Hypothesis 2. Only liars showed a difference (a decrease) in these details when comparing their baseline to their target answers.

Both the results for Hypothesis 1 and Hypothesis 2 went in the direction that would be expected by previous work (DePaulo et al., 2003), as liars appeared as vaguer, but reported less spatial and visual details, when answering the target question (lie) than the baseline question (truth). Ewens et al. (2014) found that their Small Talk baseline was ineffective, as both truth tellers and liars changed their behaviour from baseline to target answers. The present experiment reached different results, as only liars showed (statistically significant) differences between the two phases of the interview. This sheds light on the fact that it is not the baseline approach itself to be either effective or ineffective, but the way such a baseline is obtained. As predicted, if baseline questioning is very similar to target questioning, then cues to deception may be enhanced. A Comparable Truth baseline may also reduce interpersonal differences, as it is a form of within-subjects comparison that is preferred by both academics and

practitioners (Vrij, 2016; Vrij, Leal, Jupe, et al., 2018). For example, research shows that people usually decrease hands and finger movements when lying, yet there are people that may show the opposite pattern (Vrij, 2008). Just looking at specific cues may be misleading, as using always the same rule (e.g.: decrease in hands and finger movements indicate lying) may bring to decision errors. The baseline approach, if well-constructed, may partially reduce the problem. However, there are of course limitations to the Comparable Truth baseline approach. For example, it is not always possible to obtain a valid truthful baseline, and it is impossible to know whether it works in different settings, such as real-life scenario with high stakes at play. Further, we only obtained support for the hypothesis that a comparable truth baseline is effective for three out 11 dependent variables and the multivariate interaction effects were not significant. Hence, results must be taken with caution. Last, it is important to compare the Comparable Truth and the Small Talk baseline approaches directly, which is the focus of the next section.

4.4 Study III – Detecting deception through small talk and comparable truth baselines[15]

4.4.1 Background

The previous study focused on the efficacy of the Comparable Truth baseline for the elicitation of cues to deception and obtained promising results, contrarily to Ewens et al. (2014), who conducted a study on the Small Talk baseline approach. Although there are theoretical reasons to believe that this difference in the results of the two studies can be traced back to the way the baseline is obtained, it is not possible to exclude that other factors played a role. For example, the difference between our and Ewens' and colleagues (2014) results may also reside in a difference of the content of the interview or in the experimental procedure. The present experiment therefore focused on a direct comparison of the two baseline approaches in their effectiveness to elicit cues to deception. Based on the idea that the Small Talk baseline is not effective because there are fundamental differences between the baseline and the

[15] The paper relating to this study has been published as:
Palena, N., Caso, L., Vrij, A., & Orthey, R. (2018). Detecting deception through small talk and comparable truth baselines. *Journal of Investigative Psychology and Offender Profiling, 15*(2), 124-132. doi:10.1002/jip.1495. [PA2]. See appendix 3.

target sections of the interview, but that the Comparable Truth baseline is effective, as the two sections are held as similar as possible, we expected that truth tellers and liars would obtain similar similarity scores (see below, method section) in the Small Talk baseline, but truth tellers would obtain higher similarity scores than liars in the Comparable Truth baseline (Hypothesis 1). Further, since research shows that verbal lie detection is better than nonverbal lie detection (Vrij, 2016), we also predicted that the effect sizes in the Comparable Truth baseline would be more pronounced for verbal behaviours than for nonverbal behaviours.

4.4.2 Method

Sixty-nine participants (49 females and 20 males) took part in the experiment and were offered two additional points for a university examination for their participation. The procedure was identical to the previous study. The only, additional, difference is that participants in the Small Talk baseline were asked to describe their last year as a student/worker in response to the baseline question. We focused on both nonverbal and verbal behaviour. For nonverbal behaviour, we focused on the frequency per minute of hands/fingers movements, one arm movements and two arms movements. For verbal content, we analysed the occurrences of spatial, temporal, visual, audio and action details. Two coders coded 25% of the transcribed interview for the variables reported above. ICC ranged from .75 to .96, indicating high agreement between the two coders. One coder coded the remaining 75% of the transcripts. The focus of this study was on "similarity". In baseline interviews, liars are expected to show behavioural differences, thus, to be less "similar" in their behaviour than truth tellers. The similarity scores were computed as follows: for each dependent variable, we divided the lowest score by the highest score, regardless of the phase of the interview. This means that if the lowest score for a variable appeared in the baseline, we divided the score of that variable in the baseline by the score of the same variable in the target phase. The same was true for the opposite situation. Then, we multiplied the result by 100. The highest the score, the more similar participants behaved between the two phases of the interview.

4.4.3 Results and discussion

There was no Baseline (Comparable Truth vs. Small Talk) by Veracity (Truth tellers vs. Liars) interaction for nonverbal behaviours, but the interaction was significant for

one verbal cue[16]: The number of spatial details (Figure 3). Follow-up analyses showed that, for participants in the Comparable Truth baseline condition, truth tellers reported higher similarity scores for spatial details than liars (Table 2), whereas this did not happen for participants in the Small Talk condition (Table 3).

Figure 3. *Baseline by Veracity interaction for the variable spatial details.*

Hypothesis 1 was therefore supported only for Spatial details. Furthermore, three out of eight variables showed similarity scores that went in the "wrong" direction (less similarity amongst truth tellers), but for those five who went in the predicted direction, the strongest effect size was for spatial details, d = 1.20, followed by hands and finger movements, d = .54. The effect sizes for the remaining three variables were small. This means that Hypothesis 2, the effect sizes would be stronger for verbal behaviours than for nonverbal behaviours, only obtained very limited support.

[16] The analyses for verbal content did not include as covariates, initially, the total number of details in the baseline and target sections of the interview (see appendix 3). The inclusion of such covariates was asked by one reviewer when the paper was submitted for publication. However, for the sake of transparency, it is important to underline here that the main conclusion do not change whether the covariates are entered or not in the analyses. Indeed, when the covariate were not entered, the Baseline, $F(5, 61) = 31.15$, $p < .001$ and Baseline by Veracity interaction, $F(5, 61) = 2.64$, $p = .03$ multivariate effects were anyway significant, and the Veracity main effect was again not significant, $F(5, 61) = 0.65$, $p = .66$.

Table 2. *Similarity scores as a function of veracity in the comparable truth condition. Source: Palena et al. (2018a).*

	Truth tellers		Liars				
Similarity scores	*M*	*SD*	*M*	*SD*	*F*	*p*	*d*
Hands and fingers movements	77.95	23.53	65.54	22.12	2.72	.10	.54
One arm movements	69.52	27.82	66.73	26.02	.09	.75	.10
Two arms movements	52.87	39.42	54.74	35.38	.02	.88	−.05
Spatial details	78.67	15.60	52.25	26.75	7.44	.01	1.20
Temporal details	43.51	33.14	55.79	30.18	1.09	.30	−.38
Visual details	70.56	17.95	65.53	21.12	.53	.47	.25
Audio details	14.47	32.61	30.46	39.78	.51	.48	−.44
Action details	66.09	18.21	61.96	22.89	.00	.98	.20

Table 3. *Source: Similarity scores as a function of veracity in the small talk condition. Palena et al. (2018a).*

	Truth tellers		Liars				
Similarity scores	*M*	*SD*	*M*	*SD*	*F*	*p*	*d*
Hands and fingers movements	57.94	27.68	70.58	27.33	2.23	.14	−.46
One arm movements	41.52	32.16	30.33	24.57	.35	.55	.39
Two arms movements	59.77	37.98	57.58	35.78	.00	.96	.06
Spatial details	14.69	30.79	33.66	42.93	.09	.76	−.50
Temporal details	40.98	30.57	54.60	30.23	1.518	.23	−.45
Visual details	10.36	12.88	13.20	23.68	.21	.65	−.14
Audio details	41.17	50.73	33.93	45.58	.42	.52	.15
Action details	45.50	22.21	55.45	27.87	.78	.39	−.39

The appearing picture suggests that: i) the study failed to find evidence that Small Talk baseline is effective as truth tellers and liars showed a similar level of similarity between the baseline and the target phases of the interview[17]. Therefore, it may be better to drop this approach altogether (Ewens et al., 2014); ii) the Comparable Truth baseline works better, as truth tellers obtained higher similarity scores than liars. Yet, the approach is still problematic. Indeed, it worked only for one cue (spatial details), hinting at the fact that the baseline approach could be tightly connected to the content of the story. Again, the "comparability" of the two sections is paramount. Further, as Table 2 shows, truth tells did not show perfect similarity themselves, meaning that an investigator should consider the amount of "similarity" to decide whether someone is lying or telling the truth. Considering that a cut-off point is non-existent (and will likely never be obtained) the task is still difficult also when a Comparable Truth baseline is used. Nonetheless, this

[17] Bayes Factors$_{01}$ for each of the dependent variable went from 1.28 to 2.97, seemingly suggesting anecdotal evidence of the lack of an effect.

difference in similarity between truth tellers and liars may be detectable to the observers, with a positive effect on his/her lie detection accuracy. This is the focus on the next three sections.

4.5 Study IV – Observers' performance at evaluating truthfulness when provided with comparable truth or small talk baselines[18]

4.5.1 Background

The two previous studies showed that a Comparable Truth baseline can elicit cues to deception. Indeed, only liars showed (non)verbal differences between the baseline and the target phase of the interview (Palena, Caso, Carlotto, De Mizio, & Marciali, 2017), and truth tellers appear to behave more similarly than liars, at least for what concerns the amount of spatial details (Palena et al., 2018a). Yet, this does not mean that these cues are noticeable by observers when assessing senders' honesty. It is therefore needed to test whether the Comparable Truth baseline, in addition to eliciting cues to deception, also improves lie detection accuracy, when compared to the Small Talk baseline. This exploration may fall under the umbrella of "interviewing to detect deception", where specific techniques are developed to enhance cues to deception and observers' lie detection accuracy (Vrij, 2014; Vrij & Granhag, 2012). Building on theoretical reasons suggesting the inefficacy of a Small Talk baseline (Vrij, 2016), as well as on experimental results confirming this assumption (Ewens et al., 2014; Palena et al., 2018a) we developed an experiment where the Comparable Truth and the Small Talk baselines were compared in terms of their effect on observers' lie detection accuracy. This is interesting both under a theoretical perspective and an applied one. Concerning the former, the rationale is to confirm the idea whereby if reliable cues to deception appear for the Comparable Truth, but not for the Small Talk baseline, then only observers provided with the first one should have a benefit for lie detection purposes. Concerning the latter, if observers are indeed more accurate when a Comparable Truth, rather than a Small Talk baseline, is employed, then practitioners and "on-the-field" investigators should drop the latter, and use the former only if

[18] The paper relating to this study has been published as:
Caso, L., Palena, N., Vrij, A., & Gnisci, A. (2019). Observers' performance at evaluating truthfulness when provided with Comparable Truth or Small Talk Baselines. *Psychiatry, Psychology and Law*. doi:10.1080/13218719.2018.1553471. [PA3]. See appendix 4.

valid, as already suggested elsewhere (Ewens et al., 2014; Palena et al., 2018a). We predicted that observers in the Comparable Truth condition would reach higher total (Hypothesis 1), truth (Hypothesis 2a) and lie (Hypothesis 2b) accuracy rates.

4.5.2 Method

Seventy-four participants (56 females and 18 males) took part in the experiment, but one participant was excluded from the analyses as he did not follow the experimental instructions. Participants were offered one additional point for a university examination for taking part in the study. Thirty-seven participants were allocated to the Comparable Truth condition; the remaining 36 were allocated to the Small Talk condition. The experiment was based on a simple two-groups comparison, with the Baseline condition (Comparable Truth vs. Small Talk) as factor, and total, truth, lie accuracy rates, *d*' and β values as dependent variables. Twenty video stimuli were used in total: 10 for each of the two experimental condition. Hence, 10 video showed mock suspects interviewed via a Comparable Truth approach and 10 interviewed with a Small Talk approach. All the videos were randomly chosen from the experiment described in section 4.4. Therefore, interviewees described the actions as described in Palena et al. (2018a). Yet, original videos were modified as follows. A black screen with a white text indicating "Baseline" appeared and lasted for 3 s. The first sender then appeared on the screen and started answering the baseline question. Once s/he finished, a second black screen with the text "Target" appeared, also lasting 3 s. Then the first sender started answering the target question. Once s/he had finished answering, another black screen appeared and lasted for 30 s. In this time-window, participants had to make their veracity decisions. When the 30 s had expired, a high-frequency sound warned the participant that the time to evaluate the first sender was over and that the second sender was going to appear on the screen. This sequence was repeated until the 10 videos had been seen. Participants expressed their veracity decisions answering the following question: "*Do you believe the interviewee was…*". The answer alternatives were "*Telling the truth*" and "*Lying*". Senders' veracity status was counterbalanced, with each 10-target tape consisting of 5 truth-tellers and 5 liars. It is important to note that observers did not received any form of training on cues to deception/interviewing techniques. They were just informed that senders always told the truth when answering the baseline question, but that they may be either lying or

telling the truth when answering the target question (they were not informed about how many videos they were going to assess, nor they were informed about the truth/lie ratio). Participants were asked to decide whether each sender was telling the truth or lying when answering the target question, based on (non)verbal deviations from the baseline answer.

4.5.3 Results and discussion

Results showed that observers in the Comparable Truth condition obtained higher overall and lie accuracy rates, supporting Hypothesis 1 and Hypothesis 2b. The overall accuracy rate of 56.49% obtained by participants in the Comparable Truth condition was higher that chance, but not higher than meta-analyses average scores of 54% (Bond Jr. & DePaulo, 2006). On the contrary, overall accuracy of participants in the Small Talk condition of 47.41% did not differ from chance but was lower than 54%. When considering lie accuracy, the only significant difference was between the accuracy of 55.14% reached by participants in the Comparable Truth condition and meta-analytical results showing that average lie detection accuracy is around 47% (Bond Jr. & DePaulo, 2006). Further analyses using Signal Detection Theory (Stanislaw & Todorov, 1999) showed that participants in the Comparable Truth condition discriminated truth tellers from liars better than participants in the Small Talk condition. Yet, responding bias (β values) did not differ between the two condition and all participants appeared to be truth biased.

Results of this experiment seem to show that participants in the Comparable Truth condition reached higher (total and lie) accuracy rates than participants in the Small Talk baseline condition. At a first sight, this indicates that the former is better than the second for veracity assessments. Yet, looking at the results in more detail, another picture can be drawn. The fact that total accuracy rates for observers in the Comparable Truth condition did not differ from what would be expected from previous meta-analyses (Bond Jr. & DePaulo, 2006), whereas that for participants in the Small Talk baseline was lower than it, may indicate that it is not the Comparable Truth that helps veracity assessments. It may be the Small Talk that negatively influences it. In short, having a Small Talk baseline is detrimental for veracity assessments, whereas having a Comparable Truth baseline does not change much the accuracy that would be expected when no baseline is available. Hence this experiment

shows that, although better than the Small Talk baseline, the Comparable Truth baseline is still problematic. Additionally, the effect of the baseline method on observers' lie detection accuracy, if explored via Bayes Factors, only indicated anecdotal evidence (Wagenmakers et al., 2018) for the superiority of the Comparable Truth baseline over the Small Talk baseline (for total accuracy, $BF_{10} = 1.50$, for lie accuracy, $BF_{10} = 1.30$, for *d*-prime, $BF_{10} = 1.51$). Our results must be then taken with caution. Additionally, our results come from an experiment where observers where not taught valid cues to deception. This means that participants may have relied on invalid cues to deception and the picture we obtained may change when observers are trained on valid cues to deception, but future studies are needed to disentangle this. Such trainings should be based on verbal cues rather than nonverbal cues, and future studies may explore how the Comparable Truth performs when implemented with other interviewing techniques. Lastly, this experiment only explored laypersons' accuracy. Yet, for applied reasons, it is important to explore professionals' accuracy when provided with the Comparable Truth baseline, that is the focus of the next section. Section 4.7 will instead deal with observers' accuracy when informed about valid cues to deception to be coded through a checklist.

4.6 Study V – Police accuracy in truth/lie detection when judging baseline interviews[19]

4.6.1 Background

The previous study showed that a Small Talk baseline influences observers' accuracy negatively, as they obtained accuracy rates below what would be expected by previous meta-analytic work (Bond Jr. & DePaulo, 2006). Together with previous work on the Small Talk baseline (Caso, Palena, Vrij, et al., 2019; Ewens et al., 2014; Palena et al., 2018a, 2018b) the picture is clear: This approach does not work, brings to decisions errors, and should be dropped altogether. On the contrary, the Comparable Truth baseline appears to be a better option, although it is still problematic (see previous sections). The only available study comparing the effect of the two baselines

[19] The paper relating to this study has been published as:
Caso, L., Palena, N., Carlessi, E., & Vrij, A. (2019). Police accuracy in truth/lie detection when judging baseline interviews. *Psychiatry, Psychology, and Law. doi: 10.1080/13218719.2019.1642258* [PA4]. See appendix 5.

published to date showed the benefits on laypersons' accuracy when opting for a Comparable Truth baseline rather than for a Small Talk baseline. Yet, it is important to explore how practitioners such as police officers perform when provided with a Comparable Truth baseline, which is the focus of the present section. Indeed, practitioners have a different experience and may show a different pattern, perhaps focusing on different cues. This becomes particularly interesting when considering that it has been suggested that investigative interviews should include a Small Talk baseline approach (Frank et al., 2006; Inbau et al., 2013). Yet, research clearly showed the inefficacy of the Small Talk approach, hence the present study focused on a comparison of police officers' veracity assessment accuracy when provided with a Comparable Truth baseline vs. when no baseline is provided. Building on theory and experimental results presented in the previous sections on the baseline approach, we predicted that practitioners provided with a Comparable Truth baseline would obtain higher total (Hypothesis 1), true (Hypothesis 2) and lie (Hypothesis 3) accuracy rates.

4.6.2 Method

Ninety-five practitioners (42 belonging to the Italian State Police, 28 to the Financial and Economic Police, 25 to the Military Police, called Carabinieri) participated in the study. One participant was excluded from the analyses as did not followed the instructions. The procedure mirrored the previous experiment. Participants in the Comparable Truth condition judged 10 senders interviewed with a Comparable Truth approach. Participants in the no baseline condition judged the same ten senders, but the videos for this condition only showed answers to the target answer.

4.6.3 Results and discussion

Total and truth accuracy did not differ between conditions, hence Hypotheses 1 and 2 were not supported. Lie accuracy, on the contrary, was higher for participants in the Comparable Truth condition (Table 4).

Table 4. *Descriptive according to the Baseline condition.*

	Comparable Truth		No baseline	
Accuracy rates	*M*	*SD*	*M*	*SD*
Total	53.26	14.91	49.58	13.67
Truth	51.73	22.93	60.00	17.01
Lie	54.78	20.41	39.16	15.95

The descriptive reported above may at first sight suggest that the Comparable Truth condition influences lie detection accuracy positively. Yet, looking closely at the means, a different picture arises. Indeed, officers in the Comparable Truth condition obtained similar lie accuracy rates to laypersons' lie detection accuracy found in the previous study (Caso, Palena, Vrij, et al., 2019) but, since officers in the no baseline condition obtained very low accuracy rates, the effect may be explained on the basis of responding bias rather than on improved lie detection accuracy. Further, there was no effect of the Comparable Truth baseline on *d*-prime. In short, having a Comparable Truth baseline brought officers from below-chance accuracy to chance-accuracy. To test for this effect, we ran Signal Detection analyses using *c* values for responding bias. Although, historically speaking, β values is the preferred metric for responding bias, *c* values are more relevant in this context. Indeed, *c* values are less biased and are based on the logic whereby, to decide whether the signal is present or not (hence, if the stimuli is just "noise") a threshold should be reached (Stanislaw & Todorov, 1999). The analyses on the *c*' value showed that: a) it differed from 0 (no bias) for participants in the no baseline condition, but it did not for participants in the Comparable Truth baseline condition; b) participants in the no baseline condition obtained higher *c* scores than those in the Comparable Truth baseline condition (Table 5). These results hence showed that participants in the no baseline condition were truth biased[20], whereas those in the Comparable Truth condition did not show a clear bias[21].

Table 5. *Descriptive for responding bias according to the baseline condition.*

	Comparable Truth		No baseline	
Responding bias	*M*	*SD*	*M*	*SD*
c	-0.05	0.61	0.29	0.26

[20] The *c* values is a measure for responding bias, but its interpretation depends on the way it is calculated. Mathematically speaking, a *c* greater than 0 indicates a bias toward responding "no", hence "stimulus not present". The way we calculated *c* associates the presence of the stimulus as answer to the question "was the sender lying?". Therefore, for our computation, a value greater than 0 indicated a tendency toward responding "truth teller", whereas a value lower than 0 indicated a tendency toward responding "liar"

[21] With a Bayesian approach, we obtained a Bayes Factor $BF_{01} = 5.40$, showing moderate evidence in support of H0 (Wagenmakers et al., 2018)

This can be read in terms of the Comparable Truth baseline affecting observers' responding bias rather than accuracy. This is also consistent with the Truth Default Theory (Levine, 2014), which predicts that observers tend to believe the others until "deceptive triggers" appear. In the context of a Comparable Truth baseline such deceptive triggers may originate from (non)verbal deviations from the baseline which, although more pronounced for liars than for truth tellers, are also present when telling the truth (Palena et al., 2017; Palena et al., 2018a). Yet again, the results obtained in this experiment must be taken with caution for the following reasons. First, observers were not trained on valid cues to deception. Was this the case, a different picture may have appeared. Future studies should therefore explore how the Comparable Truth baseline affects observers' accuracy when they are instructed on valid cues to deception. A simple design would include three levels for the baseline condition (No baseline vs. Small Talk Baseline vs. Comparable Truth Baseline) and two levels for training (not trained vs. trained on valid cues to deception). Dependent variables may be just *d*' and *c*' values. Second, we only showed 10 senders, but research showed that Senders' competence (and personal characteristics), veracity, media, and observers' accuracy influence each other (Caso et al., 2018). The following experiment tried to deal with the first point. However, rather than training participants on valid cues to deception, we provided them with a checklist to aid participants' judgments, which mirrors real life situations where there are no economic and time resources for a specific training (Evans, Michael, et al., 2013).

4.7 Study VI - Implementing the use of a checklist with comparable truth and small talk baselines: The effect on laypersons' lie detection accuracy[22]

4.7.1 Background

The previous experiments showed that the Comparable Truth baseline, if well-constructed, may elicit cues to deception (Palena et al., 2017; Palena et al., 2018a), whereas a Small Talk baseline does not (Palena et al., 2018a). Also, our studies

[22] The paper relating to this study is in preparation for submission as:
Caso, L., Palena, N., Vrij, A., & Melocchi, L. (*in preparation*). Implementing the use of a checklist with comparable truth and small talk baselines: The effect on laypersons' lie detection accuracy. [IP3]. See appendix 6.

showed that the Small Talk baseline impaired lie detection. Observers in the previous studies were not trained on lie detection literature, neither they had any tool as a support. Both these aspects may help lie detection, as research shows that people hold wrong beliefs about cues to deception (Global Deception Research Team, 2006) and that people make poor lie detectors (Bond Jr. & DePaulo, 2006, 2008). Hence, participants of this study were requested to make their veracity decisions using a checklist. Since there are both theoretical and experimental reasons that suggest a Comparable Truth baseline outperforms a Small Talk baseline, we predicted that observers allocated to the former condition would reach higher accuracy rates than observers allocated to the latter (Hypothesis 1). We also predicted an interaction effect between type of baseline and veracity of the sender (Hypothesis 2) on the checklist scores. Participants in the comparable truth condition were expected to rate liars as more different from truth tellers than participants in the small talk condition.

4.7.2 Method

One-hundred and twenty participants (93 females and 27 males) took part in the experiment and were offer two additional points for a university examination. Participants were requested to evaluate 10 video each and decide whether the senders were telling the truth or lying. The 20 videos (10 for observers in the Comparable Truth condition and 10 for participants in the Small Talk condition) were randomly selected from the previous study reported in section 4.4 (Palena et al., 2018a). The videos were modified so that observers had time to fill the checklist and make their veracity decision. Each set of 10 clips started with the sender answering to the baseline (either Comparable Truth or Small Talk) question, followed by a 15-seconds time-window to allow the participants to fill the checklist concerning the baseline answer. Then the sender answered the target question, and the observers were again given 15 seconds to fill the checklist for this target answer. Once the 15 seconds expired, the observers were warned through a high frequency sound that it was now the time to make their veracity decisions. They were given 30 seconds for this task. Then the following interviewee appeared on the screen and the same sequence took place. At the end of the ten videos the videoclip ended. Observers were not informed about how many videos they were going to evaluate and about the truth/lie ratio. Participants were also instructed to compare baseline and target answers within each sender. That

is, they were instructed to pay attention to difference showed by the same subjects, and not differences showed across subjects. The checklist (Appendix 2) focused on: i) contextual details; ii) sensory details; iii) verifiable details; iv) action details; v) cognitive load; and vi) vagueness, and was built on the basis of previous work (Evans, Michael, et al., 2013). Before starting the actual experiment, participants were explained how to use the checklist and showed practice videos.

4.7.3 Results and discussion

Results showed that there was no baseline effect on accuracy (Table 6): Observers in the Comparable Truth condition obtained similar accuracy rates and discrimination accuracy (*d'* scores[23]) of observers in the Small Talk baseline condition. Therefore, Hypothesis 1 was not supported.

Table 6. *Accuracy rates according to the baseline condition. Source:* Caso, Palena, Vrij, and Melocchi (in preparation).

	Comparable Truth	Small Talk			
	M(SD)	*M(SD)*	*t*	*p*	*Cohen's d*
Total accuracy	55.18 (17.89)	54.31 (16.37)	.277	.78	.05
Truth accuracy	57.88 (24.76)	58.44 (24.08)	-.126	.90	-.02
Lie accuracy	52.37 (24.40)	50.21 (24.69)	.481	.63	.09

Furthermore, when considering the checklist scores as dependent variable, the interaction between Baseline (Comparable Truth vs. Small Talk) and Veracity (Truth tellers vs. Liars) was not significant, meaning that Hypothesis 2 was not supported neither. On the contrary, both the Veracity and the Baseline main effect were significant. Participants rated liars are more deceptive than truth tellers, and participants in the Comparable Truth baseline condition rated senders as more deceptive than observers in the Small Talk condition.

Observers taking part in this study obtained accuracy rates (50-58%) in the typical range of lie-detection studies (Bond Jr. & DePaulo, 2006). Also, and contrarily to

[23] *d'* values are a measure of sensitivity (in standard deviation units) used in Signal Detection Analysis. They are used to indicate the capacity to discriminate between the signal and noise.

previous studies showing an apparent advantage of the Comparable Truth baseline over the Small Talk baseline (Caso, Palena, Vrij, et al., 2019, section 4.5), the type of baseline had no effect on accuracy. This result was unexpected also when considering psychological theory showing the issues of a Small Talk baseline approach (Palena et al., 2018a; Vrij, 2016). There may be several reasons for the null results obtained in this study. First, the differences between truth tellers and liars may have not been substantial enough in the comparable truth condition to become more accurate. Second, observers may have paid attention to irrelevant cues (i.e.: observers ignored their checklists scorings when deciding). Third, the instructions about how to use the checklist were either too short or not easy to understand. Fourth, interactions between senders' and observers' characteristics may have affected the outcome (Caso et al., 2018). Fifth, this study was better powered than the previous ones with the consequence of a more precise effect size estimate. This is also supported by the fact that the Bayes Factor for the difference of total accuracy in the two baseline condition is of $BF_{10} = 4.97$ for the present experiment and of $BF_{10} = 1.50$ for the previous one (Caso, Palena, Vrij, et al., 2019). Anyway, future research is needed. For example, it is important to explore what happens between attending a stimulus (sender) and the output decision and how to guide observers in this process. Indeed, there may be the possibility that observers allocated to the Comparable Truth condition did notice that liars showed larger (non)verbal differences than truth tellers, yet they may have employed ineffective decision-making strategies. This possibility becomes even more plausible when considering that cues to deception are faint and unreliable (DePaulo et al., 2003) and that, because of this, people base their decisions on context-general information when there is low diagnostic "individuating information" (Street, 2015, p. 335). A last interpretation is worth mentioning: providing participants with a checklist levelled out the difference in accuracy rates that appears when comparing the accuracy of observers that are provided with a Comparable Truth vs. a Small Talk baseline (Caso, Palena, Vrij, et al., 2019) or when no baseline is available. In short, teaching observers what to look for (using a checklist) improves lie detection accuracy, regardless of what interviewing strategy is employed as in doing so, observers pay attention to valid cues to deception. Yet, this assumption may not hold valid. Indeed, when a Comparable Truth baseline is employed only liars showed

(statistically significant) differences between the baseline and the target section of the interview (Palena et al., 2017). Second, only when a Comparable Truth baseline is employed truth tellers obtain higher similarity scores (i.e.: maintain a more similar behaviour between the baseline and the target section of the interview) than liars (Palena et al., 2018a). Yet, research suggesting an advantage of the Comparable Truth baseline over the Small Talk baseline comes from the same (our) laboratory. Therefore, to disentangle the issue, future studies in other laboratories/research groups are needed.

4.8 Study VII – Detecting lies via a Theme-Selection strategy[24]

4.8.1 Background

To date, there is a great amount of research on deception detection and HUMINT interviewing. Using the keyword "deception detection", Google Scholar returns about 131.000 outputs. Hence, although research is a never-ending process, we have learnt much on how humans lie and how we could improve our lie detection skills. Yet, the problem is that most of research focuses on outright lies or complete fabrications, but liars prefer to tell embedded lies or a mixture of truths and lies to appear more credible (Leins et al., 2013; Vrij, 2008). This becomes particularly relevant in HUMINT settings where, to appear credible and obtain a gain, the source may opt for: i) withholding most of the information; ii) revealing a mixture of truths and lies to appear cooperative.

The main goal of this study was to develop a new interviewing strategy to detect liars and their mixtures of truths and lies. The strategy, named "Theme-Selection Approach", is based on the following assumptions: i) interviewees categorize their knowledge into several themes. In the example of a terrorist attack, such themes may be the preparation for the attack, the bomb location, the electronical devices to be used, etc (Oleszkiewicz, Granhag, & Cancino Montecinos, 2014); ii) interviewees may vary the level of precision of pieces of information they reveal according to the

[24] The paper relating to this study has been published as:
Palena, N., Caso, L., & Vrij, A. (2019). Detecting Lies via a Theme-Selection Strategy. *Frontiers in Psychology, 9*(2775). doi:10.3389/fpsyg.2018.02775. [PA6]. See appendix 7.

specific themes being discussed; iii) interviewees may opt for mixing truths for some themes and lies for other themes.

Building on the previous points, it is clear that just looking to the total amount of detail as a cue to deception (DePaulo et al., 2003) may be misleading. A between-subjects comparison between truth tellers and liars focusing on the total amount of details may be poorly effective as liars' stories would also include many details (some of which would be true). On the other hand, focusing on within-subjects comparisons is more effective (Vrij, 2016). There are several ways to make such within-subjects comparisons (Vrij, 2016). For example, the Baseline approach is a within-subjects measure that focuses on the comparison between two different statements: a baseline statement and a target statement. The first one is only used as a reference line, whereas the second one is the topic of the investigation. However, the statements provided by interviewees in baseline interviews typically focus on unrelated topics (although the level of similarity of the content, stake, etc. of the two phases can vary, see Palena et al., 2018a). Further, the baseline approach, as well as other within-subjects strategies, do not deal with the problem of mixture of truths and lies. To reach this goal, it may be effective to compare several section of a statement concerning the same topic (within-statements comparison) rather than to compare two different statements dealing with different topics, as usually done for example in the baseline approach (within-subjects comparison). The Theme-Selection Approach aims exactly at this. By organising interviewees' statement into themes, truth tellers are not expected to show a great amount of difference (given that they remember equally well all pieces of information), whereas liars are expected to show a difference in the amount of detail they report for the theme(s) they tell the truth and that for the theme(s) about which they lie (Hypothesis 1). In short, the well-known result that liars are less detailed than truth tellers (DePaulo et al., 2003) will show-up within liars' statement only. Additionally, since: a) within-subjects measures are better than between-subjects measures for lie detection purposes (Vrij, 2016); b) liars' statement taken as a whole include both true and false details, we predicted that comparing interviewees' answers to the several themes (two themes for the present experiment, within-subjects measure) is more efficient for lie detection purposes than comparing truth tellers' and

liars' answers taken as a whole (without accounting for themes, between-subjects measure).

4.8.2 Method

Seventy-three participants took part in the experiment (61 females and 12 males) and all were offered one additional credit for their participation. However, data screening showed that five participants were outliers in terms of details reported. The new sample, on which all analyses were conducted, therefore consisted of 68 participants (56 females and 12 males).

Participants took the role of a component of a criminal organisation holding knowledge about it. The information concerning the criminal organisation was divided into two main themes: A non-critical theme, concerning the structure of the criminal organisation and its component; and a critical theme, concerning information about the leader of such organisation, his routines and his hiding place. Truth tellers were asked to report all pieces of information honestly, whereas liars were asked to tell the truth about the non-critical theme but to lie about the critical theme. Memory performance was taken into account as it may affect the experimental results. Indeed, a difference in the amount of detail when comparing answers to the two themes may be due, in addition to the veracity condition, to differences in the ability to recall the information for the two themes.

Three interviewers interviewed the participants. The interviewer started introducing herself, then asked the participant to briefly tell something about themselves, such as their hobbies and interest. This question was asked to put the interviewee at ease and to avoid that the first moment of the interview were influenced by the context (such as the presence of the camera). The interviewer then asked the following free recall question: *"As you know, our agency is investigating a criminal organization led by the Passatante clan. Tell me everything you know about this criminal organization in as much detail as possible."* This was followed by the question: *"Is there anything you would like to add?"* The interviewer then asked two follow up questions to elaborate on the two themes. *"Ok, now tell me everything you know about the structure of the organization, such as its components, roles, and activities"* (non-critical theme) and *"Ok, now tell me everything you know about the Boss of the*

criminal organization" (critical theme). Additionally, the second follow-up question was followed by another open-ended question: *"Is there anything you would like to add?."* After this, the participant was thanked, and the interview ended.

The interviews were then coded using a checklist reporting all the thirty-six pieces of information included in the original story, divided in the two themes. The aim of the coding was to count how many pieces of information were disclosed by the interviewees throughout the interview.

4.8.3 Results and discussion

Truth tellers reported more pieces of information overall than liars (between-subjects measure). Also, truth tellers reported a similar amount of detail when discussing the two themes, whereas liars revealed more pieces of information for the non-critical theme, about which they were asked to tell the truth, than when discussing the critical theme, about which they were asked to lie. Therefore Hypothesis 1 was supported. This held true both when accounting for, and not doing so, memory performance.

Hypothesis 2 could instead only be partially supported as, although the effect sizes for within-subjects comparisons were larger than those for between-subjects comparisons, both were large.

This experiment showed the potential of a new within-subjects approach, the Theme-Selection Strategy, for both discriminating truth tellers from liars (by looking at differences in the amount of information revealed for different themes) and to understand what section of a statement is probably a lie (the less detailed). Yet, this is only the first study on this strategy and results must be taken with great caution. There are indeed several factors that may have played a role, such as different memory performance for different themes, different importance allocated by the interviewees to different themes, different ease to remember/recall/talk about different themes, liars lying about all themes (therefore no difference between themes may appear), and so on. Future studies should therefore try to replicate the results we obtained in this experiment also accounting for such variables.

4.9 Study VIII: Training practitioners: An experience with Bergamo police officers[25]

4.9.1 Background

Research has shown that trainings on cues to deception (Hauch et al., 2016), and interviewing techniques (K. Colwell et al., 2009; K. Colwell, James-Kangal, Hiscock-Anisman, & Phelan, 2015; L. H. Colwell et al., 2012; Hartwig et al., 2006; Oleszkiewicz, Granhag, & Kleinman, 2017a; Vrij et al., 2015) improve the elicitation of information and lie detection. Therefore, an important step is to implement those interviewing technique that are believed to be ready for use in real-life in police investigations (Vrij & Fisher, 2016).

With the aim of starting a collaboration with practitioners, Prof. Caso and I met the Public Prosecutor of Bergamo and high rank police officers and presented a two-step project with 30 practitioners. The first step aimed at exploring whether the comparable truth baseline approach improved police officers' lie detection accuracy, which is the focus of section 4.6. The second step focused on a 2-days workshop on cues to deception and interviewing techniques, which was delivered by Prof. Caso and me. The first day of training we delivered a five-hours workshop on: i) pitfalls in lie detection; ii) opportunities in lie detection; and iii) invalid and valid cues to deception. The second day of the workshop lasted six hours and included: i) lectures on interviewing techniques; ii) role-playings, where mock suspects were interrogated by real police officers, who were asked to implement the technique they had been taught; and iii) a discussion focusing on strengths and weaknesses of interviewing techniques. This last part, together with a questionnaire filled in by police officers which focused on the perceived utility of each interviewing technique for specific crimes and feedback on the course itself, are the core part of this section.

4.9.2 Method

Training Bergamo Police officers on interviewing to detect deception

[25] The paper relating to this study is in preparation for submission as:
Caso, L., Palena, N., & Monticciolo, R. Utilità delle tecniche di intervista sviluppate in ambito accademico in contesti reali: Il parere delle Forze di Polizia di Bergamo.

The main goal of the training was to understand which interviewing techniques were perceived to be useful by Italian police officers, also according to specific crimes. Additionally, it was useful to get a feedback concerning the utility of techniques developed in the laboratory for real-life situations.

The training introduced several approaches, but it mainly focused on: i) PEACE interviewing (Bull, 2018); ii) SUE (Granhag & Hartwig, 2015); iii) Cognitive Credibility Assessment, divided in the "Encouraging the interviewee to tell more", "Increase cognitive load", and "Asking unexpected questions" approaches (Vrij et al., 2015); ACID (K. Colwell et al., 2013). Practitioners had to rate the perceived utility of each presented interviewing technique on a scale from one (not useful) to 5 (very useful) for the following crimes:

- Criminal association
- Corruption
- Organised crime
- Extortion
- Cyber fraud
- Arsons
- Wounding with intent
- Homicide/attempted homicide
- Prostitution
- Heist/Theft
- Environmental crimes
- Financial crimes
- Sexual crimes
- Kidnapping
- Terrorism
- Drug trafficking
- Fraud

The 17 crimes reported above was reduced to a list of five crimes, according to the Crime Classification Manual (Douglas, Burgess, Burgess, & Ressler, 2013).

Homicide/attempted homicide, terrorism and wounding with intent were grouped together under the label “Homicide”. Prostitution and Sexual assault were grouped under the label “Sexual Crimes”. Arson and computer crimes were left on their own. All the remaining crimes were grouped under the label “Non-lethal crimes”.

4.9.3 Results and discussions

Data were explored via a linear mixed model (LMM) analysis and no hypothesis was formulated. The design was based on a 6 (Technique: PEACE vs. SUE vs. Encouraging the interviewee to tell more vs. Imposing cognitive load vs. Asking unexpected questions vs. ACID) X 5 (Crime: Homicide vs. Sexual crimes vs. Arson vs. Computer crimes vs. Non-lethal crimes) design. Technique and Crime were the fixed, within-subjects, factors. Participants’ intercepts were the random component.

The Technique by Crime interaction was not significant, $F(20, 824.04) = 1.23$, $p = .22$, meaning that there was no sufficient evidence to conclude that practitioners considered each specific technique differently effective according to specific crimes. The main effect of Technique, $F(5, 824.21) = 12.53$, $p < .001$, and that of Crime, $F(4, 824.03) = 18.21$, $p < .001$, were both significant. The SUE technique and the “asking unexpected questions” approach were considered the most effective, whereas the ACID, the “Imposing cognitive load” and the “Encouraging the interviewee to tell more” approaches received the lowest scores.

Furthermore, the techniques were believed to be more useful for violent crimes such as Homicide and Sexual crimes, than for less violent crimes such as Computer crimes.

Finally, police officers provided, on average, high scores to the questions: “How much did you like the course” ($M = 4.56$, $SD = 0.86$); “How much do you think the course was useful for your job” ($M = 4.36$, $SD = 0.85$); and “How much would you like to participate in a new course” ($M = 4.56$, $SD = 0.82$), which all ranged from one (not at all) to five (very much). This indicates that they liked the course and found it useful.

Conclusions

Although short, this workshop shed light on several important aspects. First, we received positive feedback from police officers. This is relevant and consistent with the result that Italian police officers rarely receive any form of training from their

institutions (Zappalà & Pompedda, 2017). In short, it is likely that practitioners recognised the issue and the utility of this type of trainings. Second, there has been a relevant debate concerning the ecological validity of laboratory studies and the applicability of laboratory-developed techniques into real life investigations (although research in this regard is encouraging, see Hartwig & Bond Jr., 2014; Vrij, 2014; Vrij et al., 2015), but the positive feedback we received from practitioners may indicate that they believed that their job may benefit from lab-developed techniques. Third, the analyses showed that police officers found two interviewing approaches particularly useful: The SUE and the Asking unexpected questions approaches. Likely, this may be read in light of eliciting strong evidence (e.g.: statement-evidence inconsistency) that can be fed into the prosecutor's file. Yet, we did not explore this, thus future research is needed. Last, the techniques were perceived as useful for violent crimes in particular. This result can be explained in terms of how evidence is gathered for specific crimes. Take the example of financial crimes and fraud, usually investigated by the Financial and Economic Police here in Italy (*Guardia di Finanza*). One participant belonging to this body clearly stated that the techniques were very interesting, but not very useful in his activities as trials for the types of crime his institution fights hinge on strong evidence (e.g.: bank transactions and proofs of tax evasion). In conclusion, this experience underlined that academics and practitioners can work together toward shared goals effectively and future projects like this should continue to take place.

5 General discussion

5.1 Summary of the findings

The present dissertation focused on interviewing techniques aiming at eliciting information and at detecting deception. There is a logical sequence in the presentation of the studies included in this dissertation. In real world applications, the first step is to obtain cooperation from the sources so that they will disclose unknown information to the interviewer/intelligence agency. However, sources can lie on the information they disclose. Therefore, the following studies focused on a specific interviewing technique, the Baseline approach, and on its capabilities to elicit cues to deception and improve observers' lie detection accuracy. The results, detailed in the following sections, brought to four main conclusions. First, social influence through reciprocity does not seem to increase information elicitation in Scharff interviews. Second, the baseline approach may, if well-structured, elicit a different pattern for truth tellers vs. liars, meaning that it may work as a strategy for lie detection purposes. Third, this difference between truth tellers and liars has an inconsistent effect on observers' lie detection skills and may affect responding bias rather than observers' accuracy. Fourth, evaluating interviewees' statements on the basis of the various themes discussed may prove effective for the assessment of mixtures of truths and lies. Yet, as already mentioned throughout this dissertation, results must be taken with caution because of the low sample sizes, the subsequent risk of type II error and high uncertainty of effect size estimates. Indeed, in some cases (see sections 4.5 and 4.6 for example) the results were significant but weak in terms of evidence in support of the experimental hypothesis.

5.1.1 The use of reciprocity in Scharff Interviews

Building on research showing the efficacy of the Scharff technique for information elicitation in HUMINT interviewing (Granhag, Kleinman, et al., 2016) and on the positive effect of social influence through reciprocity in investigative interviewing (Matsumoto & Hwang, 2018), the first study of this dissertation (section 4.2) tried to improve information elicitation even more through a combination of the two. However, our results showed that reciprocity did not seem to increase information disclosure in Scharff interviewing, which is inconsistent with previous results on the

positive effect of reciprocity for social influence (Cialdini, 2001, 2016; Cialdini & Goldstein, 2004; Matsumoto & Hwang, 2018). At a first sight one may conclude that reciprocity simply did not work in such a context, which is of course a possibility. Yet, it is also possible to reach a different conclusion. Reciprocity can be of different types, and at least two examples are needed here (described in Sobel, 2005). Intrinsic reciprocity, that is based on the perceived intention of our interlocutor- our manipulation of offering water likely falls under this definition- and instrumental reciprocity, where an agent reciprocates the other only if a gain is expected. In HUMINT settings the latter is more plausible as sources disclose information to fulfil their (instrumental) needs (Neequaye & Luke, 2018). Further, it is possible that, since we offered a deal to the sources, the effect of reciprocity was already exhausted for the offer of water to have any effect. In short, promising a possible gain of university points to interviewees in exchange of their collaboration overshadowed a possible positive effect of the offer of a beverage. However, since our participants always experienced both types of reciprocity, it is difficult to disentangle the effect of each of them. Indeed, the trend we found was in the expected direction, with interviewees who shared water with the interviewer revealing more pieces of information that those who were not offered the drink. Yet, no conclusion is possible as the result was not significant. Additionally, a more social form of influence such as offering water may have an indirect effect through rapport on information disclosure and may need more time to maintain its effect (Kelly et al., 2016) than the way we manipulated it.

Future studies should therefore focus on two important manipulations. First, the type of reciprocity. An experiment may be developed so that four groups are formed: One where no reciprocity is employed vs. instrumental reciprocity vs. intrinsic (social) reciprocity vs. both types of reciprocity at the same time. Contrasts analyses will then show which one of the two forms of reciprocity is better in HUMINT settings, whether intrinsic reciprocity alone is better than no reciprocity at all and if any positive effect disappears when comparing the group with both forms of reciprocity employed vs. groups where only one form is used. Second, another experiment may test if a social form of reciprocity that is presented several times enables the interrogator to build rapport that is stable over the course of the interview and whether it is effective for information elicitation. Building on our results and on Neequaye's and Luke's (2018)

theoretical assumptions, it is possible to predict that instrumental reciprocity is better suited for HUMINT settings than intrinsic reciprocity, but that intrinsic reciprocity is better than no reciprocity at all, especially when it is presented several times over the course of the interview. Will this be the case there are clear practical applications: Offering a deal to the source is the best option if possible but developing rapport and holding it throughout the interview is also paramount. Future studies will shed light on this issue.

5.1.2 The Baseline approach

Due to the absence of strong and reliable cues to deception (DePaulo et al., 2003; Vrij, 2008), recent research focused on strategies to enhance cues to deception (Vrij, 2014; Vrij & Granhag, 2012). The Baseline approach is one of such interviewing strategies, whose advocated goal is to reduce the effect of interpersonal differences (Caso et al., 2018; Vrij, 2016). Its effect on cues to deception and observers' accuracy is dealt with in the two following sections.

5.1.2.1 Effects on cues to deception

Our first two studies focused on the possibility to elicit cues to deception using the Baseline approach. A baseline, a (non)verbal reference line of how an interviewee speaks when telling the truth to which answers to investigative questions should be compared, can be obtained in two main ways. The first one is by asking non-threatening questions, unrelated to the event under investigation (Small Talk approach), which proved ineffective (Ewens et al., 2014). A second possibility is to obtain such a baseline by asking baseline questions that are very similar to investigative questions (Comparable Truth approach). In our first study (section 4.3) we seemingly found that, when a Comparable Truth baseline is used, only liars showed differences between the two phases of the interview. In particular, liars seemed to appear as vaguer and less detailed when answering the investigative question (about which they lied) than when answering the baseline question (about which they told the truth). This seems to go in favour of the baseline approach, as an investigator could pay attention to verbal deviation from the baseline as an index of deception. Yet, evidence in support of the hypothesis was anecdotal and must be taken with caution. In our second study (section 4.4) we directly compared the two type of baselines and the results suggested that truth tellers' and liars' answers differed only

in the Comparable Truth approach. That is, when interviewing with this strategy, truth tellers seem to appear to speak with a similar amount of detail when answering the two questions, whereas liars seem to show larger deviations from the baseline. This suggests again that the Comparable Truth baseline approach may have potential.

Our results on the possible potential efficacy of the Comparable Truth baseline can be traced back to cognitive theory. As outlined by Vrij (2015a) and found in meta-analytic work (Vrij, Fisher, et al., 2017), lying is usually more mentally taxing than telling the truth and this may result in vaguer and less detailed accounts concerning the investigative section of the interview than the baseline section. The same result, however, can also be read in term of strategies (DePaulo et al., 2003; Nahari et al., 2014; Vrij, 2008) in that liars reveal less detail (for the section about which they lied) on purpose to keep the story simple and to avoid giving new leads to the investigator. However, such cues did not appear when a Small Talk baseline was employed, as both truth tellers and liars deviated from the baseline, but this was due to the difference in content of the two sections (baseline vs. target) of the interview.

Summarising, our studies suggested that a Comparable Truth baseline may be a good approach to elicit different patterns between truth tellers and liars. Yet, there are some issues with this technique as well. In first instance, it only worked for three types of clues: i) vagueness, ii) spatial details; iii) visual details. Due to the content of the story our participants had to discuss, the cues an investigator would have to look for are very tightly connected to the type of event being discussed. This is not impossible to be done in real-world, yet it is a difficult task. Second, it is easy to obtain a Comparable Truth baseline that is really truthful in the laboratory, but this task is difficult (if possible, at all) in real life. Therefore, if an investigator does have a Comparable Truth baseline, s/he may benefit from it. Yet, obtaining it is almost impossible in real life. Further, in the case the baseline itself includes some false information its efficacy drops. Additionally, it is also difficult to obtain a baseline that is of about the same detail of the event under investigation, and as easy to remember as it. The rising picture therefore provides theoretical leads (if the baseline and the target sections are really comparable, this may work for the elicitation of cues to deception) but low applicability to real-world settings. This is also clear from the

effect of the baseline approach on observers' lie detection accuracy, which is the focus of the next section.

5.1.2.2 Effects observers' accuracy

The previous studies suggested that a Comparable Truth baseline can elicit cues to deception such as differences in (non)verbal behaviour between baseline and target answers. A logical next step is to explore whether such differences are noticeable to, and increase accuracy of, observers. We ran three studies aiming at this. In the first study (section 4.5), we found partial support for the fact that a Comparable Truth is better than a Small Talk baseline in that the former seemingly brought to better total and lie accuracy rates than the latter (although the evidence was anecdotal, see section 4.5). Further, observers in the Comparable Truth condition obtained accuracy rates significantly better than chance. Observers in the Small Talk approach did not, which discourages the use of a Small Talk baseline (Frank et al., 2006; Inbau et al., 2013). Yet, the Comparable Truth baseline approach also comes with problems. It is very strictly related to the content of the story and, as said in the previous section, is very difficult to obtain in real life.

Another issue is that we obtained inconsistent results in our second study (section 4.6), where we compared police officers' accuracy rates when the Comparable Truth baseline was provided vs. when no baseline was present. First, police officers did not reach higher total accuracy rate with a baseline vs. without it. Second, the better performance in terms of lie detection accuracy can be read in terms of an effect on responding bias rather than on lie detection accuracy.

In the third study (section 4.7) we found that participants provided with a checklist obtained similar accuracy rates, regardless of the baseline condition (Comparable Truth vs. Small Talk) they were assigned to. This result, however, may be due to participants failing to follow the checklist

It is difficult to disentangle the inconsistent results we obtained; thus, further research is needed. Yet, it is possible to summarise our results as follows. First, due to the fact that people change their behaviour over time (Buller & Burgoon, 1996; Stiff et al., 1994) and in repeated interviews (Granhag & Strömwall, 2002), and considering the role of the content of the various topic discussed (Ewens et al., 2014; Palena et al.,

2018a), the Small Talk baseline affects observers' accuracy negatively as *both* truth tellers and liars show (non)verbal differences in this context. Hence, the Small Talk baseline should be dropped altogether. Second, a Comparable Truth baseline limits this issue (as liars show larger differences than truth tellers) and it is a better within-subjects measure. Yet, it does not improve observers' accuracy rates compared to when non baseline is available as observers in the former condition remain within the usual 50%-55% range found in meta-analytic work (DePaulo et al., 2003). Third, differences in accuracy rates between the two baseline conditions (Small Talk vs. Comparable Truth) can be traced back to the Small Talk reducing expected accuracy rates (DePaulo et al., 2003) rather than the Comparable Truth increasing it. Fourth, the fact that a Comparable Truth baseline elicits cues to deception but does not improve observers' accuracy can be due to the fact that such elicited cues are too weak to be noticed by observers. Fifth, the same result may be also due to observers' not paying attention to reliable cue as our participants did not receive any form of training and may have based their decisions on wrong (or unreliable) cues to deception. This may also explain our results on the inefficacy of the checklist we provided to our participants in one of our studies (Caso et al., in preparation).

In conclusion, future studies should try to disentangle this issue building on the aspects discussed in this dissertation (e.g.: associating the Comparable Truth to other techniques such as the ACID, training interviewees on reliable cues to deception, etc.), as well as prefer a different within-subjects measure over the Comparable Truth approach. Indeed, it may be that a Comparable Truth baseline is not very effective also because of the way it is obtained (comparing two different answers: one for the baseline section and one for the target section of the interview). A possible solution may be to compare interviewees' answers *within* statements rather than *between* statements. This may work also when we do not know whether a section of the interview is true or not (i.e.: a truthful baseline is missing) as the cue to look for is changes in the quantity and/or quality of interviewee's answers when discussing different topics. This is the focus of the next section.

5.1.3 Dealing with mixtures of truths and lies

People often tell a mixture of truths and lies (Leins et al., 2013) and within-subjects comparisons made between statements, such as in the Comparable Truth approach,

are problematic. These two points brought us to the development of a novel within-subject measure. We started from the idea that people categorize their experiences and knowledge in themes (Oleszkiewicz, Granhag, & Cancino Montecinos, 2014) and advocated that, if the characteristics of the themes are not too different (e.g.: they can be remembered with the same ease/difficulty) then quantitative and qualitative differences between the various themes discussed within the interview (and within-statements) should be larger for liars than for truth tellers. Indeed, truth tellers should report a similar amount of detail of a similar quality for all themes as there are no factors influencing it. Liars, on the contrary, are expected to report less detail for the theme(s) about which they lied than for the theme(s) about which they told the truth, because of strategic (Granhag & Hartwig, 2015; Nahari et al., 2014) and cognitive (Vrij, 2015a) reasons. In one of our experiments (section 4.8, Palena, Caso, & Vrij, 2019) we found exactly this. Also, the differences for liars were pronounced, and this is potentially a benefit under an applied perspective. Our results must be taken with caution though, as they are the output of just one study. Yet, if the Theme-Selection approach will consistently bring to the same output, then it may be refined with the final aim of being applied to real-world settings. Here, an investigator should look for changes in the quality and quantity of detail, regardless of whether s/he has a truthful reference point- as it is instead done in the Baseline approach. Furthermore, due to theoretical reasons reported above, the Theme-Selection strategy may prove effective to detect *what* the interviewee is lying about, not only *if* s/he is lying or not. This is a clear advantage both in criminal and HUMINT settings. Yet, we are quite far from this goal at the moment.

Changes in details over time already proved effective in airport setting (Ormerod & Dando, 2015), but a difference in themes was never explored. The difficulty, however, concerns how to categorize themes, which future research should explore. An effective tactic already mentioned in this dissertation may be to combine several techniques at the same time, in this case the Theme-Selection approach with another suitable technique, which would increase its efficacy for lie detection purposes. Yet, to conclude, it is important to note (as specified in appendix 7) that there may be other reasons for differences between the several themes being discussed, such as the sensitive nature of different themes.

5.2 Contributions: Theoretical and Practical

Our studies provide us with theoretical leads that can be used for future research. We found that reciprocity did not seem to improve the efficacy of the Scharff technique. A theoretical explanation for this result builds on the possibility that an instrumental form of reciprocity overshadows an intrinsic form of reciprocity in HUMINT settings. Hence, it is possible that different settings benefit from different types of reciprocity. Intrinsic reciprocity may play a less important role in intelligence settings as the attention of the source is on something else: Obtain a gain, without losing too much. This information dilemma is the core of the exchange between the source(s) and the investigator(s) (Neequaye & Luke, 2018). If gains and losses are at play, then intrinsic reciprocity through offers such as water may of course foster a positive atmosphere, but it may also have no effect at all on the investigator's goal: increase sources' information disclosure. Yet, recognizing the important role of instrumental offers can be strategically used by the investigator who, to increase sources' disclosure, may play on a bargain to be offered to the source. In other settings when no instrumental offer is done, intrinsic reciprocity may have a positive effect on disclosure, but future studies should explore this. Practical application of results from our first study suggest that an investigator could employ both types of reciprocity, but s/he should also be aware of the effects reported above they have. Since a positive atmosphere is paramount, investigators may put the interviewee at ease also through intrinsic reciprocity (Soufan, 2011) and then employ instrumental reciprocity to increase information disclosure. Neequaye and Luke (2018) offer an interesting interpretation of the information management dilemma sources experience, from which effective social influence strategies can be developed. Special efforts should be put to increase disclosure for guarded information (high cost of sharing and low cost of withholding) and for high-stake information (high cost of sharing and withholding). That is, if the source is offered nothing, then it is very unlikely that s/he will reveal guarded information. But if the source is offered something in exchange of disclosure (instrumental reciprocity), then s/he may weight off the situation and decide to reveal guarded information. Another important factor takes place here: Tie and affiliation with fellow criminals. When the tie between an interviewee and his/her fellow criminals is strong it is more difficult to increase information disclosure because of

the tie itself and on fidelity to one's criminal organization. This, in turn, may make information gathering difficult also when specific strategies are employed. The investigator may thus need to get trust from the interviewee and to increase rapport. Thus, before employing intrinsic reciprocity, the investigator should also contain the effect of affiliation, which may impair information disclosure.

Concerning the baseline approach, we found that a Small Talk Baseline is ineffective as both truth tellers and liars change their (non)verbal behaviour when moving from baseline to investigative answers, which replicates previous studies (Ewens et al., 2014). We therefore empirically supported the theoretical assumption that this approach is ineffective as there are fundamental differences between baseline and target answer that overshadow any effect of veracity on (non)verbal behaviour differences. Such fundamental differences include, but are not limited to, cognitive processes, stakes, emotions, memory, and content of the story. Clearly, the Small Talk baseline must be dropped altogether in real-life settings as it brings to decision errors (e.g.: false positives).

As far as for the Comparable Truth approach, we found partial support for the theoretical assumption whereby liars show larger difference between the two phases of the interview than truth tellers. This is interesting under a theoretical point of view as it can be read as an indirect support for theories of deception focusing on cognitive processes and interviewees' strategies (Nahari et al., 2014; Vrij, 2015a) predicting that lying affects cues to deception such as level of detail. Yet, under a practical point of view the Comparable Truth approach is still problematic and our idea is that it is not ready to be implemented in real life. Indeed, although we found that truth tellers did not change their (non)verbal behaviour between the two phases of the interview, whereas liars do (Palena et al., 2017), observers' accuracy did not benefit from it. It is plausible that the reason for this can be traced back to differences that are present, but too weak to be noticeable by the observers. Additionally, as another study showed (Caso, Palena, Carlessi, & Vrij, 2019), the Comparable Truth seemingly affects responding bias and pushes observers towards a lie bias. Indeed, observers in the Comparable Truth condition never reached accuracy rates higher than would be expected by meta-analytic work (DePaulo et al., 2003). In essence, this approach also brings the risk of decision errors. Another possibility for the low accuracy is that our

observers did not benefit from a Comparable Truth baseline because they were not trained about valid cues to deception (i.e.: they may have paid attention to irrelevant cues, vanishing any possible positive effect of the technique); yet, we believe that the first interpretation is more plausible. In short, the Comparable Truth baseline should not be used in real-life settings as well.

The idea that we should make within-subjects comparison is still valid, but our results went in the direction of comparing interviewees' answers *within* statements rather than *between* statements. This would also reduce problems relating to the construction of a "baseline" itself which, as previously mentioned, is difficult to obtain in real life. Our results on the Theme-Selection strategy were promising in that they supported two main theoretical assumptions: The first one, that liars show larger differences than truth tellers, a result that we replicated in the study on the Theme-Selection approach. The second one, that within-subjects measures are preferable over between-subjects measures (Vrij, 2016, 2018a), as comparing interviewees' answers within statements by focusing on the level of detail for the two themes they spoke about brought to larger differences than comparing liars' whole statements to truth tellers' whole statements. Our results also suggested that it may be possible that a baseline may be not needed. In short, the investigator may not need a truthful point and then evaluate if answers to investigative questions are quantitatively and/or qualitatively different. The investigator may pay attention to variations over time of specific cues, and associate larger variability with higher likelihood of deception. These changes over the course of an interview were also found to be effective in airport settings (Ormerod & Dando, 2015). Yet, before investigators use "variation over time" to decide whether an interviewee is lying or not, future research and replications of our results are paramount. In case the Theme-Selection strategy will hold valid, the practical value is that investigator may also be able to detect *what* information the interviewee is lying about, which can be then fed into the prosecutor's case file (for legal proceedings) or into decision-making policies in intelligence settings.

5.3 **Limitations of the studies**

There were several limitations in our studies. First, we did not measure rapport in our Scharff experiment, which could have been a good mediator. Second, all our studies were lab-studies, with clear subsequent limitations. Of course, in laboratory studies

the experimenters can exert a great control over the variables at play, but the generalizability of the results is limited. One reason for this is the effect of stakes. Although we tried to mirror real-life HUMINT and criminal interviews, all experiments were based on a role-play and the participant could leave the experiment at any moment for ethical reasons. Therefore, important factors such as bargain at play in the Scharff experiment reduced the generalizability of the studies. Stakes may also affect our result in another manner, therefore replication with high-stakes situations is recommended. However, we advise against doing that in experiments on the Comparable Truth approach due to its characteristics. Comparable means that the two phases of the interview are similar, so it implies that stakes are also similar and if they do are, then higher or lower stakes should not play an important role on the differences that are expected to appear between the two phases of the interview. Effort should instead be directed to interviews tactics such as the Theme-Selection approach. Another limitation, particularly present in this last experiment, is that the technique worked because we kept the number of pieces of information, and their ease to be remember, as similar as possible between the two phases of the interview. The technique appeared promising in this situation, but questions remain open as to whether it still will when the themes embedded in a story are different in detail, ease to be remembered, etc. For example, a liar may be labelled as a truth teller in case very few differences will appear between two themes, but such lack of differences may be due to the fact that for the theme about which s/he told the truth fewer details were available. Lastly, the experiments included a low number of participants and no a priori design analyses were done. This was caused also by the lack of previous experiments on reciprocity in Scharff techniques, the comparable truth baseline approach, and the Theme-Selection strategy. Consequently, there are at least two things to bear in mind: i) there is the risk that studies were underpowered; ii) the effect size estimates we obtained may show a large degree of uncertainty. Future research should explore all these aspects.

5.4 **Final remarks**

The present dissertation provided useful insight for future research and some practical contribution for real life settings. Nowadays, intelligence agencies need effective strategies to interview sources for the sake of national security. However,

interviewees are often not willing to provide information to the investigator and their agencies. Hence, it is paramount to conduct research in this regard, and psychology can give an important contribution in this sense. Given our results, it is too premature to tell practitioners that if they employ any form of reciprocity, they will obtain more information from their source. We do not want to give practitioners unrealistic hope. Yet, social influence tactics as intrinsic reciprocity (e.g.: offering a beverage to the interviewee) could be used by the investigators as they come at no cost. Further, there is evidence that it is already used by real-life interviewers who recognise its potential (Brandon et al., 2019; Brandon, Wells, & Seale, 2018; Soufan, 2011). The use of instrumental reciprocity must be instead seen with more caution. Indeed, the risk here is that employing this type of social influence will have high cost consequences. For example, an interviewee who is offered a penalty discount if s/he provide information may end up either by remaining silent or by providing false information. Consequently, practitioners would have high cost (the offer made to the source) but low (if any) gain. The use of instrumental reciprocity should therefore carefully considered before any interview.

The need to detect lies is also essential in criminal justice settings. There are several interviewing strategies that are believed to be effective (Caso & Palena, 2018, appendix 8), some of which are based on within-subjects comparisons and believed to be ready for use in real-life settings (Vrij, 2016; Vrij & Fisher, 2016). One approach that received great publicity is the use of a baseline (Frank et al., 2006). Yet, as the reader has seen throughout this dissertation, employing this approach is risky. The baseline approach comes with a lot of issues and even when the baseline is obtained through a comparable truth, advantages in terms of lie detection are too small and inconsistent. One of the main contributions of this work is that it shed light on the fact that it is unadvisable to use the (comparable truth) baseline approach on its own. It wold be better if practitioners either drop it altogether or use it in pair with other approaches. Also, when possible and as shown in section 4.8 and in appendix 7, it may be even better if comparisons are made within-statements rather than within-subjects. Yet, this may not be enough. Experimental findings, although very informative, may still lack of ecological validity and are often based on single, sometime unrealistic, goals (Nahari et al., 2019). As an example, research is often

either focused on information elicitation or lie detection, when the things clearly go hand by hand. This is not to say that research is superfluous. Rather, the contrary is true. Yet, efforts should be made to develop effective research design which, whenever possible, should see the collaboration of academics and practitioners.

References

Abbe, A., & Brandon, S. E. (2013). The role of rapport in investigative interviewing: A review. *Journal of Investigative Psychology and Offender Profiling, 10*(3), 237-249. doi:10.1002/jip.1386

Alison, L., Alison, E., Elntib, S., & Noone, G. (2010). *ORBIT (Observing Rapport Based Interpersonal Techniques): A Manual for Assessing and Coding Interpersonal Rapport.* Internal Document, University of Liverpool, UK.

Alison, L., Alison, E., Elntib, S., & Noone, G. (2012). *Educing information through interpersonally competent interviewing: An assessment tool for coding interviews with high value suspects. coding manual.* Internal Document.

Alison, L., Alison, E., Noone, G., Elntib, S., & Christiansen, P. (2013). Why tough tactics fail and rapport gets results: Observing Rapport-Based Interpersonal Techniques (ORBIT) to generate useful information from terrorists. *Psychology, Public Policy, and Law, 19*(4), 411. doi:10.1037/a0034564

Anderson, D. E., DePaulo, B. M., & Ansfield, M. E. (2002). The Development of Deception Detection Skill: A Longitudinal Study of Same-Sex Friends. *Personality and Social Psychology Bulletin, 28*(4), 536-545. doi:10.1177/0146167202287010

Bond, C. F., Levine, T. R., & Hartwig, M. (2015). New findings in nonverbal lie detection *Deception detection: Current challenges and new directions* (pp. 37-58). Chichester, UK: Wiley.

Bond Jr., C. F., & DePaulo, B. M. (2006). Accuracy of deception judgments. *Personality and Social Psychology Review, 10*(3), 214-234. doi:10.1207/s15327957pspr1003_2.

Bond Jr., C. F., & DePaulo, B. M. (2008). Individual differences in judging deception: Accuracy and bias. *Psychological bulletin, 134*(4), 477. doi:10.1037/0033-2909.134.4.477.

Boothby, E. J., Clark, M. S., & Bargh, J. A. (2014). Shared Experiences Are Amplified. *Psychological Science, 25*(12), 2209-2216. doi:10.1177/0956797614551162

Bowman, M. N. (2010). Truth or consequences: Self-incriminating statements and informant veracity. *NML Rev., 40*, 225.

Brandon, S. E. (2011). Impacts of psychological science on national security agencies post-9/11. *American Psychologist, 66*(6), 495.

Brandon, S. E. (2014). Towards a Science of Interrogation. *Applied Cognitive Psychology, 28*(6), 945-946. doi:10.1002/acp.3090

Brandon, S. E., Arthur, J. C., Ray, D. G., Meissner, C. A., Kleinman, S. M., Russano, M. B., & Wells, S. (2019). The High-Value Detainee Interrogation Group (HIG): Inception, evolution, and impact. In M. A. Staal & S. C. Harvey (Eds.), *Operational Psychology: A New Field to Support National Security and Public Safety*: Praeger.

Brandon, S. E., Wells, S., & Seale, C. (2018). Science-based interviewing: Information elicitation. *Journal of Investigative Psychology and Offender Profiling, 15*(2), 133-148. doi:10.1002/jip.1496

Brandt, D. R., Miller, G. R., & Hocking, J. E. (1980a). Effects of self-monitoring and familiarity on deception detection. *Communication Quarterly, 28*(3), 3-10. doi:10.1080/01463378009369370

Brandt, D. R., Miller, G. R., & Hocking, J. E. (1980b). The truth deception attribution: Effects of familiarity on the ability of observers to detect deception. *Human Communication Research, 6*(2), 99-110. doi:10.1111/j.1468-2958.1980.tb00130.x

Bull, R. (2018). *PEACE-ful Interviewing/Interrogation: What Research Can Tell Us.* Paper presented at the Diversity in Harmony–Insights from Psychology: Proceedings of the 31st International Congress of Psychology.

Bull, R., & Soukara, S. (2010). Four studies of what really happens in police interviews. In G. D. Lassiter & C. A. Meissner (Eds.), *Police interrogations and false confessions: Current research, practice, and policy recommendations* (pp. 81-95). Washington, DC: American Psychological Association.

Buller, D. B., & Burgoon, J. K. (1996). Interpersonal deception theory. *Communication theory, 6*(3), 203-242.

Buller, D. B., Strzyzewski, K. D., & Hunsaker, F. G. (1991). Interpersonal deception: II. The inferiority of conversational participants as deception detectors. *Communication Monographs, 58*(1), 25-40. doi:10.1080/03637759109376212

Burgoon, J. K., & Buller, D. B. (1994). Interpersonal deception: III. Effects of deceit on perceived communication and nonverbal behavior dynamics. *Journal of Nonverbal behavior, 18*(2), 155-184.

Cabral, L., Ozbay, E. Y., & Schotter, A. (2014). Intrinsic and instrumental reciprocity: An experimental study. *Games and Economic Behavior, 87*, 100-121. doi:10.1016/j.geb.2014.05.001

Caso, L., Gnisci, A., Vrij, A., & Mann, S. (2005). Processes underlying deception: an empirical analysis of truth and lies when manipulating the stakes. *Journal of Investigative Psychology and Offender Profiling, 2*(3), 195-202. doi::10.1002/jip.32

Caso, L., Maricchiolo, F., Bonaiuto, M., Vrij, A., & Mann, S. (2006). The impact of deception and suspicion on different hand movements. *Journal of Nonverbal behavior, 30*(1), 1-19. doi:10.1007/s10919-005-0001-z.

Caso, L., Maricchiolo, F., Livi, S., Vrij, A., & Palena, N. (2018). Factors affecting Observers' Accuracy when Assessing Credibility: The Effect of the Interaction between Media, Senders' Competence and Veracity. *The Spanish Journal of Psychology, 21*, 1-10. doi:10.1017/sjp.2018.54.

Caso, L., Morganti, F., Palena, N., & Vrij, A. (under review). Dual-Task, working memory and deception.

Caso, L., & Palena, N. (2018). Tecniche di interrogatorio con soggetti adulti: una rassegna internazionale. *Rassegna Italiana di Criminologia, 2*, 130-140.

Caso, L., Palena, N., Carlessi, E., & Vrij, A. (2019). Police accuracy in truth/lie detection when judging baseline interviews. *Psychiatry, Psychology and Law*, 1-10. doi:10.1080/13218719.2019.1642258

Caso, L., Palena, N., Vrij, A., & Gnisci, A. (2019). Observers' performance at evaluating truthfulness when provided with Comparable Truth or Small Talk Baselines. *Psychiatry, Psychology and Law*. doi:10.1080/13218719.2018.1553471.

Caso, L., Palena, N., Vrij, A., & Melocchi, L. (in preparation). Implementing the use of a checklist with comparable truth and small talk baselines: The effect on laypersons' lie detection accuracy.

Caso, L., & Vrij, A. (2009). *L'interrogatorio giudiziario e l'intervista investigativa: metodi e tecniche di conduzione (Judicial interrogation and investigative interviewing: methods and techniques)*. Bologna: Il Mulino.

Caso, L., Vrij, A., Mann, S., & De Leo, G. (2006). Deceptive responses: The impact of verbal and non-verbal countermeasures. *Legal and Criminological Psychology, 11*(1), 99-111. doi:10.1348/135532505X49936

Cheng, K. H. W., & Broadhurst, R. (2005). The Detection of Deception: The Effects of First and Second Language on Lie Detection Ability. *Psychiatry, Psychology and Law, 12*(1), 107-118. doi:10.1375/pplt.2005.12.1.107

Christ, S. E., Van Essen, D. C., Watson, J. M., Brubaker, L. E., & McDermott, K. B. (2009). The Contributions of Prefrontal Cortex and Executive Control to Deception: Evidence from Activation Likelihood Estimate Meta-analyses. *Cerebral Cortex, 19*(7), 1557-1566. doi:10.1093/cercor/bhn189

Cialdini, R. (2001). *Influence: Science and practice (4th ed.)* (Vol. 3). Boston, MA: Allyn and Bacon.

Cialdini, R. (2016). *Pre-Suasion: A revolutionary way to influence and persuade*. New York, NY: Simon and Schuster.

Cialdini, R., & Goldstein, N. J. (2004). Social Influence: Compliance and Conformity. *Annual Review of Psychology, 55*(1), 591-621. doi:10.1146/annurev.psych.55.090902.142015

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences. 2nd Ed.*: Hillsdale, NJ: Erlbaum Associates.

Colwell, K., Hiscock-Anisman, C., & Fede, J. (2013). Assessment Criteria Indicative of Deception: An Example of the New Paradigm of Differential Recall Enhancement. In B. S. Cooper, D. Griesel, & M. Ternes (Eds.), *Applied Issues in Investigative Interviewing, Eyewitness Memory, and Credibility Assessment* (pp. 259-291). New York, NY: Springer New York.

Colwell, K., Hiscock-Anisman, C., Memon, A., Colwell, L. H., Taylor, L., & Woods, D. (2009). Training in Assessment Criteria Indicative of Deception to Improve Credibility Judgments. *Journal of Forensic Psychology Practice, 9*(3), 199-207. doi:10.1080/15228930902810078

Colwell, K., Hiscock-Anisman, C., Memon, A., Rachel, A., & Colwell, L. H. (2007). Vividness and spontaneity of statement detail characteristics as predictors of witness credibility. *American Journal of Forensic Psychology, 25*(1), 5.

Colwell, K., Hiscock-Anisman, C. K., Memon, A., Taylor, L., & Prewett, J. (2007). Assessment Criteria Indicative of Deception (ACID): An integrated system of investigative interviewing and detecting deception. *Journal of Investigative Psychology and Offender Profiling, 4*(3), 167-180.

Colwell, K., Hiscock, C. K., & Memon, A. (2002). Interviewing techniques and the assessment of statement credibility. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition, 16*(3), 287-300.

Colwell, K., James-Kangal, N., Hiscock-Anisman, C., & Phelan, V. (2015). Should Police Use ACID? Training and Credibility Assessment Using Transcripts Versus

Recordings. *Journal of Forensic Psychology Practice, 15*(3), 226-247. doi:10.1080/15228932.2015.1035187

Colwell, L. H., Colwell, K., Hiscock-Anisman, C. K., Hartwig, M., Cole, L., Werdin, K., & Youschak, K. (2012). Teaching Professionals to Detect Deception: The Efficacy of a Brief Training Workshop. *Journal of Forensic Psychology Practice, 12*(1), 68-80. doi:10.1080/15228932.2012.629592

Cuddy, A. J., Fiske, S. T., & Glick, P. (2008). Warmth and competence as universal dimensions of social perception: The stereotype content model and the BIAS map. *Advances in experimental social psychology, 40*, 61-149. doi:10.1016/S0065-2601(07)00002-0

Davis, M., & Hadiks, D. (1995). Demeanor and credibility *Semiotica* (Vol. 106, pp. 5).

De Leo, G., Scali, M., & Caso, L. (2005). *La testimonianza. Problemi, metodi e strumenti nella valutazione dei testimoni*. Bologna: Il mulino.

Debey, E., Verschuere, B., & Crombez, G. (2012). Lying and executive control: An experimental investigation using ego depletion and goal neglect. *Acta Psychologica, 140*(2), 133-141. doi:10.1016/j.actpsy.2012.03.004

Department of the Army. (2006). *Human intelligence collector operations (Field Manual 2-22.3)*. (1494811952). Wasinghton, DC: Headquarters, Department of the Army.

DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological bulletin, 129*(1), 74-118. doi:10.1037/0033-2909.129.1.74.

Douglas, J. E., Burgess, A. W., Burgess, A. G., & Ressler, R. K. (2013). *Crime classification manual: A standard system for investigating and classifying violent crime*: John Wiley & Sons.

Dujmovic, N. (2005). Fifty years of studies in Intelligence: Building an 'Intelligence literature'. *Studies in Intelligence, 49*(4), 6-13.

Ekman, P. (2001). *Telling Lies. Clues to deceit in the marketplace, politics, and marriage.* New York, NY: Norton.

Ekman, P., & Friesen, W. V. (1969). Nonverbal leakage and clues to deception. *Psychiatry, 32*(1), 88-106.

Evans, J. R., Houston, K. A., Meissner, C. A., Ross, A. B., LaBianca, J. R., Woestehoff, S. A., & Kleinman, S. M. (2014). An Empirical Evaluation of Intelligence-gathering Interrogation Techniques from the United States Army Field Manual. *Applied Cognitive Psychology, 28*(6), 867-875. doi:10.1002/acp.3065

Evans, J. R., Meissner, C. A., Brandon, S. E., Russano, M. B., & Kleinman, S. M. (2010). Criminal versus HUMINT interrogations: The importance of psychological science to improving interrogative practice. *The Journal of Psychiatry & Law, 38*(1-2), 215-249.

Evans, J. R., Meissner, C. A., Ross, A. B., Houston, K. A., Russano, M. B., & Horgan, A. J. (2013). Obtaining guilty knowledge in human intelligence interrogations: Comparing accusatorial and information-gathering approaches with a novel experimental paradigm. *Journal of Applied Research in Memory and Cognition, 2*(2), 83-88.

Evans, J. R., Michael, S. W., Meissner, C. A., & Brandon, S. E. (2013). Validating a new assessment method for deception detection: Introducing a Psychologically Based Credibility Assessment Tool. *Journal of Applied Research in Memory and Cognition, 2*(1), 33-41. doi:10.1016/j.jarmac.2013.02.002

Ewens, S., Vrij, A., Jang, M., & Jo, E. (2014). Drop the small talk when establishing baseline behaviour in interviews. *Journal of Investigative Psychology and Offender Profiling, 11*(3), 244-252. doi:doi.org/10.1002/jip.1414.

Ewens, S., Vrij, A., Leal, S., Mann, S., Jo, E., Shaboltas, A., . . . Houston, K. (2016). Using the Model Statement to Elicit Information and Cues to Deceit from Native Speakers, Non-native Speakers and Those Talking Through an Interpreter. *Applied Cognitive Psychology, 30*(6), 854-862. doi:doi:10.1002/acp.3270

Fallon, M. (2014). Collaboration Between Practice and Science Will Enhance Interrogations. *Applied Cognitive Psychology, 28*(6), 949-950. doi:doi:10.1002/acp.3091

Feeley, T. H., deTurck, M. A., & Young, M. J. (1995). Baseline familiarity in lie detection. *Communication Research Reports, 12*(2), 160-169. doi:10.1080/08824099509362052

Fein, R. A., Lehner, P., & Vossekuil, B. (2006). *Educing information-interrogation: Science and art, foundations for the future*. NATIONAL DEFENSE INTELLIGENCE COLL WASHINGTON DC CENTER FOR STRATEGIC INTELLIGENCE RESEARCH.

Festinger, L. (1954). A Theory of Social Comparison Processes. *Human Relations, 7*(2), 117-140. doi:10.1177/001872675400700202

Fisher, R. P., & Geiselman, R. E. (1992). *Memory enhancing techniques for investigative interviewing: The cognitive interview*. Springfield, IL: Charles C Thomas Publisher.

Fiske, S. T., Cuddy, A. J., & Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends in cognitive sciences, 11*(2), 77-83. doi:10.1016/j.tics.2006.11.005

Frank, M. G., Maroulis, A., & Griffin, D. J. (2013). The voice. In D. Matsumoto, M. G. Frank, & H. S. Hwang (Eds.), *Nonverbal communication Science and Applications* (pp. 53-74). Thousand Oaks, CA: Sage.

Frank, M. G., Yarbrough, J. D., & Ekman, P. (2006). Investigative interviewing and the detection of deception. In T. Williamson (Ed.), *Investigative interviewing: Rights, research and regulation* (pp. 229-255). Cullompton, Devon: Willian.

Gilovich, T., Savitsky, K., & Medvec, V. H. (1998). The illusion of transparency: Biased assessments of others' ability to read one's emotional states. *Journal of personality and social psychology, 75*(2), 332-346.

Global Deception Research Team. (2006). A world of lies. *Journal of cross-cultural psychology, 37*(1), 60-74.

Global Terrorism Index. (2017). Measuring and understanding the impact of terrorism. Retrieved from http://visionofhumanity.org/app/uploads/2017/11/Global-Terrorism-Index-2017.pdf

Gnisci, A., Caso, L., & Vrij, A. (2010). Have you made up your story? The effect of suspicion and liars' strategies on reality monitoring. *Applied Cognitive Psychology, 24*(6), 762-773. doi:doi:10.1002/acp.1584

Goodman-Delahunty, J., & Howes, L. M. (2016). Social persuasion to develop rapport in high-stakes interviews: qualitative analyses of Asian-Pacific practices. *Policing and Society, 26*(3), 270-290. doi:10.1080/10439463.2014.942848

Gouldner, A. W. (1960). The Norm of Reciprocity: A Preliminary Statement. *American Sociological Review, 25*(2), 161-178. doi:10.2307/2092623

Granhag, P. A. (2010). *The Strategic Use of Evidence (SUE) technique: A scientific perspective.* Paper presented at the High Value Detainee Interrogation Group (HIG; FBI). HIG Research Symposium: Interrogation in the European Union., Washington, DC.

Granhag, P. A., & Hartwig, M. (2015). The strategic use of evidence technique: A conceptual overview. In P. A. Granhag, V. Aldert, & B. Verschuere (Eds.), *Detecting Deception: Current Challenges and Cognitive Approaches* (pp. 231-251). Chichester, UK: Wiley.

Granhag, P. A., Kleinman, S. M., & Oleszkiewicz, S. (2016). The Scharff technique: On how to effectively elicit intelligence from human sources. *International Journal of Intelligence and CounterIntelligence, 29*(1), 132-150. doi:10.1080/08850607.2015.1083341

Granhag, P. A., & Luke, T. J. (2018). How to interview to elicit concealed information: Introducing the Shift-of-Strategy (SoS) approach. In J. P. Rosenfeld (Ed.), *Detecting Concealed Information and Deception: Recent Developments* (pp. 272-295). San Diego, USA: Academic Press.

Granhag, P. A., Montecinos, S. C., & Oleszkiewicz, S. (2015). Eliciting intelligence from sources: The first scientific test of the Scharff technique. *Legal and Criminological Psychology, 20*(1), 96-113. doi:10.1111/lcrp.12015

Granhag, P. A., Oleszkiewicz, S., & Kleinman, S. (2016). Eliciting information from small cells of sources. *Journal of Policing, Intelligence and Counter Terrorism, 11*(2), 143-162. doi:10.1080/18335330.2016.1215507

Granhag, P. A., Oleszkiewicz, S., Strömwall, L. A., & Kleinman, S. M. (2015). Eliciting intelligence with the Scharff technique: Interviewing more and less cooperative and capable sources. *Psychology, Public Policy, and Law, 21*(1), 100-110. doi:10.1037/law0000030

Granhag, P. A., & Strömwall, L. A. (2002). Repeated interrogations: verbal and non-verbal cues to deception. *Applied Cognitive Psychology, 16*(3), 243-257. doi:10.1002/acp.784

Granhag, P. A., Vrij, A., & Meissner, C. A. (2014). Information Gathering in Law Enforcement and Intelligence Settings: Advancing Theory and Practice. *Applied Cognitive Psychology, 28*(6), 815-816. doi:doi:10.1002/acp.3093

Granhag, P. A., Vrij, A., & Verschuere, B. (2015). *Detecting deception: Current challenges and cognitive approaches*: John Wiley & Sons.

Hartwig, M., Anders Granhag, P., & Strömwall, L. A. (2007). Guilty and innocent suspects' strategies during police interrogations. *Psychology, Crime & Law, 13*(2), 213-227. doi:10.1080/10683160600750264

Hartwig, M., & Bond Jr, C. F. (2011). Why do lie-catchers fail? A lens model meta-analysis of human lie judgments. *Psychological bulletin, 137*(4), 643-659. doi:10.1037/a0023589

Hartwig, M., & Bond Jr., C. F. (2014). Lie Detection from Multiple Cues: A Meta-analysis. *Applied Cognitive Psychology, 28*(5), 661-676. doi:10.1002/acp.3052

Hartwig, M., Granhag, P. A., & Luke, T. (2014). Strategic use of evidence during investigative interviews: The state of the science. In D. C. Raskin, C. R. Honts, & J. C. Kircher (Eds.), *Credibility Assessment* (pp. 1-36). Waltham, MA: Academic Press.

Hartwig, M., Granhag, P. A., Strömwall, L. A., & Doering, N. (2010). Impression and information management: On the strategic self-regulation of innocent and guilty

suspects. *The Open Criminology Journal, 3*(1), 10-16. doi:10.2174/1874917801003010010

Hartwig, M., Granhag, P. A., Strömwall, L. A., & Kronkvist, O. (2006). Strategic Use of Evidence During Police Interviews: When Training to Detect Deception Works. *Law and human behavior, 30*(5), 603-619. doi:10.1007/s10979-006-9053-9

Hartwig, M., Meissner, C. A., & Semel, M. D. (2014). Human intelligence interviewing and interrogation: Assessing the challenges of developing an ethical, evidence-based approach. In R. Bull (Ed.), *Investigative interviewing* (pp. 209-228): Springer.

Hauch, V., Sporer, S. L., Michael, S. W., & Meissner, C. A. (2016). Does training improve the detection of deception? A meta-analysis. *Communication Research, 43*(3), 283-343. doi:0.1177/0093650214534974

High Value Detainee Interrogation Group. (2016). *Interrogation Best Practices*. Retrieved from https://www.fbi.gov/file-repository/hig-report-august-2016.pdf/view

Il Sole 24 Ore. (2018). Anonymous contro il governo, attacco hacker il 5 novembre. Retrieved from https://www.ilsole24ore.com/art/tecnologie/2018-11-01/anonymous-contro-governo-attacco-hacker-5-novembre-111050.shtml?uuid=AEgYTLZG

Inbau, F. E., Reid, J. E., Buckley, J. P., & Jayne, B. C. (2013). *Criminal interrogation and confessions (5th ed.)*. Burlington, MA: Jones & Bartlett Learning.

Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological Review, 88*(1), 67-85. doi:10.1037/0033-295X.88.1.67

Jundi, S., Vrij, A., Mann, S., Hillman, J., & Hope, L. (2015). 'I'm a photographer, not a terrorist': the use of photography to detect deception. *Psychology, Crime & Law, 21*(2), 114-126. doi:10.1080/1068316X.2014.935776

Kelln, B. R. C., & Ellard, J. H. (1999). An Equity Theory Analysis of the Impact of Forgiveness and Retribution on Transgressor Compliance. *Personality and Social Psychology Bulletin, 25*(7), 864-872. doi:10.1177/0146167299025007008

Kelly, C. E., Miller, J. C., & Redlich, A. D. (2016). The dynamic nature of interrogation. *Law and human behavior, 40*(3), 295-309. doi:10.1037/lhb0000172

Kleinberg, B., Toolen, Y., Vrij, A., Arntz, A., & Verschuere, B. (2018). Automated verbal credibility assessment of intentions: The model statement technique and predictive modeling. *Applied Cognitive Psychology, 32*(3), 354-366. doi:10.1002/acp.3407

Kleinke, C. L. (1986). Gaze and eye contact: A research review. *Psychological bulletin, 100*(1), 78-100. doi:10.1037/0033-2909.100.1.78

Knieps, M., Granhag, P. A., & Vrij, A. (2013). Back to the future: Asking about mental images to discriminate between true and false intentions. *The Journal of psychology, 147*(6), 619-640.

Köhnken, G. (2004). Statement validity analysis and the 'detection of the truth'. In P. A. Granhag & L. Strömwall (Eds.), *The detection of deception in forensic contexts* (pp. 41-63). Cambridge, England: Cambridge University Press.

Kraut, R. (1980). Humans as Lie Detectors. *Journal of Communication, 30*(4), 209-218. doi:10.1111/j.1460-2466.1980.tb02030.x

Leal, S., Vrij, A., Deeb, H., & Jupe, L. (2018). Using the model statement to elicit verbal differences between truth tellers and liars: The benefit of examining core and

peripheral details. *Journal of Applied Research in Memory and Cognition*. doi:10.1016/j.jarmac.2018.07.001

Leal, S., Vrij, A., Warmelink, L., Vernham, Z., & Fisher, R. P. (2015). You cannot hide your telephone lies: Providing a model statement as an aid to detect deception in insurance telephone calls. *Legal and Criminological Psychology, 20*(1), 129-146. doi:10.1111/lcrp.12017

Leary, T. (1955). The Theory and Measurement Methodology of Interpersonal Communication. *Psychiatry, 18*(2), 147-161. doi:10.1080/00332747.1955.11023002

Leins, D. A., Fisher, R. P., & Ross, S. J. (2013). Exploring liars' strategies for creating deceptive reports. *Legal and Criminological Psychology, 18*(1), 141-151. doi:10.1111/j.2044-8333.2011.02041.x

Leins, D. A., Fisher, R. P., & Vrij, A. (2012). Drawing on Liars' Lack of Cognitive Flexibility: Detecting Deception Through Varying Report Modes. *Applied Cognitive Psychology, 26*(4), 601-607. doi:doi:10.1002/acp.2837

Leins, D. A., Fisher, R. P., Vrij, A., Leal, S., & Mann, S. (2011). Using sketch drawing to induce inconsistency in liars. *Legal and Criminological Psychology, 16*(2), 253-265.

Levine, T. R. (2014). Truth-Default Theory (TDT):A Theory of Human Deception and Deception Detection. *Journal of Language and Social Psychology, 33*(4), 378-392. doi:10.1177/0261927x14535916

Loftus, E. F. (2011). Intelligence gathering post-9/11. *American Psychologist, 66*(6), 532.

Luke, T. J. (2019). Lessons from Pinocchio: Cues to deception may be highly exaggerated. *Perspectives on Psychological Science, 14*(4), 646-671.

Mac Giolla, E., Granhag, P. A., & Liu-Jönsson, M. (2013). Markers of good planning behavior as a cue for separating true and false intent. *PsyCh Journal, 2*(3), 183-189. doi:10.1002/pchj.36

Maguire, M., & John, T. (1995). *Intelligence, surveillance and informants: Integrated approaches*.

Mann, S., Vrij, A., Shaw, D. J., Leal, S., Ewens, S., Hillman, J., . . . Fisher, R. P. (2013). Two heads are better than one? How to effectively use two interviewers to elicit cues to deception. *Legal and Criminological Psychology, 18*(2), 324-340. doi:10.1111/j.2044-8333.2012.02055.x

Matsumoto, D., & Hwang, H. C. (2018). Social influence in investigative interviews: The effects of reciprocity. *Applied Cognitive Psychology, 32*(2), 163-170. doi:10.1002/acp.3390

May, L., Granhag, P. A., & Oleszkiewicz, S. (2014). Eliciting Intelligence Using the Scharff-Technique: Closing in on the Confirmation/Disconfirmation-Tactic. *Journal of Investigative Psychology and Offender Profiling, 11*(2), 136-150. doi:10.1002/jip.1412

McCornack, S. A., Morrison, K., Paik, J. E., Wisner, A. M., & Zhu, X. (2014). Information manipulation theory 2: A propositional theory of deceptive discourse production. *Journal of Language and Social Psychology, 33*(4), 348-377.

Miller, W. R., & Rollnick, S. (2012). *Motivational interviewing: Helping people change*. New York: Guilford press.

Mitchell, R. W. (1986). *A framework for discussing deception*. New Your, NY: State University of New York.

Nahari, G., Ashkenazi, T., Fisher, R. P., Granhag, P.-A., Hershkowitz, I., Masip, J., . . . Vrij, A. (2019). 'Language of lies': Urgent issues and prospects in verbal lie detection research. *Legal and Criminological Psychology, 24*(1), 1-23. doi:doi:10.1111/lcrp.12148

Nahari, G., Vrij, A., & Fisher, R. P. (2014). Exploiting liars' verbal strategies by examining the verifiability of details. *Legal and Criminological Psychology, 19*(2), 227-239. doi:10.1111/j.2044-8333.2012.02069.x.

NATO. (2018). *Glossary of terms and definitions - AAP-06 Edition 2018*. Retrieved from

Neequaye, D. A., & Luke, T. J. (2018). Management of disclosure-costs in intelligence interviews. *PsyArXiv. November, 28*. doi:10.31234/osf.io/tfp2c

Oleszkiewicz, S. (2016). *Eliciting human intelligence: A conceptualization and empirical testing of the Scharff technique.*

Oleszkiewicz, S., Granhag, P. A., & Cancino Montecinos, S. (2014). The Scharff-technique: Eliciting intelligence from human sources. *Law and human behavior, 38*(5), 478-489. doi:10.1037/lhb0000085

Oleszkiewicz, S., Granhag, P. A., & Kleinman, S. M. (2014). On Eliciting Intelligence from Human Sources: Contextualizing the Scharff-Technique. *Applied Cognitive Psychology, 28*(6), 898-907. doi:10.1002/acp.3073

Oleszkiewicz, S., Granhag, P. A., & Kleinman, S. M. (2017a). Eliciting information from human sources: Training handlers in the Scharff technique. *Legal and Criminological Psychology, 22*(2), 400-419. doi:10.1111/lcrp.12108

Oleszkiewicz, S., Granhag, P. A., & Kleinman, S. M. (2017b). Gathering human intelligence via repeated interviewing: further empirical tests of the Scharff technique. *Psychology, Crime & Law, 23*(7), 666-681. doi:10.1080/1068316X.2017.1296150

Ormerod, T. C., & Dando, C. J. (2015). Finding a needle in a haystack: Toward a psychologically informed method for aviation security screening. *Journal of Experimental Psychology: General, 144*(1), 76-84. doi:10.1037/xge0000030

Palena, N., Caso, L., Carlotto, G., De Mizio, L., & Marciali, M. (2017). Efficacia della tecnica di baselining nella valutazione della credibilità (Efficacy of the baseline technique when assessing credibility). *Giornale italiano di psicologia, 44*(4), 905-916. doi:10.1421/88773

Palena, N., Caso, L., Granhag, P. A., Orthey, R., Monticciolo, R., & Vrij, A. (in preparation). Reciprocity: Does it helps in Scharff interviews?

Palena, N., Caso, L., & Vrij, A. (2019). Detecting Lies via a Theme-Selection Strategy. *Frontiers in Psychology, 9*(2775). doi:10.3389/fpsyg.2018.02775

Palena, N., Caso, L., Vrij, A., & Orthey, R. (2018a). Detecting deception through small talk and comparable truth baselines. *Journal of Investigative Psychology and Offender Profiling, 15*(2), 124-132. doi:10.1002/jip.1495

Palena, N., Caso, L., Vrij, A., & Orthey, R. (2018b). *Interviewing suspects with small talk and comparable truth baselines.* Paper presented at the EAPL, Turky, Finland.

Porter, S., & ten Brinke, L. (2008). Reading Between the Lies:Identifying Concealed and Falsified Emotions in Universal Facial Expressions. *Psychological Science, 19*(5), 508-514. doi:10.1111/j.1467-9280.2008.02116.x

Redlich, A. D. (2007). Military Versus Police Interrogations: Similarities and Differences. *Peace and Conflict: Journal of Peace Psychology, 13*(4), 423-428. doi:10.1080/10781910701665741

Rind, B., & Strohmetz, D. (1999). Effect on Restaurant Tipping of a Helpful Message Written on the Back of Customers' Checks. *Journal of Applied Social Psychology, 29*(1), 139-144. doi:10.1111/j.1559-1816.1999.tb01378.x

Rosenfeld, J. P. (2018). *Detecting concealed information and deception: Recent developments*: Academic Press.

Senate Select Committee in Intelligence. (2014). *The United States senate report on CIA detention interrogation program*.

Shaw, D. J., Vrij, A., Leal, S., Mann, S., Hillman, J., Granhag, P. A., & Fisher, R. P. (2015). Mimicry and Investigative Interviewing: Using Deliberate Mimicry to Elicit Information and Cues to Deceit. *Journal of Investigative Psychology and Offender Profiling, 12*(3), 217-230. doi:10.1002/jip.1438

Shumate, S., & Borum, R. (2006). Psychological support to defense counterintelligence operations. *Military Psychology, 18*(4), 283-296.

Sinnigen, W. G. (1961). The roman secret service. *The Classical Journal, 57*(2), 65-72.

Sistema di Informazione per la Sicurezza della Repubblica. (2018). Attacco hacker PEC: Governo già al lavoro da tempo. Retrieved from http://www.sicurezzanazionale.gov.it/sisr.nsf/archivio-notizie/attacco-hacker-pec-governo-gia-al-lavoro-da-tempo.html

Sobel, J. (2005). Interdependent preferences and reciprocity. *Journal of economic literature, 43*(2), 392-436. doi:10.1257/0022051054661530

Soufan, A. (2011). *The black banners: Inside the hunt for Al Qaeda*: Penguin UK.

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers, 31*(1), 137-149. doi:10.3758/bf03207704

Steller, M. (1989). Recent developments in statement analysis *Credibility assessment* (pp. 135-154): Springer.

Stiff, J., Corman, S., Krizek, B., & Snider, E. (1994). Individual Differences and Changes in Nonverbal Behavior:Unmasking the Changing Faces of Deception. *Communication Research, 21*(5), 555-581. doi:10.1177/009365094021005001

Street, C. N. H. (2015). ALIED: Humans as adaptive lie detectors. *Journal of Applied Research in Memory and Cognition, 4*(4), 335-343. doi:https://doi.org/10.1016/j.jarmac.2015.06.002

Swanner, J. K., Meissner, C. A., Atkinson, D. J., & Dianiska, R. E. (2016). Developing diagnostic, evidence-based approaches to interrogation. *Journal of Applied Research in Memory and Cognition, 5*(3), 295-301.

Tekin, S., Granhag, P. A., Strömwall, L. A., & Vrij, A. (2016). How to make perpetrators in denial disclose more information about their crimes. *Psychology, Crime & Law, 22*(6), 561-580. doi:10.1080/1068316X.2016.1168425

Tickle-Degnen, L., & Rosenthal, R. (1990). The nature of rapport and its nonverbal correlates. *Psychological inquiry, 1*(4), 285-293. doi:10.1207/s15327965pli0104_1

Toliver, R. (1997). The interrogator. *Pennsylvania: Schiffer Publishing*.

Undeutsch, U. (1967). Beurteilung der glaubhaftigkeit von aussagen. *Handbuch der psychologie, 11*, 26-181.

United States Government Accountability Office (US GAO). (2011). *Aviation Security: TSA is taking steps to valdidate the science underlying its passengers behavior detection program, but efforts may not be comprehensive*.

VanVoorhis, C. W., & Morgan, B. L. (2007). Understanding power and rules of thumb for determining sample sizes. *Tutorials in quantitative methods for psychology, 3*(2), 43-50.

Varouhakis, M. (2013). What is being published in intelligence? A study of two scholarly journals. *International Journal of Intelligence and CounterIntelligence, 26*(1), 176-189.

Vernham, Z., Vrij, A., & Leal, S. (2018). Collective interviewing: The use of a model statement to differentiate between pairs of truth-tellers and pairs of liars. *Legal and Criminological Psychology, 23*(2), 214-229. doi:doi:10.1111/lcrp.12136

Vrij, A. (2006). Challenging interviewees during interviews: The potential effects on lie detection. *Psychology, Crime & Law, 12*(2), 193-206. doi:10.1080/10683160512331331319

Vrij, A. (2008). *Detecting lies and deceit: Pitfalls and opportunities, second edition*. Chichester: John Wiley and Sons.

Vrij, A. (2014). Interviewing to detect deception. *European Psychologist, 19*, 184-194. doi:10.1027/1016-9040/a000201.

Vrij, A. (2015a). A cognitive approach to lie detection. In P. A. Granhag, A. Vrij, & B. Verschuere (Eds.), *Detecting Deception: Current Challenges and Cognitive Approaches* (pp. 205-230). Chichester: John Wiley and Sons.

Vrij, A. (2015b). Verbal lie detection tools: Statement validity analysis, reality monitoring and scientific content analysis. In P. A. Granhag, A. Vrij, & B. Verschuere (Eds.), *Detecting Deception: Current Challenges and Cognitive Approaches* (pp. 3-35). Chichester: John Wiley and Sons.

Vrij, A. (2016). Baselining as a Lie Detection Method. *Applied Cognitive Psychology, 30*(6), 1112-1119. doi:10.1002/acp.3288

Vrij, A. (2018a). Deception and truth detection when analyzing nonverbal and verbal cues. *Applied Cognitive Psychology*. doi:10.1002/acp.3457

Vrij, A. (2018b). Verbal Lie Detection Tools From an Applied Perspective. In J. P. Rosenfeld (Ed.), *Detecting Concealed Information and Deception: Recent Developments* (pp. 297-327). San Diego, USA: Academic Press.

Vrij, A., Akehurst, L., Soukara, S., & Bull, R. (2002). Will the truth come out? The effect of deception, age, status, coaching, and social skills on CBCA scores. *Law and human behavior, 26*(3), 261-283.

Vrij, A., Akehurst, L., Soukara, S., & Bull, R. (2004). Let me inform you how to tell a convincing story: CBCA and reality monitoring scores as a function of age, coaching, and deception. *Canadian Journal of Behavioural Science / Revue canadienne des sciences du comportement, 36*(2), 113-126. doi:10.1037/h0087222

Vrij, A., Edward, K., & Bull, R. (2001). Police officers' ability to detect deceit: The benefit of indirect deception detection measures. *Legal and Criminological Psychology, 6*(2), 185-196. doi:10.1348/135532501168271

Vrij, A., & Fisher, R. P. (2016). Which lie detection tools are ready for use in the criminal justice system? *Journal of Applied Research in Memory and Cognition, 5*(3), 302-307.

Vrij, A., Fisher, R. P., & Blank, H. (2017). A cognitive approach to lie detection: A meta-analysis. *Legal and Criminological Psychology, 22*(1), 1-21. doi:10.1111/lcrp.12088

Vrij, A., & Granhag, P. A. (2012). Eliciting cues to deception and truth: What matters are the questions asked. *Journal of Applied Research in Memory and Cognition, 1*(2), 110-117. doi:10.1016/j.jarmac.2012.02.004.

Vrij, A., Leal, S., & Fisher, R. P. (2018). Verbal deception and the Model Statement as a lie detection tool. *Frontiers in Psychiatry, 9*, 492. doi:10.3389%2Ffpsyt.2018.00492

Vrij, A., Leal, S., Fisher, R. P., Mann, S., Dalton, G., Jo, E., . . . Houston, K. (2018). Sketching as a Technique to Eliciting Information and Cues to Deceit in Interpreter-Based Interviews. *Journal of Applied Research in Memory and Cognition, 7*(2), 303-313. doi:10.1016/j.jarmac.2017.11.001

Vrij, A., Leal, S., Granhag, P. A., Mann, S., Fisher, R. P., Hillman, J., & Sperry, K. (2009). Outsmarting the liars: The benefit of asking unanticipated questions. *Law and human behavior, 33*(2), 159-166.

Vrij, A., Leal, S., Jupe, L., & Harvey, A. (2018). Within-subjects verbal lie detection measures: A comparison between total detail and proportion of complications. *Legal and Criminological Psychology, 23*(2), 265-279. doi:10.1111/lcrp.12126.

Vrij, A., Leal, S., Mann, S., Dalton, G., Jo, E., Shaboltas, A., . . . Houston, K. (2017). Using the model statement to elicit information and cues to deceit in interpreter-based interviews. *Acta Psychologica, 177*, 44-53. doi:10.1016/j.actpsy.2017.04.011

Vrij, A., Leal, S., Mann, S., Vernham, Z., & Brankaert, F. (2015). Translating theory into practice: Evaluating a cognitive lie detection training workshop. *Journal of Applied Research in Memory and Cognition, 4*(2), 110-120. doi:10.1016/j.jarmac.2015.02.002

Vrij, A., Mann, S., & Fisher, R. P. (2006). Information-gathering vs accusatory interview style: Individual differences in respondents' experiences. *Personality and Individual Differences, 41*(4), 589-599. doi:10.1016/j.paid.2006.02.014

Vrij, A., Meissner, C. A., Fisher, R. P., Kassin, S. M., Morgan III, C. A., & Kleinman, S. M. (2017). Psychological perspectives on interrogation. *Perspectives on Psychological Science, 12*(6), 927-955. doi:10.1177/1745691617706515

Vrij, A., & Winkel, F. W. (1991). Cultural patterns in Dutch and Surinam nonverbal behavior: An analysis of simulated police/citizen encounters. *Journal of Nonverbal behavior, 15*(3), 169-184. doi:10.1007/bf01672219

Wagenmakers, E.-J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., . . . Morey, R. D. (2018). Bayesian inference for psychology. Part II: Example applications with JASP. *Psychonomic Bulletin & Review, 25*(1), 58-76. doi:10.3758/s13423-017-1323-7

Weinberger, S. (2010). Airport security: Intent to deceive? *Nature News, 465*(7297), 412-415.

Zappalà, A., & Pompedda, F. (2017). Criminal interrogation in Italy: Legal procedures and practices *International developments and practices in investigative interviewing and interrogation* (pp. 180-188): Routledge.

Zuckerman, M., DePaulo, B. M., & Rosenthal, R. (1981). Verbal and nonverbal communication of deception *Advances in experimental social psychology* (Vol. 14, pp. 1-59): Elsevier.

Appendices

In these appendices the reader will find the papers corresponding to its specific paragraph.

Appendix 1: Palena, N., Caso, L., Granhag, P. A., Orthey, R., Monticciolo, R., & Vrij, A. Reciprocity: Does it help in Scharff interviews? [IP2]. Paragraph 4.2.

Appendix 2**.** Palena, N., Caso, L., Carlotto, G., De Mizio, L., & Marciali, M. (2017). Efficacia della tecnica di baselining nella valutazione della credibilità (Efficacy of the baseline technique when assessing credibility). *Giornale italiano di psicologia, 44*(4), 905-916. doi:10.1421/88773. [PA1]. Paragraph 4.3.

Appendix 3**:** Palena, N., Caso, L., Vrij, A., & Orthey, R. (2018). Detecting deception through small talk and comparable truth baselines. *Journal of Investigative Psychology and Offender Profiling, 15*(2), 124-132. doi:10.1002/jip.1495. [PA2]. Paragraph 4.4.

Appendix 4: Caso, L., Palena, N., Vrij, A., & Gnisci, A. (2019). Observers' performance at evaluating truthfulness when provided with Comparable Truth or Small Talk Baselines. *Psychiatry, Psychology and Law*. doi:10.1080/13218719.2018.1553471. [PA3]. Paragraph 4.5.

Appendix 5: Caso, L., Palena, N., Carlessi, E., & Vrij, A. (2019). Police accuracy in truth/lie detection when judging baseline interviews. *Psychiatry, Psychology, and Law*. doi: 10.1080/13218719.2019.1642258. [PA4]. Paragraph 4.6.

Appendix 6: Caso, L., Palena, N., Vrij, A., & Melocchi, L. (in preparation). Implementing the use of a checklist with comparable truth and small talk baselines: The effect on laypersons' lie detection accuracy. [IP3]. Paragraph 4.7.

Appendix 7: Palena, N., Caso, L., & Vrij, A. (2019). Detecting Lies via a Theme-Selection Strategy. *Frontiers in Psychology, 9*(2775). doi:10.3389/fpsyg.2018.02775. [PA6]. Paragraph 4.8.

Appendix 1: Palena, N., Caso, L., Granhag, P. A., Orthey, R., Monticciolo, R., & Vrij, A. Reciprocity: Does it help in Scharff interviews? [IP2]

Khalid Sheikh Mohammed, the alleged "brain" behind 9/11 terrorist attack, was waterboarded by the CIA at least 183 times and eventually provided information, but this then proved to be false (Vrij et al., 2017). To date, there are still some proponents of torture, whose advocated goals are at least three (Rejali, 2009): i) to intimidate; ii) to obtain a confession; iii) to gather intelligence. Perhaps, one of the ideas behind the use of torture is that physical and psychological coercion can overcome sources' resistance (e.g., counter-interrogation strategies).

Psychological research, however, showed that torture is- besides inhuman- ineffective and that information obtained through torture is unreliable (Peters, 1996; Rejali, 2009). Evidence based approaches are better, and they proved effective for several goals, such as reliable recollection of memories, deception detection and information gathering in intelligence settings (Swanner, Meissner, Atkinson, & Dianiska, 2016; Vrij et al., 2017). Most of them share the underlying principle whereby the source must be treated fairly, and that rapport with the source should be obtained (Abbe & Brandon, 2013).

Also, 'reading' interviewees' mind can be effective as it may bypass sources' counter-interrogation strategies: Perspective-taking plays a very important role in such settings (Granhag & Hartwig, 2008).

Among non-coercive techniques there is the Scharff technique, which is discussed in the next section.

The Scharff Technique

"*What did he get out of me? There is no doubt in my mind that he did extract something, but I haven't the slightest idea what"*.

This is a quote by Colonel Hubert Zemke, a prisoner of war (POW) interrogated by Hans Scharff, an interrogator of the German Luftwaffe during world war II (Toliver, 1997). Scharff was perhaps one of the most effective interrogators of all times, who managed to obtain actionable intelligence by considering the perspective of his interrogees (Granhag, Kleinman, & Oleszkiewicz, 2016; Toliver, 1997). Scharff

concluded that sources have three prevalent counter-interrogation strategies: a) I *will not tell much*; b) *I will try to figure out what they are looking for and I will not tell them what they want*; c) *it is useless to withhold or deny known information.* When Scharff encountered sources, he refrained from harsh interrogation techniques and avoided direct questions and torture.

His approach involved extensive preparation and a cunning interviewing strategy. When Scharff met sources, he set the interview as a conversation, whereby the source was given the opportunity to add information. In preparing the interview before its beginning, Scharff managed to give an impression of already knowing most of the information, whereas by downplaying any new information provided by the source he made it quite difficult for the source to understand what his information objectives were. In doing so, he overcame sources' counter-interrogation strategies and managed to obtain new information.

There is a recent strand of research aiming to conceptualize Scharff's approaches into a technique, and five general tactics have been conceived (Granhag et al., 2016). Friendly approach: The interviewer does not play on authority. Rather, s/he tries to put the source at ease, shows (and expects) respect, and uses adaptive behaviours (Alison, Alison, Noone, Elntib, & Christiansen, 2013; Alison, Giles, & McGuire, 2015; Birtchnell, 2002); Do not press for information: The interviewer does not ask direct questions. Rather, the source is given the opportunity to add information or to (dis)confirm claims made by the interviewer. This tactic also recognises the importance of sources' autonomy (Miller & Rollnick, 2012).; I already know it all strategy: Scharff started the interview by presenting the sources with the information he already held. In doing so, he gave the impression of already knowing a relevant amount of information, and this has at least two effects. First, if the source wants to appear as cooperative, s/he must provide some new information. Second, the source can conclude that the interviewer holds additional information that s/he (the interviewer) has not told. Therefore, the source can add something which is, in reality, new to the interviewer; Use of (dis)confirmation: Rather than asking direct questions, the source is presented with some claims that s/he can confirm or disconfirm. Here, there are again at least two possible positive effects. First, it is easier for the source to confirm something rather than tell a brand-new piece of information as this is

perceived as being less compliant with the interrogator compared to answering direct questions. Second, the interrogator may use (dis)confirmation for his/her benefit in that s/he can present information about which they have some intelligence, but no certainty. This: a) gives the impression that the interviewer already knows the information concerning a specific claim, increasing the illusion-of-knowing-it-all; b) gives the interviewee the opportunity to just confirm it, which results in the interrogator accumulating information which are, in fact, new to the interviewer.

The Scharff technique has also been compared to the Direct Approach (DA), where the interview is conducted in a business-like manner, with open, direct questions being asked (Department of the Army, 2006). When comparing the Scharff technique to the Direct Approach, results in favour of the former appear (Granhag et al., 2016; Oleszkiewicz, Granhag, & Cancino Montecinos, 2014; Oleszkiewicz, Granhag, & Kleinman, 2014). First, in Scharff settings the source believes that the interviewer held more information prior to the interview than in direct approach settings. Second, interviewees interrogated with the Scharff technique find it more difficult to understand what are the interviewer's (information) objectives. Third, the Scharff technique elicits more new information than the direct approach, also because of the two previous effects. Fourth, interviewees interviewed via the Scharff technique underestimate, whereas those interviewed with the direct approach overestimate, the amount information they revealed. Interestingly, the Scharff technique outperforms the direct approach also with less cooperative sources (Granhag, Oleszkiewicz, Strömwall, & Kleinman, 2015) and when considering the precision of the information revealed (Oleszkiewicz, Granhag, & Cancino Montecinos, 2014). A source, indeed, can deal with the interrogation cost-benefit dilemma (Neequaye & Luke, 2018) by providing information which are vague or by balancing differently the detail and veracity of information they provide for different topics discussed within the same interview (Palena, Caso, & Vrij, 2019). For example, instead of revealing the specific area of the city where a bomb will be planted, sources may just reveal which is the target city.

The sum of the aforementioned evidence supports the potential of the Scharff technique in HUMINT interrogations. Considering that in HUMINT settings the sources are prevalently only *semi*-cooperative, a possible 'next step' is to combine the

Scharff technique with social influence strategies. Such strategies, indeed, showed to have some potential in investigative interviewing settings, especially when dealing with cooperation and information disclosure (Goodman-Delahunty, Martschuk, & Dhami, 2014), which will be dealt with in the next section.

Rapport and Social Influence

Rapport.

A relevant theoretical interpretation of rapport comes from the model outlined by Tickle-Degnen and Rosenthal (1990). According to the authors' view, rapport consists of mutual attention (being in sync), positivity (positive affects) and coordination. Positivity can also be read in light of social judgment (Cuddy, Fiske, & Glick, 2008; Fiske, Cuddy, & Glick, 2007) such as warmth and competence (Abbe & Brandon, 2013). According to Abbe and Brandon (2013), rapport can be beneficial in interviewing and intelligence settings to create working alliance, elicit information, and as a form of social influence. Rapport also showed to be beneficial to obtain cooperation from the sources (Bull & Soukara, 2010; Department of the Army, 2006) which is in turn important to elicit information from them. Kelly, Miller, and Redlich (2016) found that rapport had a positive impact on sources' cooperation, whereas confrontation has a detrimental effect which can last up to 15 minutes. Goodman-Delahunty et al. (2014) interviewed 123 experienced interrogators on the theme of rapport-building and found that liking and reciprocity (which will be dealt with in the next section) were the most used principles used to create rapport with a source. Hence, rapport and cooperation are crucial prerequisites of a successful interrogation.

Social influence.

A way to obtain rapport is to employ social influence tactics. Cialdini (2001) identified six form of social influence: reciprocity, scarcity, commitment and consistency, social proof, authority, and liking. Cialdini and Goldstein (2004) define reciprocity as *"[…] the rule that obliges us to repay others for what we have received from them"* (p. 599). The principle of reciprocation is valid across cultures (Gouldner, 1960) and plays an important role in our relations with others (Kelln & Ellard, 1999). Rind and Strohmetz (1999) found that reciprocity increased tips received by restaurant

employees. Social influence tactics, in general, showed their potential in various situations (Cialdini, 2001), but also the way they are framed before they are enacted plays an important role: before changing minds, it is important to change states of mind (Cialdini, 2016).

Given the potential of social influence tactics, researchers started to explore their application within interviewing settings. Vrij et al. (2018) offered water to interviewees to increase rapport. More in detail, Matsumoto and Hwang (2018) directly tested how the application of the reciprocity principle influenced information disclosure with three different ethnic/cultural groups, where some people had to tell the truth, whereas others had to lie, about a mock crime. Their manipulation consisted of offering (or not) a bottle of water to interviewees before the investigative interview started. The author differentiated between relevant and irrelevant details and measured their plausibility. Additionally, the authors measured the level of rapport as they recognised its potential mediating effect between the experimental condition and information disclosure. The authors found that reciprocity influenced information disclosure, and that it interacted with whether the interviewees were telling the truth or lying. Truth tellers who had been offered water, indeed, revealed less relevant details than those who were not, whereas liars showed the opposite pattern. For liars, the effect was also mediated by rapport. The authors therefore concluded that the use of social influences tactics, such as reciprocity, has practical relevance.

The Present Study

Considering research showing the beneficial effect of the Scharff technique in eliciting information from semi-cooperative sources and that showing the potential positive effect of reciprocity in interviewing settings, the main goal of this study was to combine the two approaches. Indeed, if reciprocity makes liars disclose more relevant information and increases rapport, and rapport can increase interviewees' cooperation, then it is possible that employing the reciprocity principle in Scharff settings can make the semi-cooperative sources become more cooperative and reveal new information.

Reciprocity, however, can be established in several ways. First, the interviewer can offer a drink to the interviewee, who can then decide whether to drink or not, as in

Matsumoto and Hwang (2018). Offering water: a) causes the reciprocity principle to take place; b) gives autonomy to the interviewees, since it is their decision whether to drink the water or not. Autonomy is also important for rapport building (Alison et al., 2013) and it is one of the principles of the motivational interview (Miller & Rollnick, 2012). We will refer to this type of reciprocity as "autonomy" from here therein. Second, the interviewer can share the water with the interviewee (both the interviewer and the interviewees drink the water poured by the interviewer). Sharing is still a form of reciprocity, but it has also additional effects. Boothby, Clark, and Bargh (2014) found that shared experiences are amplified. In particular, the authors had the participants either tasting chocolate at the same time of another person (a confederate) or not. Participants in the first condition evaluated the chocolate as more likeable than participants in the second condition. Sharing, additionally, may also increase rapport and make the interviewer more likeable. Furthermore, research shows that sharing attention increases social facilitation and the cognitive resources allocated to the task at play (Shteynberg, 2015). We will refer to this type of reciprocity as "sharing" from here therein. Hence, the potential positive effect of reciprocity in investigative interviewing could be increased by sharing something with source rather than by offering something to him/her. This effect on information disclosure can be both direct and mediated by interviewer's likeability.

Considering the research outlined above we predicted the following effects. First, applying the reciprocity principle in HUMINT interviews increases the amount of new information revealed by the interviewees. Also, the feeling that *obliges us to repay others for what we have received from them"* (Cialdini & Goldstein, 2004, p. 599) encouraged by reciprocity may refrain interviewees from lying. Therefore, when looking at information disclosure, we hypothesise that interviewees in the "sharing" condition would reveal a higher proportion of true new information than interviewees in the "autonomy" condition and that, in turn, interviewees in the autonomy condition will reveal a higher proportion of true new pieces of information than interviewees who were not offered water (Hypothesis 1). Second, we predicted that the effect of reciprocity on information disclosure is mediated by interviewer's likeability and the stress experienced by the interviewee during the interview. In particular, we expected that: a) interviewees in the "sharing" condition would like the interviewer more than

interviewees in the "autonomy" condition who, in turn, like the interviewer more than interviewees who had not been offered water (from here therein, "no reciprocity"); b) liking the interviewer more increases information disclosure; c) increased interviewer's likeability decreases the stress experienced by the interviewees; d) lower stress increases information disclosure (Hypothesis 2). Figure 1 depicts the hypothesized path diagram.

Method

Participants

Sixty university students (55 females and 5 males) took part in the experiment. Age ranged from 20 years old to 46 years old (M = 23.42, SD = 4.55). Participants were equally distributed across the three experimental conditions. All procedures performed in this study were in accordance with the ethical standards of the institutional and/or national research committee and with APA regulations.

Design

The experiment featured 3 (Reciprocity: No Reciprocity vs. Autonomy vs. Sharing, between-subjects) between-subjects design. The amount of new detail revealed (including confirmed pieces of information, see "coding" section) was the dependent variable. Two additional variables were measured and entered as mediators into the model and were computed as follows. The degree of interviewer's likeability was measured on a 7-points likert scale which could range from 1 (not at all) to 7 (very much). The same was true for and the degree of stress experienced by the interviewees.

Procedure

A list of participants was obtained during university lectures. Participants were informed that they would take part in an experiment mirroring a HUMINT investigative interview. Upon arrival they were fully briefed by the experimenter, who told them that the experiment would last about 60-75 minutes and gave them the following written information. Participants would have to play the role of a component of a criminal organization who did not commit any criminal nor violent act themselves, but whose role was to be the organization accountant. The police

started to dismantle the organization last spring and arrested some of its components, including the participant. Participants were asked to imagine that they were given the opportunity to receive a penalty discount if they helped the police to reach their aim to dismantle the organization and to arrest its boss. However, participants were informed that they had to strike a balance between not providing too much nor too little information, a common manipulation in this field (Granhag et al., 2016; Oleszkiewicz, Granhag, & Cancino Montecinos, 2014). To appear cooperative, they had to reveal some information, but since they were told to have a negative attitude toward the police and did not want their fellow criminal to be arrested, they should not tell everything they knew. Participants were also informed that the police already held some information, also because rumours said that the police had intercepted the communications between the criminals. However, participants did not know exactly what information the interviewer already knew. Further, participants were informed that according to their performance in balancing the amount of information revealed they would obtain from one to three additional points for a university exam (eventually, all participants received 2 points).

As soon as the participants confirmed that they understood their task, they were given 15 minutes to study the material describing the structure and activities of the criminal organization, as well as information concerning its components, similar to Palena et al. (2019). The material also included three pictures: a) a photo of one of the components of the organisation; b) a record of a bank transaction; c) a photo depicting the boss' car. Once the participants had finished studying the material, the experimenter evaluated participants' memory for each piece of information in a repetitive manner. That is, the experimenter asked a sequence of questions, one for each piece of information, and noted on a checklist any piece of information that the participant may have forgotten or misremembered. Once the experimenter ended reading the last question, she asked again the questions concerning any piece of information which the participant did not remember or misremembered. The process was repeated until the participant remembered all pieces of information correctly. After the memory evaluation was over the interview, which is reported below, started.

Interview

A female psychology student acted as the interviewer. She was trained for the purpose of the experiment and performed simulated interviews before interviewing actual participants. This was done to make sure that she followed the structure of the interview, kept a constant demeanour, and did not improvise during the interview (Oleszkiewicz, Granhag, & Cancino Montecinos, 2014; Palena et al., 2019). The interviewer was blind to the hypotheses of the experiment and to the experimental conditions.

The interview started with the interviewer introducing herself and inviting the interviewee to take their seat. Then, the interviewer stressed the importance of the meeting and empathized with the interviewee's uncomfortable situation. The experimental manipulation then took place. For the "no reciprocity" condition, the interview started immediately. For the "autonomy" condition, the interviewer invited the interviewee to take a glass of water before starting the interview, by saying *"I would like you to be at ease, please have a glass of water before we start"*. The interviewer did not drink in this condition. For the "sharing" condition, the interviewer shared the water with the interviewee by saying *"I would like you to be at ease, let's have a glass of water before we start".* From here therein, the interview did not differ between the three experimental condition.

The interviewer started by presenting known information to the interviewee in a conversational manner. Known information concerned the structure of the criminal organisation, its components and some of its activities. In total, the interviewer presented 16 known pieces of information. Three of them where presented via photographic evidence (the same picture included in the material the participants had to study). The interviewer then sympathised again with the interviewee and said that they intended to be faithful to the deal they had agreed on (the participant receiving a penalty discount/between 1 and 3 points for a university examination). The interviewer then asked the following open-ended question: *"How could you help us with additional information?"*. Once the participants had finished answering, the interviewer presented four confirmation claims which the participant could (dis)confirm. Then, the interviewer asked the following open-ended question *"Is there anything you would like to add?"* and the interview then ended.

Post-interview Questionnaire

Once the interview was over, the experimenter asked the participants to fill in a post-interview questionnaire. The questions concerned how much motivated to succeed in the experiment the participant felt; how likeable the interviewer was; how pleasant the interviewing room was; how stressful the interview was; all on a likert scale from 1 (not at all) to 7 (very much).

Further, the interviewee was asked how much information they believed the interviewer already knew prior to the interview, on a likert scale from 1 (none) to 7 (all); also, the interviewee was asked to tick each of the 29 pieces of information that they believed the interviewer knew prior to the interview. Finally, they were asked to rate their perceived importance/relevance of each piece of information, were 1 = little important, 2 = moderately important, and 3 = very important.

Coding

In first instance, all interviews were transcribed. Then, two coders, both blind to experimental hypotheses and conditions, coded 30 transcripts (50%) for the presence of each piece of information. For this purpose, a grid including the 29 pieces of information provided to the interviewees was created. Sixteen pieces of information were known to the interviewer prior to the interview, whereas 13 were unknown to the interviewer. Of these new pieces of information, four were presented as confirmation claims. The coders also coded for the veracity of each piece of information, which was considered as true, false (in case the interviewee falsified a piece of information included in the story), inconsistent (in case the interviewee reported the same piece of information once truthfully and once lying), or "mixture" (in case the interviewee revealed a piece of information which was partially true and partially false). An inter-rater reliability analysis on new and confirmation claims information only was run, and showed high agreement between the two coders, $ICC_{(2,2)} = .91$. At this point, any disagreement between the two coders was discussed and resolved. The first coder then coded the remaining 30 transcripts.

Results

Manipulation Checks

Participants' motivation was above the mid-point (M = 6.33, SD = .60). Further, for the Scharff technique to be effective, it is important that the interviewer successfully gives the impression of already knowing it all (Granhag et al., 2016). Therefore, we explored perceived previous interviewer's knowledge as rated by interviewees, which was above the mid-point as well (M = 5.75, SD = 1.00). Also, it has been suggested that room features, such as "cosiness" (Heyer & Traufetter, 2011) may influence disclosure (Dawson, Hartwig, Brimbal, & Denisenkov, 2017; Miwa & Hanyu, 2006). Our participants' evaluation of setting pleasantness was above the mid-point as well (M = 5.63, SD = .99). Motivation, Welch's $F(2, 36.58) = 2.38$, $p = .11$, perceived interviewer's knowledge, Welch's $F(2, 36.22) = .54$, $p = .59$, and setting pleasantness, $F(2, 57) = .72$, $p = .49$, did not differ significantly between reciprocity conditions.

Lastly, since our hypothesis focused on the proportion of true information, it was important to explore the occurrences of untrue pieces of information. A first descriptive analysis showed that interviewees revealed a very low amount of untruthful (false, inconsistent and mixed collapsed together) information (M = .33, SD = .68, 95% CI [.15, .51]). Forty-four participant only reported truthful information, 14 reported at least one untruthful piece of information, and two participants reported two and four untruthful pieces of information respectively. Due to the low frequency of untruthful information, it was inadvisable to run any analysis considering the effect of reciprocity of lying. Therefore, hypothesis testing focused on the total (raw) number of new pieces of information revealed without accounting for veracity.

Information Disclosure Mediation Model

A mediation analysis was conducted (Fig. 1) with reciprocity (no reciprocity vs. autonomy vs. sharing) as the factor. Interviewer's likeability and stress were the mediators. Additionally, interviewer's likeability was also set as a predictor of stress. The amount of new information revealed by the interviewees was the dependent variable. Reciprocity had no effect, neither direct nor mediated, on information disclosure (Table 1). Therefore, neither of the two Hypotheses was supported. Interviewees in the "no reciprocity" condition, (M = 5.45, SD = 1.76, 95% CI [4.62, 6.27]), those in the "autonomy" condition, (M = 5.40, SD = 2.09, 95% CI [4.42, 6.37]), and those in the "sharing" condition, (M = 6.30, SD = 2.58, 95% CI [5.09, 7.50]), all

disclosed a similar amount of new detail (Figure 2). The only significant effect appeared for the difference in perceived stress between participants in the "autonomy" (M = 3.75, SD = 1.51, 95% CI [3.04, 4.46]) and those in the "sharing" (M = 4.75, SD = 1.44, 95% CI [4.07, 5.43]) conditions, b = 1.048, p = .03 (Table 1). Further, those participants who revealed more (sharing condition), only reported on average about 48% of new pieces of information (Fig. 2).

Further Explorations

Known vs. unknown pieces of information.

Mediation model.

The idea behind the use of reciprocity in the present study was that it makes interviewees more cooperative and more compliant to the interviewer. Yet, analysing the pieces of information which were objectively new (unknown) to the interviewer may not give the full picture. The rationale behind this is that if reciprocity influences interviewees' cooperation, then it makes sense to consider that truly cooperative participants may reveal information that they themselves believe to be unknown to the interviewer. For this reason, we re-run the same analyses as those for hypothesis testing, but with the amount of information perceived to be unknown to the interviewer as the dependent variable[26]. Again, no significant result appeared (Table 2). Participants in the no reciprocity condition, those in the autonomy condition, and those in the sharing condition, all reported a similar proportion of information they perceived as unknown to the interviewer (Figure 3).

Balancing between known and unknown pieces of information.

A repeated measure ANOVA with Reciprocity as the between-subjects factor (no reciprocity vs. autonomy vs. sharing), perceived interviewer's knowledge (information perceived as known vs. unknown) as the within-subjects factor, and the proportion of revealed information as the dependent variable was conducted. The reciprocity condition main effect, $F(2, 56) = .53$, $p = .58$, $\eta_p^2 = .02$, and the reciprocity

[26] A proportion was used here as participants may differ in regard of what information they thought were new, and what were known, to the interviewer.

by perceived interviewer's knowledge interaction, $F(2, 56) = 1.02$, $p = .36$, $\eta_p^2 = .03$ were not significant. The main effect of perceived interviewer's knowledge was significant, $F(1, 56) = 62.47$, $p < .001$, $\eta_p^2 = .53$. Participants reported a higher proportion of information they perceived to be known to the interviewer ($M = .51$, $SD = .17$, 95% CI [.47, .55]) than the proportion of information they perceived as unknown ($M = .30$, $SD = .20$, 95% CI [.25, .36]) to the interviewer.

Participants' accuracy of prior interviewer's knowledge.

One of the principles of the Scharff technique is the illusion-of-knowing-it-all strategy. We therefore analysed interviewee's accuracy for old information, confirmation claims and new information individually. Participants were quite accurate in evaluating old information as known to the interviewer ($M = .73$, $SD = .17$, 95% CI [.68, .77]) and new information as unknown to the interviewer ($M = .88$, $SD = .16$, 95% CI [.84, .92]). However, interviewees were mostly inaccurate at evaluating information presented in confirmation claims as actually new to the interviewer ($M = .15$, $SD = .23$, 95% CI [.09, .21]).

Importance of information: Cognitive dissonance and comparison of interviewees vs. police officers' coding.

Obtaining the indication of the importance from both interviewees and police officers gives us the possibility to: i) test whether interviewees underestimated the importance of the information they thought they had revealed during the interview, which can be interpreted in terms of cognitive dissonance (Festinger, 1957); ii) explore whether interviewees underestimated or overestimated the importance of information compared to the ratings provided by police officers.

To test whether interviewees under or overestimated the importance of the pieces of information s/he had revealed, a linear mixed model analysis was conducted. Disclosure (whether the interviewee revealed or not each piece of information) was the fixed factor. The value allocated to each piece of information was the dependent variable. Intercepts were the random component. The test revealed a nonsignificant effect for the factor Disclosure (revealed vs. not revealed), $F(1, 1720.76) = 3.77$, $p = .052$. Interviewees rated the information they had revealed ($M = 2.47$, $SD = .70$) of a

similar importance of that they had not revealed (*M* = 2.40, *SD* = .74). As the descriptive show, information was rated as more than moderately important, on average.

To explore whether the interviewees underestimated or overestimated the pieces of information compared to police officers' evaluations, we firstly obtained coding from two experienced officers, who evaluated the importance of each piece of information the same way as the interviewees did. Then, a mean value for each piece of information for interviewees' ratings, and another mean value for police officers' ratings[27] were obtained. Subsequently, we subtracted the mean value (for each piece of information) provided by police officers to the mean value provided by interviewees (interviewees' mean – police officers' mean). Therefore, a negative score indicate that the interviewees *underestimated* the importance of a specific piece of information as rated by police officers, whereas a positive score would indicate that the interviewees *overestimated* the relative importance. Then, we assigned a value of zero if the two means corresponded, a value of one if the interviewees underestimated the relative importance of a specific piece of information, and a value of two if the interviewees overestimated the relative importance. Interviewees underestimated 65.5% of pieces of information and overestimated the remaining 34.5%, but this did not reach statistical significance, χ^2 (1) = 2.79, p = .09.

Discussion

In this experiment, we tested the effect of reciprocity on information disclosure in Scharff interviewing. We found that reciprocity (Cialdini, 2001, 2016; Cialdini & Goldstein, 2004), had no effect on information disclosure, neither on objectively nor on perceived (by the interviewee) new pieces of information. Both the direct and the mediated effect were not significant. Our hypotheses were therefore not supported, and our results did not fit with previous work on the effect of reciprocity on information disclosure in investigative interviews (Matsumoto & Hwang, 2018).

[27] We evaluated the agreement between the two police officers using weighted Cohen's kappa, which was of 0.62.

One possible reason for the lack of significance may be due to the Scharff setting itself. Indeed, in HUMINT interviewing, interviewees usually engage in a cost-benefit analysis, weighting what to reveal to obtain a compensation, and what to withhold to defend one's own group or goals (Neequaye & Luke, 2018). Further, participants in our study as well as participants in previous studies on the Scharff technique (Granhag et al., 2015; Oleszkiewicz, Granhag, & Cancino Montecinos, 2014; Oleszkiewicz, Granhag, & Kleinman, 2014) were requested to strike a balance between providing too little and too much information, and they were offered something as a reward according to their performance (e.g.: money, additional points, leniency in the role-playing scenario, etc.). Therefore, looking at this picture closely, the Scharff itself embeds a form of reciprocity: Rewarding the interviewee. This may have caused a ceiling effect whereby applying the reciprocity principle through additional offers (such as water, in our case) no longer works because the effect of reciprocity has already been exhausted by the possibility to receive a reward. Hence, additive effects of different forms of reciprocity may not increase cooperation.

Another possible explanation for the lack of an effect lies in the type of reciprocity we employed. Offering water can be seen as a form of (positive) affective reciprocity (Forgas, 2011; Pike & Sillars, 1985), whereas the bargain at play in HUMINT/Scharff interviewing can sometime be seen as more of an instrumental form of reciprocity, where there is a forward-looking decision making process guiding the decision to reveal information or not. This distinction has already been given elsewhere. Sobel (2005), proposes a difference between "intrinsic reciprocity", where a person responds back to a kind act with a kind behaviour himself/herself (but it can also work in a negative way, where un unkind act is punished with another unkind act as a response) and "instrumental reciprocity". In the former case, the intention of the other person is usually at the core of the responding behaviour. In the latter case, the long-term gain is usually considered. The forward-looking behaviour may have been more relevant here. The strategy behind this latter case (instrumental reciprocity) is based on selfishness, whereby the actions performed by an agent are informed by the expected response of our opponent (Cabral, Ozbay, & Schotter, 2014; Sobel, 2005). Additionally, our experiment (as all but one study on the Scharff technique, see

Oleszkiewicz, Granhag, & Kleinman, 2017) was based on a single interview. However, in HUMINT settings it is more common that sources are interviewed multiple times. Therefore, long-term effects of principles such as the forward-looking behaviour which may be at play with instrumental reciprocity, become even more relevant. Future studies should focus on the two different forms of reciprocity (socio-affective vs. instrumental) and how they affect the outcome in single and repeated interviews. It is plausible that intrinsic reciprocity increases information elicitation through rapport because of the perceived warmth of the interviewer (Abbe & Brandon, 2013; Cuddy et al., 2008; Fiske et al., 2007); instrumental reciprocity, on the contrary, can increase information elicitation via a direct effect based on a cognitive (selfishness and cost-benefit analysis) rather than affective component. This question, however, remains open for future studies. A last explanation for our null findings is that reciprocity just does not work in settings like this.

Additional analyses showed that our participants revealed about 50% of the information they believed to be known to the interviewer, and about 30% of information they thought to be unknown to the interviewer. This makes sense and fit with the extant literature (Granhag & Luke, 2018). Indeed, in doing so, the interviewee can appear as cooperative when, in reality, s/he is mainly providing information thought to be known to the interviewer. Considering that our interviewees reached high accuracy in figuring out what pieces of information were old and new, but that they reached low accuracy in classifying confirmations as information new to the interviewer, future studies should focus on how to improve the "illusion of knowing it all" strategy and in refining the "confirmation/disconfirmation" tactic (Granhag et al., 2016), so that the interviewee believes that the interviewer knows almost everything.

Furthermore, we found no support for a cognitive dissonance effect whereby interviewees would underestimate the importance of information they revealed. One possible reason for this may be due to the content of each specific piece of information. For example, information such as where the head of the criminal organisation hides may be easily evaluated as more important than, for example, the

fact that the interviewee worked as an accountant within the organisation. This difference may be so clear that it overrode any possible cognitive dissonance effect. Also, it is possible that the interviewees had decided their strategy (what information to disclose and what information to withhold) before the interview and they then stuck to it. However, this needs further exploration.

Lastly, we did not find a significant difference when comparing interviewees' and police officers' ratings of the importance of information, which suggests that interviewees where skilful in deciphering what information is perceived as important and what as not important by police officers. Future studies may therefore try to explore this further so that most important information (e.g.: information objectives needed to prevent a terrorist attack) are elicited.

Limitations

One limitation of the present experiment is that we only measured how much the interviewees liked the interviewer, and used this measure as a mediator, but did not measure rapport (Alison et al., 2014). Rapport may have worked better than likeability as a mediator, especially considering that our participants consistently rated the interviewee as likeable (hence, we cannot exclude a ceiling effect). Future studies can therefore explore the effect of reciprocity when rapport, rather than likeability, is entered as a mediator.

We also did not actively manipulate interviewer's likeability. Future studies can therefore explore the effect of reciprocity in Scharff interviews when also manipulating interviewer's likeability so that both main effects and their interaction may be explored. Furthermore, we only applied the reciprocity principle at the beginning of the interview: Future studies should explore how reciprocity affects information disclosure when it is employed later in the interview, and when it is employed in several occasions such as in repeated interviews.

We also had only one interviewer. Although a structured interview was followed, it is important that future studies explore the possible effect of interviewer's

characteristics in HUMINT interviewing, even when the interview follows a structured protocol.

Lastly, we did not explore whether the interviewees perceived reciprocity as a voluntary form of social influence enacted by the interviewer, which may have negatively affected information disclosure, which future studies should also explore.

References

Abbe, A., & Brandon, S. E. (2013). The role of rapport in investigative interviewing: A review. *Journal of Investigative Psychology and Offender Profiling, 10*(3), 237-249. doi:10.1002/jip.1386

Alison, L., Alison, E., Noone, G., Elntib, S., & Christiansen, P. (2013). Why tough tactics fail and rapport gets results: Observing Rapport-Based Interpersonal Techniques (ORBIT) to generate useful information from terrorists. *Psychology, Public Policy, and Law, 19*(4), 411. doi:10.1037/a0034564

Alison, L., Alison, E., Noone, G., Elntib, S., Waring, S., & Christiansen, P. (2014). The efficacy of rapport-based techniques for minimizing counter-interrogation tactics amongst a field sample of terrorists. *Psychology, Public Policy, and Law*, *20*(4), 421-430. doi:10.1037/law0000021

Alison, L., Giles, S., & McGuire, G. (2015). Blood from a stone: Why rapport works and torture doesn't in'enhanced'interrogations. *Investigative Interviewing: Research and Practice, Special Issue*, 5-23.

Birtchnell, J. (2002). Psychotherapy and the interpersonal octagon. *Psychology and Psychotherapy: Theory, Research and Practice, 75*(3), 349-363. doi:10.1348/147608302320365235

Boothby, E. J., Clark, M. S., & Bargh, J. A. (2014). Shared Experiences Are Amplified. *Psychological Science, 25*(12), 2209-2216. doi:10.1177/0956797614551162

Bull, R., & Soukara, S. (2010). Four studies of what really happens in police interviews. In G. D. Lassiter & C. A. Meissner (Eds.), *Police interrogations and false*

confessions: Current research, practice, and policy recommendations (pp. 81-95). Washington, DC: American Psychological Association.

Cabral, L., Ozbay, E. Y., & Schotter, A. (2014). Intrinsic and instrumental reciprocity: An experimental study. *Games and Economic Behavior, 87*, 100-121. doi:https://doi.org/10.1016/j.geb.2014.05.001

Cialdini, R. (2001). *Influence: Science and practice (4th ed.)* (Vol. 3). Boston, MA: Allyn and Bacon.

Cialdini, R. (2016). *Pre-Suasion: A revolutionary way to influence and persuade*. New York, NY: Simon and Schuster.

Cialdini, R., & Goldstein, N. J. (2004). Social Influence: Compliance and Conformity. *Annual Review of Psychology, 55*(1), 591-621. doi:10.1146/annurev.psych.55.090902.142015

Cuddy, A. J., Fiske, S. T., & Glick, P. (2008). Warmth and competence as universal dimensions of social perception: The stereotype content model and the BIAS map. *Advances in experimental social psychology, 40*, 61-149. doi:10.1016/S0065-2601(07)00002-0

Dawson, E., Hartwig, M., Brimbal, L., & Denisenkov, P. (2017). A room with a view: Setting influences information disclosure in investigative interviews. *Law and human behavior, 41*(4), 333. doi:10.1037/lhb0000244

Department of the Army. (2006). *Human intelligence collector operations (Field Manual 2-22.3)*. (1494811952). Wasinghton, DC: Headquarters, Department of the Army.

Festinger, L. (1957). *A theory of cognitive dissonance* (Vol. 2): Stanford university press.

Fiske, S. T., Cuddy, A. J., & Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends in cognitive sciences, 11*(2), 77-83. doi:10.1016/j.tics.2006.11.005

Forgas, J. P. (2011). Affective influences on self-disclosure: Mood effects on the intimacy and reciprocity of disclosing personal information. *Journal of personality and social psychology, 100*(3), 449. doi:10.1037/a0021129

Goodman-Delahunty, J., Martschuk, N., & Dhami, M. K. (2014). Interviewing High Value Detainees: Securing Cooperation and Disclosures. *Applied Cognitive Psychology, 28*(6), 883-897. doi:10.1002/acp.3087

Gouldner, A. W. (1960). The Norm of Reciprocity: A Preliminary Statement. *American Sociological Review, 25*(2), 161-178. doi:10.2307/2092623

Granhag, P. A., & Hartwig, M. (2008). A new theoretical perspective on deception detection: On the psychology of instrumental mind-reading. *Psychology, Crime & Law, 14*(3), 189-200. doi:10.1080/10683160701645181

Granhag, P. A., Kleinman, S. M., & Oleszkiewicz, S. (2016). The Scharff technique: On how to effectively elicit intelligence from human sources. *International Journal of Intelligence and CounterIntelligence, 29*(1), 132-150. doi:10.1080/08850607.2015.1083341

Granhag, P. A., & Luke, T. J. (2018). How to interview to elicit concealed information: Introducing the Shift-of-Strategy (SoS) approach. In J. P. Rosenfeld (Ed.), *Detecting Concealed Information and Deception: Recent Developments* (pp. 272-295). San Diego, USA: Academic Press.

Granhag, P. A., Oleszkiewicz, S., Strömwall, L. A., & Kleinman, S. M. (2015). Eliciting intelligence with the Scharff technique: Interviewing more and less cooperative and capable sources. *Psychology, Public Policy, and Law, 21*(1), 100.

Heyer, J. A., & Traufetter, G. (2011). The Breivik interrogations: Norway mas-sacre suspect reveals all but motive. *Spiegel Magazine.* Retrieved from http://www.spiegel.de/international/spiegel/the-breivik-interrogations-norway-mas-sacre-suspect-reveals-all-but-motive-a-793923.htm

Kelln, B. R. C., & Ellard, J. H. (1999). An Equity Theory Analysis of the Impact of Forgiveness and Retribution on Transgressor Compliance. *Personality and Social Psychology Bulletin, 25*(7), 864-872. doi:10.1177/0146167299025007008

Kelly, C. E., Miller, J. C., & Redlich, A. D. (2016). The dynamic nature of interrogation. *Law and human behavior, 40*(3), 295-309. doi:10.1037/lhb0000172

Matsumoto, D., & Hwang, H. C. (2018). Social influence in investigative interviews: The effects of reciprocity. *Applied Cognitive Psychology, 32*(2), 163-170. doi:10.1002/acp.3390

Miller, W. R., & Rollnick, S. (2012). *Motivational interviewing: Helping people change*. New York: Guilford press.

Miwa, Y., & Hanyu, K. (2006). The Effects of Interior Design on Communication and Impressions of a Counselor in a Counselling Room. *Environment and Behavior, 38*(4), 484-502. doi:10.1177/0013916505280084

Neequaye, D. A., & Luke, T. J. (2018). Management of disclosure-costs in intelligence interviews. *PsyArXiv. November, 28*. doi:10.31234/osf.io/tfp2c

Oleszkiewicz, S., Granhag, P. A., & Cancino Montecinos, S. (2014). The Scharff-technique: Eliciting intelligence from human sources. *Law and human behavior, 38*(5), 478. doi:10.1037/lhb0000085

Oleszkiewicz, S., Granhag, P. A., & Kleinman, S. M. (2014). On eliciting intelligence from human sources: Contextualizing the Scharff-technique. *Applied Cognitive Psychology, 28*(6), 898-907. doi:10.1002/acp.3073

Oleszkiewicz, S., Granhag, P. A., & Kleinman, S. M. (2017). Gathering human intelligence via repeated interviewing: further empirical tests of the Scharff technique. *Psychology, Crime & Law, 23*(7), 666-681. doi:10.1080/1068316X.2017.1296150

Palena, N., Caso, L., & Vrij, A. (2019). Detecting Lies via a Theme-Selection Strategy. *Frontiers in Psychology, 9*(2775). doi:10.3389/fpsyg.2018.02775

Peters, E. (1996). *Torture*. Philadelphia: University of Pennsylvania Press.

Pike, G. R., & Sillars, A. L. (1985). Reciprocity of Marital Communication. *Journal of Social and Personal Relationships, 2*(3), 303-324. doi:10.1177/0265407585023005

Rejali, D. (2009). *Torture and democracy*: Princeton University Press.

Rind, B., & Strohmetz, D. (1999). Effect on Restaurant Tipping of a Helpful Message Written on the Back of Customers' Checks. *Journal of Applied Social Psychology, 29*(1), 139-144. doi:doi:10.1111/j.1559-1816.1999.tb01378.x

Shteynberg, G. (2015). Shared attention. *Perspectives on Psychological Science, 10*(5), 579-590. doi:10.1177/1745691615589104

Sobel, J. (2005). Interdependent preferences and reciprocity. *Journal of economic literature, 43*(2), 392-436. doi:10.1257/0022051054661530

Swanner, J. K., Meissner, C. A., Atkinson, D. J., & Dianiska, R. E. (2016). Developing diagnostic, evidence-based approaches to interrogation. *Journal of Applied Research in Memory and Cognition, 5*(3), 295-301.

Tickle-Degnen, L., & Rosenthal, R. (1990). The nature of rapport and its nonverbal correlates. *Psychological inquiry, 1*(4), 285-293. doi:10.1207/s15327965pli0104_1

Toliver, R. (1997). The interrogator. *Pennsylvania: Schiffer Publishing*.

Vrij, A., Leal, S., Fisher, R. P., Mann, S., Dalton, G., Jo, E., . . . Houston, K. (2018). Sketching as a Technique to Eliciting Information and Cues to Deceit in Interpreter-Based Interviews. *Journal of Applied Research in Memory and Cognition, 7*(2), 303-313. doi:10.1016/j.jarmac.2017.11.001

Vrij, A., Meissner, C. A., Fisher, R. P., Kassin, S. M., Morgan III, C. A., & Kleinman, S. M. (2017). Psychological perspectives on interrogation. *Perspectives on Psychological Science, 12*(6), 927-955. doi:10.1177/1745691617706515

Figure 1. *Mediation path diagram.*

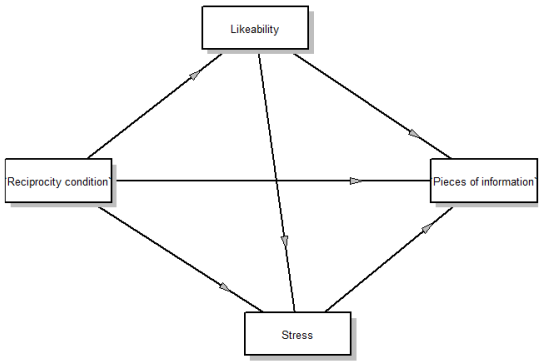


Table 1. *Indirect, direct, and total effects on the amount of information revealed (confirmation and new information together).*

Type	Effect	Estimate	SE	95% C.I. (a)		β	z	p
				Lower	Upper			
Indirect	Reciprocity condition1 ⇒ Likeability ⇒ Pieces of information	-0.006	0.046	-0.097	0.085	-0.001	-0.132	0.8951
	Reciprocity condition1 ⇒ Stress ⇒ Pieces of information	0.006	0.086	-0.162	0.173	0.001	0.068	0.9458
	Reciprocity condition2 ⇒ Likeability ⇒ Pieces of information	-0.018	0.057	-0.131	0.094	-0.004	-0.319	0.7494
	Reciprocity condition2 ⇒ Stress ⇒ Pieces of information	-0.179	0.199	-0.569	0.211	-0.039	-0.901	0.3677
	Reciprocity condition1 ⇒ Likeability ⇒ Stress ⇒ Pieces of information	0.003	0.020	-0.037	0.042	5.984e-4	0.135	0.8927
	Reciprocity condition2 ⇒ Likeability ⇒ Stress ⇒ Pieces of information	0.008	0.022	-0.035	0.052	0.002	0.371	0.7107

Table 1. *Indirect, direct, and total effects on the amount of information revealed (confirmation and new information together).*

Type	Effect	Estimate	SE	95% C.I. (a)		β	z	p
				Lower	Upper			
Component	Reciprocity condition1 ⇒ Likeability	0.050	0.366	-0.668	0.768	0.020	0.137	0.8914
	Likeability ⇒ Pieces of information	-0.122	0.240	-0.592	0.348	-0.066	-0.510	0.6098
	Reciprocity condition1 ⇒ Stress	-0.034	0.499	-1.012	0.944	-0.009	-0.068	0.9457
	Stress ⇒ Pieces of information	-0.171	0.171	-0.507	0.165	-0.134	-0.998	0.3185
	Reciprocity condition2 ⇒ Likeability	0.150	0.366	-0.568	0.868	0.061	0.410	0.6822
	Reciprocity condition2 ⇒ Stress	1.048	0.500	0.068	2.028	0.292	2.097	0.0360
	Likeability ⇒ Stress	-0.320	0.176	-0.665	0.025	-0.219	-1.818	0.0690

Table 1. *Indirect, direct, and total effects on the amount of information revealed (confirmation and new information together).*

Type	Effect	Estimate	SE	95% C.I. (a)		β	z	p
				Lower	Upper			
Direct	Reciprocity condition1 ⇒ Pieces of information	0.048	0.662	-1.251	1.346	0.010	0.072	0.9428
	Reciprocity condition2 ⇒ Pieces of information	1.089	0.687	-0.258	2.436	0.238	1.585	0.1129
Total	Reciprocity condition1 ⇒ Pieces of information	0.050	0.668	-1.260	1.360	0.011	0.075	0.9404
	Reciprocity condition2 ⇒ Pieces of information	0.900	0.668	-0.410	2.210	0.197	1.347	0.1781

Note. (a) Confidence intervals computed with method: Standard (Delta method). (b) Reciprocity condition 1 = "No reciprocity" – "Autonomy"; Reciprocity condition 2 = "Shared" – "Autonomy".

Figure 2. *Means and standard deviations for the amount of information revealed in each reciprocity condition.*

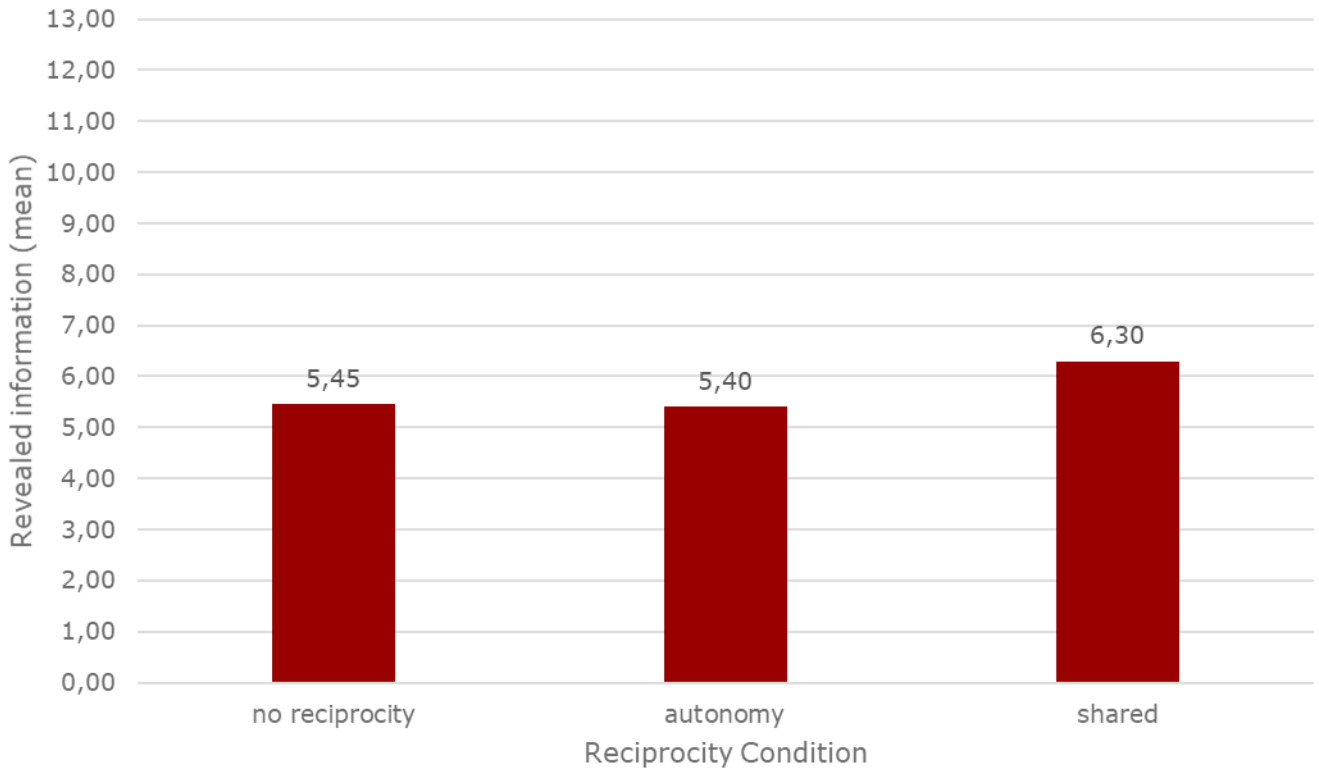


Table 2. *Indirect, direct, and total effects on the amount of revealed information perceived to be unknown to the interviewer.*

Type	Effect	Estimate	SE	95% C.I. (a)		β	z	p
				Lower	Upper			
Indirect	Reciprocity condition1 ⇒ Likeability ⇒ Pieces of information perceived as unknown	2.397e-4	0.002	-0.004	0.004	5.768e-4	0.115	0.9082
	Reciprocity condition1 ⇒ Stress ⇒ Pieces of information perceived as unknown	6.061e-4	0.009	-0.018	0.019	0.001	0.066	0.9477
	Reciprocity condition2 ⇒ Likeability ⇒ Pieces of information perceived as unknown	4.290e-4	0.003	-0.005	0.006	0.001	0.161	0.8723
	Reciprocity condition2 ⇒ Stress ⇒ Pieces of information perceived as unknown	-0.018	0.018	-0.053	0.017	-0.042	-0.996	0.3191
	Reciprocity condition1 ⇒ Likeability ⇒ Stress ⇒ Pieces of information perceived as unknown	3.187e-4	0.002	-0.004	0.005	7.669e-4	0.135	0.8924
	Reciprocity condition2 ⇒ Likeability ⇒ Stress ⇒ Pieces of information perceived as unknown	5.704e-4	0.002	-0.004	0.005	0.001	0.234	0.8146

Table 2. *Indirect, direct, and total effects on the amount of revealed information perceived to be unknown to the interviewer.*

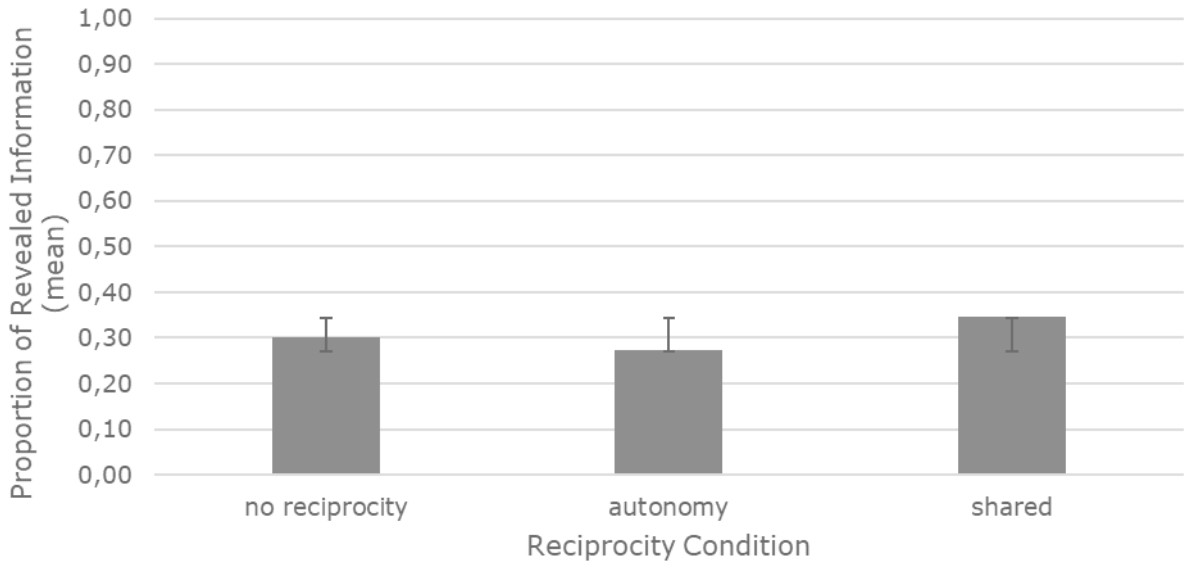
Type	Effect	Estimate	SE	95% C.I. (a)		β	z	p
				Lower	Upper			
Component	Reciprocity condition1 ⇒ Likeability	0.050	0.366	-0.668	0.768	0.020	0.137	0.8914
	Likeability ⇒ Pieces of information perceived as unknown	0.005	0.022	-0.039	0.048	0.028	0.216	0.8294
	Reciprocity condition1 ⇒ Stress	-0.033	0.499	-1.010	0.945	-0.009	-0.066	0.9476
	Stress ⇒ Pieces of information perceived as unknown	-0.018	0.016	-0.050	0.013	-0.159	-1.168	0.2429
	Reciprocity condition2 ⇒ Likeability	0.089	0.371	-0.638	0.817	0.036	0.241	0.8094
	Reciprocity condition2 ⇒ Stress	0.965	0.505	-0.025	1.955	0.267	1.910	0.0562
	Likeability ⇒ Stress	-0.345	0.177	-0.692	0.003	-0.237	-1.944	0.0519

Table 2. *Indirect, direct, and total effects on the amount of revealed information perceived to be unknown to the interviewer.*

Type	Effect	Estimate	SE	95% C.I. (a)		β	z	p
				Lower	Upper			
Direct	Reciprocity condition1 ⇒ Pieces of information perceived as unknown	0.028	0.061	-0.091	0.147	0.067	0.462	0.6444
	Reciprocity condition2 ⇒ Pieces of information perceived as unknown	0.090	0.063	-0.034	0.214	0.213	1.419	0.1560
Total	Reciprocity condition1 ⇒ Pieces of information perceived as unknown	0.029	0.061	-0.091	0.150	0.070	0.474	0.6352
	Reciprocity condition2 ⇒ Pieces of information perceived as unknown	0.073	0.062	-0.049	0.195	0.173	1.172	0.2411

Note. (a) Confidence intervals computed with method: Standard (Delta method). (b) Reciprocity condition 1 = "No reciprocity" – "Autonomy"; Reciprocity condition 2 = "Shared" – "Autonomy".

Figure 3. *Means and standard deviations of the amount of revealed information perceived to be unknown to the interviewer in each reciprocity condition.*



Appendix 2. Palena, N., Caso, L., Carlotto, G., De Mizio, L., & Marciali, M. (2017). Efficacia della tecnica di baselining nella valutazione della credibilità (Efficacy of the baseline technique when assessing credibility). *Giornale italiano di psicologia, 44*(4), 905-916. doi:10.1421/88773. [PA1]

Riassunto

La ricerca psicologica si è a lungo occupata dell'analisi del comportamento umano al fine di trovare degli indizi di menzogna, indizi che però sono deboli ed inaffidabili. D'altra parte, è necessario trovare dei metodi per valutare al meglio la credibilità delle persone intervistate. Tra le tecniche di intervista proposte, c'è l'uso di una *baseline* che possa essere utilizzata come riferimento. La letteratura ha mostrato l'inefficacia di una *small talk baseline*, ma è stato ipotizzato che una *comparable truth baseline* possa invece avere del potenziale. Questa ricerca ne ha quindi esplorato l'efficacia.

Efficacy of the baseline technique when assessing credibility

Abstract

Introduction

Research showed that a small talk baseline is ineffective for veracity assessments. On the contrary, it has been suggested that a comparable truth may be. In this experiment, we tested its efficacy.

Method

Thirty-seven participants took part in the experiment. Mean age was of M = 27.06 years and SD = 8.14. All participants performed a set of actions about which they were then interviewed. Factors were veracity condition (between) and time of the interview (within), whereas dependent variables were several verbal and nonverbal behaviours.

Results

A univariate interaction effect between time and veracity appeared for vagueness, $F(1, 35) = 4.59$, $p < .05$, visual, $F(1, 35) = 4.68$, $p < .05$, and spatial details, $F(1, 35) =$

7.10, $p < .02$. Follow-up analyses demonstrated that liars only showed differences during the interview.

Conclusion

Our results support the idea that a comparable truth baseline may have potential as an interviewing technique for veracity assessments. Its efficacy can be rooted in interpersonal dynamics and intrapersonal differences, as different persons may show different cues to deception.

Introduzione

La ricerca di indizi legati all'inganno ha alle sue spalle circa trent'anni di sperimentazione. In principio l'attenzione era volta alla tentata costruzione di strumentazioni atte a valutare l'onestà di un sospettato. Tra queste, sicuramente la più conosciuta è il poligrafo, comunemente chiamato macchina della verità (Vrij, 2008). Altre strumentazioni che registrano parametri psicofisiologici sono state inventate e testate nella loro efficacia, e tra queste c'è il recente utilizzo dei potenziali evento relati (Rosenfeld, Soskins, Bosh & Ryan, 2004).

Un altro filone di ricerca è quello che si focalizza sulla ricerca di indizi di menzogna nel linguaggio del corpo, nelle espressioni facciali e nel contenuto verbale di un resoconto (Ekman, 2001; Ekman, O'Sullivan, Friesen e Shrer, 1991). In effetti, essere in grado di valutare la credibilità di un interlocutore senza l'assistenza di apparecchiature elettroniche sarebbe più adatto per una valutazione fatta in tempo reale, utile ad esempio nella raccolta di informazioni da parte dell'intelligence (Ormerod e Dando, 2014).

Sfortunatamente, la ricerca ha mostrato che sebbene esistano degli indizi comportamentali e verbali legati alla menzogna, questi sono inaffidabili (DePaulo *et al.*, 2003; Vrij, 2008). In effetti, alcune ricerche hanno mostrato che la maggior parte della varianza nell'accuratezza degli osservatori potrebbe risiedere nell'emittente piuttosto che nel ricevente (Levine, 2010; Levine *et al.*, 2011). Ovvero, l'accuratezza nel discriminare verità da bugia non risiede tanto nella bravura dell'osservatore, quanto piuttosto nella trasparenza del mittente (Levine *et al.*, 2011).

Questo è inevitabilmente legato alla presenza di differenze interpersonali: ciò che per una persona potrebbe mostrare inganno, per un'altra potrebbe essere indice di onestà. Per questo motivo sono necessarie delle tecniche di intervista che permettano di migliorare l'accuratezza degli osservatori (Vrij, 2014). E difatti, al giorno d'oggi, la ricerca non è più orientata alla scoperta di nuovi indizi di menzogna, quanto allo sviluppo di tecniche di intervista investigativa che permettano di aumentare le differenze tra quando si mente e quando si è onesti (Deeb *et al.*, 2016; Vrij e Granhag, 2012). Tra queste ci sono alcune tecniche di utilizzo strategico delle prove, o SUE (Tekin *et al.*, 2015) ed altre che si basano più sul processamento cognitivo delle informazioni, che in diversi casi è maggiore per chi mente che per chi è onesto (Vrij, 2015). Infine, una tecnica spesso proposta (soprattutto dalle forze di polizia statunitensi) è l'utilizzo di una baseline, ovvero di un riferimento o linea base comportamentale (Frank, Yarbrough e Ekman, 2006; Inbau, Reid, Buckley e Jayne, 2013). L'idea è che, al fine di ridurre gli errori di valutazione dovuti alle differenze interpersonali, un investigatore dovrebbe prima porre delle domande che non mettano in ansia l'intervistato, per registrarne i pattern comportamentali e verbali. Una volta ottenuta tale linea base, o *baseline*, l'intervistatore dovrebbe cominciare a porre domande investigative e se il comportamento cambia, concludere che l'intervistato sta probabilmente mentendo. L'idea non è concettualmente sbagliata alla base, ma è errato il modo in cui viene ottenuta tale baseline. Come Ewens e colleghi (2014) e Vrij (2008) sottolineano, un tale tipo di baseline potrebbe aumentare gli errori di giudizio più che diminuirli, perché il coinvolgimento emotivo e cognitivo legato a quando si parla di sé stessi o di aspetti neutrali è ovviamente diverso dal coinvolgimento che si ha quando si parla di fatti e situazioni sotto indagini di polizia (Ewens *et al.*, 2014), così come parlare di questioni imbarazzanti comporta dei cambiamenti comportamentali rispetto a quando si parla di situazioni neutre (Kleinke, 1986). Inoltre, la ricerca ha mostrato che il comportamento cambia sia tra interviste ripetute nel tempo che durante un'unica intervista (Caso, Maricchiolo, Bonaiuto, Vrij e Mann, 2006; Gnisci, Caso e Vrij, 2010; Granhag e Strömwall, 2002). Ewens e colleghi (2014) hanno valutato l'efficacia di una tecnica di *baselining* chiamata *small talk*, dove la baseline è costruita proprio attraverso domande neutre in quanto riferite a questioni non legate all'evento sotto indagine, come proposto da Frank e colleghi

(Frank, Yarbrough e Ekman, 2006). Gli autori, nel loro studio, hanno chiesto ai partecipati di rispondere ad alcune domande riguardanti il proprio lavoro (mentendo o essendo onesti, a seconda della condizione sperimentale) dopo aver risposto ad una domanda di baseline riferita al rievocare ciò che i partecipanti avevano letto nel consenso informato che avevano firmato. I risultati hanno mostrato che sia gli onesti che i disonesti cambiavano i propri comportamenti tra le diverse fasi dell'intervista. Nello specifico, gli autori avevano analizzato il livello di carico cognitivo e di tentato controllo (dei comportamenti non verbali, del proprio eloquio ecc). Essi hanno quindi concluso che tale tipo di baseline è inefficace in quanto se è vero che i disonesti cambiano il proprio comportamento tra le due fasi dell'intervista, anche gli onesti lo fanno, e ciò può portare ad errori interpretativi (soprattutto falsi positivi). Ciò nonostante, gli autori sostengono che utilizzare una baseline più adatta, che chiamano "*comparable truth*" potrebbe invece risultare efficace. Per *comparable truth* si intende una baseline che sia paragonabile alla fase target dell'intervista rispetto al coinvolgimento emotivo, cognitivo, al setting dell'intervista, al contenuto del racconto e così via. Vrij e Mann (2001) sono i primi ad aver esplorato l'utilizzo di una tale tecnica di *baselining* in un contesto reale. Hanno difatti analizzato il video di un'intervista di un assassino (reale) ed hanno trovato che egli ha mostrato cambiamenti tra le fasi dell'intervista. Gli autori concludono che quindi la baseline *comparable* sembra in effetti avere un certo potenziale. D'altra parte, non è possibile sapere se anche un onesto avrebbe mostrato gli stessi cambiamenti, così come non è possibile trarre alcuna conclusione dal momento che lo studio era focalizzato su un singolo individuo.

Per questi motivi, è stato sviluppato l'esperimento in essere per esplorare l'efficacia di una *comparable truth baseline*. Oltre alle variabili di tipo non verbale, è stato analizzato il contenuto verbale del racconto dei partecipanti, dal momento che gli aspetti verbali paiono essere più efficaci che quelli non verbali quando si tratta di valutare l'onestà di una persona (Bond e DePaulo, 2006). Alcune variabili sono state misurate su scala likert, mentre altre sono state misurate su frequenza assoluta.

Se la *comparable truth baseline* è efficace, i disonesti dovrebbero mostrare delle differenze maggiori rispetto agli onesti tra i due tempi dell'intervista. Quindi, si prevede un'interazione significativa tra il momento dell'intervista e la condizione di

onestà dei partecipanti sia per quanto riguarda le variabili misurate su scala likert (Ipotesi 1) che per quelle misurate con frequenza assoluta (Ipotesi 2).

Metodo

Partecipanti

Il campione era composto da 37 partecipanti (24 femmine e 13 maschi). L'età era compresa tra 21 e 52 anni, con una media *M* = 27.06 ed una deviazione standard *DS* = 8.14.

Disegno sperimentale

Il disegno sperimentale prevedeva in primo luogo un'analisi multivariata della varianza (MANOVA) con disegno misto, ed in secondo luogo l'analisi degli effetti univariati (ANOVA) per ognuna delle singole variabili dipendenti. Le variabili indipendenti erano il momento dell'intervista (fase di *baseline* vs. fase target, within) e la condizione di veridicità (onesto vs. disonesto, between). Per quanto riguarda la variabile momento dell'intervista, i partecipanti rispondevano (sempre onestamente) ad una domanda relativa ad un primo set di azioni che era stato chiesto loro di eseguire (fase di *baselining*) e ad un'altra domanda (onestamente o mentendo, a seconda della condizione sperimentale) riguardante un secondo set di azioni che era stato chiesto loro di eseguire (fase target). Per quanto riguarda la condizione di veridicità i partecipanti, nella sola domanda target, dovevano rispondere onestamente o mentendo alla domanda posta. Le variabili dipendenti erano: livello di carico cognitivo, rigidità del corpo, dettaglio, struttura logica del racconto, produzione non strutturata e vaghezza, tutte calcolate su scala likert a quattro punti. Inoltre, è stata analizzata la frequenza assoluta dei dettagli visivi, auditivi, spaziali, temporali e d'azione. Sono state condotte due MANOVAs, seguite dalle relative analisi univariate: una per le variabili calcolate su scala likert ed una per quelle misurate con frequenza assoluta.

Procedura

Una lista di studenti interessati a partecipare è stata ottenuta durante alcune lezioni universitarie, e ai partecipanti sono stati offerti due punti aggiuntivi per un esame universitario. I partecipanti sono anche stati informati riguardo la durata

dell'esperimento, che si aggirava attorno ai 40 minuti. Per prima cosa, i partecipanti hanno letto e firmato il consenso informato per la partecipazione. È stato poi detto loro che avrebbero eseguito alcune azioni riguardo alle quali sarebbero poi stati intervistati e videoregistrati (per la videoregistrazione è stato ottenuto il consenso). Le azioni erano divise in due set. Nel primo set di azioni i partecipanti ricevevano una busta chiusa, la quale conteneva una password per accedere ad un pc. Non appena lo sperimentatore lasciava la stanza, i partecipanti accedevano ad un laptop e leggevano un file che indicava loro di cercare un cd-rom all'interno di uno zaino che era presente nella stanza. I partecipanti dovevano visionare un video presente nel cd-rom ove compariva una persona che informava loro di cercare una chiave nascosta nello stesso zaino dal quale avevano preso il CD. Tale chiave serviva per aprire una cassetta di sicurezza che conteneva ulteriori informazioni. Queste istruivano i partecipanti a mandare una e-mail ad uno specifico destinatario ed attendere l'arrivo di una persona che li avrebbe riconosciuti (un confederato). L'incontro con il confederato serviva come punto di riferimento per differenziare la fase di *baseline* da quella di target. Tutto ciò che era accaduto prima dell'incontro con il confederato faceva parte degli eventi riguardanti la *baseline*, mentre tutto ciò che è accaduto dopo faceva parte degli eventi riguardanti la fase target dell'intervista. Quest'ultima, in caso di reale intervista investigativa, sarebbe la situazione sotto indagini di polizia.

Il secondo insieme di azioni che i partecipanti dovevano compiere cominciava con il leggere le istruzioni che il confederato aveva loro consegnato, nascoste all'interno di un quotidiano. Tali indicazioni istruivano i partecipanti ad entrare in una stanza adiacente a quella dove avevano cominciato l'esperimento al fine di cercare una chiave USB attaccata ad un attaccapanni. Inoltre, i partecipanti dovevano cercare un libro nascosto in un armadio presente all'interno della stanza il quale nascondeva al suo interno un'altra chiavetta USB. Il compito dei partecipanti era di scambiare le due chiavette, tenere la seconda con loro fino alla fine dell'esperimento, nascondere quotidiano e istruzioni (insieme al libro con la prima chiavetta USB nascosta in esso) all'interno dell'armadio e tornare nella prima stanza per attendere l'intervistatore.

Successivamente, un intervistatore entrava nella stanza, si presentava e cominciava l'intervista. In primis egli poneva una domanda aperta ai partecipanti riguardante il primo set di azioni. Terminata la risposta dei partecipanti, l'intervistatore si

allontanava con una scusa dicendo che sarebbe ritornato a breve. In quel momento entrava il confederato il quale forniva ulteriori istruzioni al partecipante. Per prima cosa, lo informava che l'intervistatore non era a conoscenza di ciò che stava accadendo, ovvero del fatto che il confederato stesse dando istruzioni aggiuntive ai partecipanti. In aggiunta, i partecipanti allocati alla condizione "onesti" venivano informati di dover continuare a rispondere onestamente alle domande dell'intervistatore. Al contrario, i partecipanti nella condizione "bugiardi" venivano istruiti a mentire riguardo a ciò che avevano fatto dopo l'incontro con il confederato, ma di continuare a rispondere in maniera onesta ad ogni eventuale altra domanda riguardate ciò che era successo prima dell'incontro con lo stesso. La bugia doveva consistere in una serie di 4-5 azioni ed essere abbastanza dettagliata da risultare credibile come alibi. Infine, il confederato si accertava che i partecipanti avessero capito le istruzioni e si allontanava dalla stanza, dando loro cinque minuti per preparare la storia. Poco dopo, l'intervistatore tornava scusandosi nuovamente per l'essersi allontanato. A questo punto egli rivolgeva una seconda domanda aperta ai partecipanti, riguardo ciò che era successo dopo l'incontro con il confederato. Alla fine dell'intervista i partecipanti hanno compilato un questionario riguardante le informazioni demografiche e il livello di motivazione (su una scala a 5 punti, dove 5 corrispondeva al massimo livello di motivazione). Infine, hanno ricevuto un *debriefing* rispetto agli obiettivi dello studio.

Coding

Un *coder*, che non era a conoscenza degli obiettivi dell'esperimento e delle condizioni sperimentali, ha codificato tutti i video per l'ammontare, su una scala che andava da 1 (completamente assente) a 4 (molto presente), del livello di carico cognitivo (quanto l'intervistato appare concentrato nel ponderare ciò che dice), rigidità del corpo (quanto la persona pare rigida, ferma, e con i muscoli tesi), dettagli (quanto è dettagliato il resoconto dell'intervistato), struttura logica del racconto (quanto il resoconto dell'intervistato è logico e internamente coerente), produzione non strutturata (quanto il resoconto è flessibile e non rigido nella sua struttura interna) e vaghezza (quanto l'intervistato appare vago nell'esposizione del resoconto). Inoltre, è stata codificata la frequenza assoluta dei dettagli visivi (ciò che il partecipante ha visto: c'era una sedia rossa), auditivi (ciò che è stato udito: ho sentito un rumore), spaziali (riferimenti a

luoghi oppure organizzazioni spaziali: la sedia era a sinistra del tavolo), temporali (riferimenti a momenti precisi o all'organizzazione temporale degli avvenimenti: è arrivato dopo che avevo chiuso la porta) e di azione (ciò che è stato fatto: ho aperto il cassetto) Un secondo *coder* ha codificato il 25% dei video al fine di valutare l'accordo inter-giudice (ICC), prassi precedentemente utilizzata in esperimenti sulla valutazione della credibilità che richiedevano una fase di coding (Ewens *et al*., 2014; Nahari & Vrij, 2015; Shaw *et al*., 2015). Il più basso livello di accordo è stato raggiunto per la variabile dettagli temporali, ICC = .86, mentre il più alto per la variabile dettagli d'azione, ICC = .96.

Risultati

È stata condotta una prima MANOVA con il momento dell'intervista (*baseline* vs. target) come variabile indipendente within e con la condizione di veridicità (onesto vs. disonesto) come variabile indipendente between, e con le variabili dipendenti livello di carico cognitivo, rigidità, dettaglio, coerenza logica, produzione non strutturata e vaghezza. A livello multivariato, l'interazione *momento dell'intervista * condizione di veridicità* non è risultata statisticamente significativa, $F(6, 30) = .86$, $p = .53$. A livello univariato l'interazione è stata significativa per la variabile "vaghezza", $F(1, 35) = 4.59$, $p = .04$. L'ipotesi 1 è stata quindi supportata solo per questa variabile. D'altra parte, al fine di meglio interpretare un'interazione, sono necessarie delle analisi di follow-up, la cui variabile indipendente è il tempo dell'intervista e la variabile dipendente il livello di vaghezza del mittente

Quando i soli onesti sono stati presi in considerazione, non è apparsa una differenza significativa tra i due momenti dell'intervista, $F(1, 18) = .00$, $p = 1$. I mittenti sono infatti risultati egualmente vaghi nella fase dei baselining ($M = 1.68$, $DS = 1.05$) ed in quella target ($M = 1.68$, $DS = .94$). Quando solo i disonesti sono stati presi in considerazione, le analisi di follow-up hanno mostrato come i mittenti tendessero ad essere più vaghi nella fase di target, dove mentivano ($M = 2.33$, $DS = 1.13$) rispetto a quella di baselining, dove dicevano la verità ($M = 1.72$, $DS = .82$), $F(1, 17) = 6.25$, $p = .02$, *Cohen's d* = .61.

Una seconda MANOVA è stata condotta con la frequenza assoluta dei dettagli visivi, auditivi, spaziali temporali e di azione come variabili dipendenti.

A livello multivariato, l'interazione *momento dell'intervista * condizione di veridicità* ha solo approcciato la significatività statistica, $F(5, 31) = 2.43$, $p = .057$. A livello univariato, tale interazione è risultata essere significativa per i dettagli spaziali, $F(1, 35) = 7.10$, $p = .01$ e quelli visivi, $F(1, 35) = 4.68$, $p = .04$, in parziale supporto dell'ipotesi 2.

Le analisi di follow-up hanno mostrato che i partecipanti assegnati alla condizione onesto hanno mantenuto un comportamento verbale simile tra le due fasi dell'intervista quando venivano confrontati questi tipi di dettagli. Infatti, la frequenza dei dettagli spaziali riportati nella fase di *baselining* ($M = 7.63$, $DS = 2.69$) non è apparsa significativamente diversa da quella relativa alla fase *target* ($M = 8.10$, $DS = 2.72$), $F(1, 18) = .95$, $p = .34$. Similmente, la frequenza dei dettagli visivi riportati nella fase di *baselining* ($M = 17.63$, $DS = 6.51$) non si è discostata da quella nella fase *target* ($M = 19.68$, $DS = 8.23$) in maniera significativa, $F(1, 18) = .86$, $p = .36$.

D'altra parte, quando sono stati analizzati i dati riguardanti i soli partecipanti nella condizione disonesto, le analisi hanno mostrato una differenza significativa tra le due fasi di intervista sia per i dettagli spaziali, $F(1, 17) = 5.97$, $p = .03$, che per quelli visuali, $F(1, 17) = 4.80$, $p = .04$ (Tabella 1).

Inserire Tabella 1 qui

Discussioni

In questo esperimento abbiamo testato l'efficacia di una *comparable truth baseline* nel valutare la credibilità della persona intervistata. I risultati hanno evidenziato che solo i partecipanti nella condizione di disonestà hanno mostrato dei cambiamenti significativi tra le due fasi di intervista, effetto che è emerso dalle analisi univariate di follow-up. Nello specifico, nel momento in cui i partecipanti hanno cominciato a mentire, essi hanno mostrato un maggior livello di vaghezza, ma una minore quantità di dettagli spaziali e visuali. Tali risultati vanno nella direzione proposta dalla letteratura, dal momento che è stato evidenziato come l'atto di mentire porti ad un maggiore impegno cognitivo e ad una minore quantità di dettagli nell'eloquio (DePaulo *et al.*, 2003; Vrij, 2008).

Le ipotesi riguardanti una maggior efficacia di una *comparable truth baseline* rispetto ad una *small talk baseline* avanzate in questo studio e proposte precedentemente (Ewens *et al.*, 2014; Vrij, 2008) hanno quindi ricevuto supporto e paiono evidenziare che non è la tecnica di *baselining* ad essere inefficace, quanto piuttosto il modo con cui questa viene creata. Infatti, al contrario di quanto proposto dalle forze di polizia e da alcuni accademici (Frank, Yarbrough e Ekman, 2006; Inbau, Reid, Buckley e Jayne, 2013) quando una *baseline* non è paragonabile alla fase target dell'intervista essa non solo risulta inefficace (Ewens *et al.*, 2014) ma anzi aumenta il rischio di falsi positivi, dal momento che anche gli onesti, mostrando cambiamenti tra le varie fasi dell'intervista, potrebbero essere erroneamente giudicati bugiardi. Ciò probabilmente accade in quanto il coinvolgimento emotivo e cognitivo e la posta in gioco sono molto diversi tra quando si parla di questioni irrilevanti rispetto a quando si parla di aspetti sotto indagine, ove il livello di sospettosità dell'intervistatore cambia (Caso, Gnisci, Vrij e Mann, 2005).

È plausibile che l'efficacia di una *comparable truth baseline* sia in parte dovuta al fatto che questa permette di ridurre gli errori dovuti a differenze interpersonali ed al livello di trasparenza del nostro interlocutore. In effetti qualcuno potrebbe cambiare il proprio comportamento in modo inaspettato, ad esempio essendo più dettagliato quando è disonesto rispetto a quando dice la verità. D'altra parte, avendo a disposizione un metro di confronto che sia efficace, non ci dovrebbero essere motivi per i quali, essendo onesta, una persona mostri dei cambiamenti molto evidenti tra due fasi dell'intervista. Ciò non significa che gli onesti non cambiano il comportamento durante un'intervista, anzi, sappiamo che è vero il contrario (Caso, Maricchiolo, Bonaiuto, Vrij e Mann, 2006; Gnisci, Caso e Vrij, 2010; Granhag e Strömwall, 2002; Kleinke, 1986). I risultati di questo studio paiono però supportare l'ipotesi che, comunque, queste differenze siano maggiori e più evidenti nel momento in cui si comincia a mentire.

Sebbene i risultati di questo studio siano positivi, è necessario che ricerche future valutino se una *comparable truth baseline* è efficace anche al cambiare del contenuto dell'intervista. Da un punto di vista teorico, se si continua a mantenere costante la coerenza delle varie fasi dell'intervista, non dovrebbero esserci motivi per i quali essa perda di efficacia. D'altra parte, è importante verificare che questo sia di fatto il caso.

Ricerche future potrebbero ad esempio valutare l'efficacia di una *comparable truth baseline* quando il tema dell'intervista riguarda aspetti che non siano prettamente una serie di azioni come quelle presenti in questa ricerca, ma ad esempio le attività che sono state svolte durante la giornata (Vrij e Mann, 2001). Inoltre, è importante verificare che questa tecnica sia efficace anche quando la posta in gioco aumenta, ovvero in casi reali, al fine di affinarla sempre di più, così come appare evidente la necessità di sviluppare un metodo di scoring utilizzabile in tempo reale. Infatti, nonostante le indagini di polizia possono essere registrate e codificate in un secondo momento, sarebbe più pratico sviluppare un metodo di valutazione utilizzabile in tempo reale. Un altro importante passo da fare è quello di integrare insieme la tecnica di *baselining* ed altre tecniche che hanno già mostrato la loro efficacia, come ad esempio l'*Assessment Criteria Indicative of Deception* (Colwell, Hiscock-Anisman e Fede, 2013).

È inoltre importante sottolineare un altro aspetto: il ruolo della preparazione all'intervista. I partecipanti di questo studio hanno avuto cinque minuti per preparare la loro storia. Alcune situazioni di vita reale sono simili al setting utilizzato in questo esperimento, dove la preparazione all'intervista è limitata o completamente assente. Ciò accade ad esempio quando i mittenti vengono intervistati durante un fermo (improvviso) di polizia, ad esempio per un controllo. D'altra parte, ci sono altre situazioni dove è possibile che la persona intervistata abbia molto tempo per preparare la sua storia. Una persona che mente, anche quando preparata, mostra comunque alcuni indizi di menzogna causati da diversi processi psicologici, sociali e comunicativi, tra i quali il carico cognitivo ed emotivo (Vrij, 2015), il non aver vissuto effettivamente l'evento, con il conseguente ruolo della memoria, ed altri (cfr. DePaulo et al., 2003 e Vrij, 2008). D'altra parte, anche se l'efficacia della tecnica di *comparable truth baseline* sarà verosimilmente mantenuta- anche tenendo in considerazione il ruolo del *motivational impairment effect* (DePaulo & Kirkendol, 1989) - studi futuri dovrebbero esplorare l'efficacia della stessa quando ai mittenti viene data la possibilità di prepararsi più a fondo per l'intervista.

Inoltre, considerando che questo studio è il primo che valuta sperimentalmente l'effetto della tecnica di *comparable truth baseline*, il prossimo passo che ricerche

future dovrebbero fare è affinare la tecnica stessa e valutare quanti soggetti possono essere correttamente identificati come onesti o bugiardi attraverso il suo utilizzo.

Infine, questa ricerca ha sottolineato importanti aspetti applicativi. Primo, visto il comune utilizzo della tecnica di *baselining* da parte delle forze di polizia, è essenziale che una *comparable truth baseline* (una volta affinata) vada a sostituirsi all'utilizzo della *small talk baseline*, dal momento che è evidente come quest'ultima non funzioni (Ewens e colleghi, 2014). Secondo, una *comparable truth baseline* potrebbe offrire la possibilità di supportare la presa di decisione nei singoli casi sotto indagine se verrà adeguatamente sviluppata e sarà utilizzata insieme ad altre tecniche di indagine che già hanno mostrato la loro efficacia, come il SUE *(*Tekin *et al.*, 2015*)*

Referenze

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

Caso L., Gnisci A., Vrij A., Mann S. (2005). Processes underlying deception: an empirical analysis of truth and lies when manipulating the stakes. *Journal of Investigative Psychology and Offender Profiling*, *2*(3), 195-202.

Caso L., Maricchiolo F., Bonaiuto M., Vrij A., Mann S. (2006). The impact of deception and suspicion on different hand movements. *Journal of Nonverbal behavior, 30*(1), 1-19.

Colwell K., Hiscock-Anisman C., Fede J. (2013). Assessment criteria indicative of deception: An example of the new paradigm of differential recall enhancement. In *Applied issues in investigative interviewing, eyewitness memory, and credibility assessment*. Springer: New York, pp. 259-291.

Deeb H., Vrij A., Hope L., Mann S., Granhag P. A., Lancaster G. L. J. (2016). Suspects' consistency in statements concerning two events when different question formats are used. *Journal of Investigative Psychology and Offender Profiling*, doi: 10.1002/jip.1464.

DePaulo, B. M., & Kirkendol, S. E. (1989). The motivational impairment effect in the communication of deception. In J. C. Yuille (Ed.) *Credibility Assessment* (pp. 51-70). Dordrecht, the Netherlands: Kluwer.

DePaulo B. M., Lindsay J. J., Malone B. E., Muhlenbruck L., Charlton K., Cooper H. (2003). Cues to deception. *Psychological Bulletin, 129*(1), 74-118.

Ekman P. (2001). *Telling lies*. New York: Norton.

Ekman P., O' Sullivan M., Friesen W. V., Shrer K. (1991). Face, Voice, and Body in detecting deceit. *Journal of Nonverbal Behavior, 15,* 125-135.

Ewens S., Vrij A., Jang M., Jo E. (2014). Drop the small talk when establishing baseline behaviour in interviews. *Journal of Investigative Psychology and Offender Profiling*. DOI: 0.1002/jip.1414.

Frank M. G., Yarbrough J. D., Ekman P. (2006). Investigative interviewing and the detection of deception. In T. Williamson (Eds.), *Investigative interviewing: Rights, research and regulation*. Cullompton, Devon: Willan, pp. 229-255.

Gnisci A., Caso L., Vrij A. (2010). Have you made up your story? The effect of suspicion and liars' strategies on reality monitoring. *Applied cognitive psychology*, *24*(6), 762-773.

Granhag P. A., Strömwall L. A. (2002). Repeated interrogations: Verbal and non-verbal cues to deception. *Applied Cognitive Psychology, 16,* 243–257. doi: 10.1002/acp.784

Inbau F. E., Reid J. E., Buckley J. P., Jayne B. C. (2013*). Criminal interrogation and confessions*, 5th edition. Burlington, MA: Jones & Bartlett Learning.

Kleinke C. L. (1986). Gaze and eye contact: a research review. *Psychological bulletin, 100,* 78. doi: 10.1037/0033-2909.100.1.78.

Levine T. R. (2010). A few transparent liars: Explaining 54% accuracy in deception detection experiments. In C. Salmon (Ed.), *Communication Yearbook 34.* Sage, pp. 44-61.

Levine T. R., Serota K. B. Shulman H., Clare D. D., Park H. S., Shaw A. S., Shim J. C., Lee J. H. (2011). Sender demeanor: Individual differences in sender believability have a powerful impact on deception detection judgments. *Human Communication Research, 37,* 377-403.

Nahari, G., & Vrij, A. (2015). Can someone fabricate verifiable details when planning in advance? It all depends on the crime scenario. *Psychology, Crime & Law, 21*(10), 987-999.

Ormerod T. C., Dando C. J. (2015). Finding a needle in a haystack: Toward a psychologically informed method for aviation security screening. *Journal of Experimental Psychology: General*, *144*(1), 76.

Rosenfeld J. P., Soskins M., Bosh G., Ryan A. (2004). Simple, effective countermeasures to P300-based test of detection of concealed information. *Psychophysiology, 41,* 205-219.

Shaw, D. J., Vrij, A., Leal, S., Mann, S., Hillman, J., Granhag, P. A., & Fisher, R. P. (2015). Mimicry and investigative interviewing: Using deliberate mimicry to elicit information and cues to deceit. *Journal of Investigative Psychology and Offender Profiling*, *12*(3), 217-230.

Tekin S., Granhag P. A., Stromwall L., Mac Giolla E., Vrij A., Hartwing M. (2015). Interviewing Strategically to Elicit Admissions From Guilty Suspects. *Law and Human Behavior, 39* (3), 244-252.

Vrij A. (2008). *Detecting lies and deceit: Pitfalls and opportunities.* Chichester: John Wiley and Sons.

Vrij A. (2014). Interviewing to detect deception. *European Psychologist*, *19*(3), 184-194. DOI: 10.1027/1016-9040/a000201.

Vrij A. (2015). A cognitive approach to lie detection. In P. A. Granhag, A. Vrij, & B. Verschuere (Ed), *Detecting deception: Current challenges and cognitive approaches*. Chichester: John Wiley and Sons, pp. 205-230.

Vrij A., Granhag P. A. (2012). Eliciting cues to deception and truth: What matters are the questions asked. *Journal of Applied Research in Memory and Cognition*, *1*(2), 110-117.

Vrij A., & Mann S. (2001). Telling and detecting lies in a high-stake situation: the case of a convicted murderer. *Applied Cognitive Psychology*, *15*, 187–203.

Tabella 1. *Cambiamenti delle variabili dipendenti a frequenza assoluta tra i due tempi dell'intervista per i partecipanti nella condizione disonesto.*

	F (1, 17)	*p*	*M* (T1)	*SD* (T1)	*M* (T2)	*SD* (T2)	*d*
Dettagli Spaziali	5.97	.03	7.05	3.94	4.67	3.83	-. 57
Dettagli Visivi	4.81	.04	18.83	9.93	14.33	7.42	-. 47

Appendix 3: Palena, N., Caso, L., Vrij, A., & Orthey, R. (2018). Detecting deception through small talk and comparable truth baselines. *Journal of Investigative Psychology and Offender Profiling, 15*(2), 124-132. doi:10.1002/jip.1495. [PA2]

Abstract

The present experiment investigates similarities in participants' nonverbal and verbal behaviours when responding to baseline and investigative questions, comparing two different types of baselines. Police literature suggests to obtain a baseline through small talk, whereas academic literature underlines the importance of baseline and investigative themes to be comparable. First, a baseline was obtained (either small talk or comparable), then the investigative questioning started. During the investigative questioning participants either truthfully reported a set of actions they had performed or lied about them. Findings revealed that truth tellers and liars in the small talk condition did not differ in their level of similarity when responding to the baseline and investigative questions. In the comparable truth condition, levels of verbal similarity between the baseline and investigative questions were higher for truth tellers than for liars but only for one variable: spatial detail. Results therefore showed that a small talk baseline should not be used to assess interviewees' credibility, and that a comparable truth baseline, although better than a small talk baseline, is still problematic.

Keywords: detecting deception, baseline, similarity scores

Meta-analyses into (non)verbal cues to deception have shown that such cues are faint and unreliable (DePaulo, Lindsay, Malone, Muhlenbruck, Charlton & Cooper, 2003; Hartwig & Bond, 2011). Consequently, researchers started to focus on specific interview techniques to enhance existing or elicit new cues to deceit (Vrij & Granhag, 2012). These include cognitive lie detection (Vrij, 2014, 2015; Vrij, Fisher, Blank, Leal & Mann, 2016), the Assessment Criteria Indicative of Deception (Colwell, Hiscock-Anisman, Memon, Taylor & Prewett, 2007; Colwell, Hiscock-Anisman, Memon, Colwell, Taylor & Woods, 2009), and the Strategic Use of Evidence (Granhag & Hartwig, 2008, 2015).

These techniques have been developed within the academic setting. On the contrary, a strategy frequently proposed and used within the police culture is the use of a small talk baseline (Ewens, Vrij, Jang & Jo, 2014; Frank, Yarbrough & Ekman, 2006; Inbau, Reid, Buckley & Jayne, 2013). The core of the baseline technique is that interviewers should in first instance evaluate interviewees' baseline behaviour using non-threatening questions (small talk) and then assess if differences in behaviours between this phase and the investigative phase of the interview arise. If this is the case, then it is thought that deception may be occurring.

Frank and colleagues (2006) proposed such an implementation of the baseline technique, called the Improving Interpersonal Evaluations for Law Enforcements and National Security Technique (Frank *et al*., 2006). Others believe that this version of the technique is misleading (Ewens *et al*., 2014, Moston & Engelberg, 1993). The problem is that behaviours in response to non-threatening questions cannot be compared to behaviours in response to questions asked in the investigative interview. The baseline component is a relatively low-stakes situation whereas the interview is a relatively high-stakes situation and people show different behaviours in low versus high stakes situations (Hartwig & Bond, 2014; Ioannou & Hammond, 2015). In fact, behaviours are influenced by many factors other than deception. People react differently when they are accused of wrongdoing than when they are unchallenged (Vrij, 2006), and show different behaviours when they are interviewed by different people (Vrij & Winkel, 1991). Moreover, behaviour is topic-related. People respond

differently when discussing a topic that embarrasses them than when discussing a neutral topic (Kleinke, 1986), and they respond differently when they discuss a topic they care about or is important to them than when they discuss a topic with which they have less personal involvement (Davis & Hadiks, 1995; Matarazzo, Wiens, Jackson, & Manaugh, 1970). Finally, people's behaviour sometimes changes over time in the same interview (Buller & Burgoon, 1996; Burgoon, Buller, White, Afifi, & Buslig, 1999; Stiff, Corman, Krizek, & Snider, 1994), or, if they are interviewed on more than one occasion, changes may occur over repeated interviews (Granhag & Strömwall, 2002).

Therefore, when researchers wish to compare a person's deceptive nonverbal response with a truthful nonverbal response from the same person, they need to make sure that the deceptive and truthful responses are taken from the same interview setting; that the person talks about similar topics in the deceptive and truthful parts; and that these parts were discussed within a short period of time from each other. In that respect, Vrij (2008) and Ewens *et al*. (2014) refer to a *comparable truth* baseline. Comparable means that the baseline the investigator uses must be similar in content, stakes, and cognitive and emotional involvement to investigative questions. Additionally, the contexts within which the baseline and the target questioning occur must be similar (Vrij, 2008).

In the present experiment we compared the effect of a comparable truth baseline with the effect of a small talk baseline, typically used by police and proposed by Frank *et al*. (2006). We compared a series of nonverbal and verbal behaviours displayed during baseline and the interview and calculated a similarity score: The higher the score, the higher the similarity. We did not expect the similarity scores for nonverbal or verbal behaviours to differ between truth tellers and liars in the small talk baseline condition, due to problems associated with this type of baseline comparison outlined above. The comparable truth baseline technique might work, and perhaps even more so for verbal behaviours, as these are typically more diagnostic cues to deceit than nonverbal behaviours (Bond & DePaulo, 2006). We therefore predicted the following interaction effect: Truth tellers and liars will obtain similar similarity scores in the small talk baseline condition but truth tellers will obtain higher similarity scores in the comparable truth baseline condition (Hypothesis 1). The effect sizes in the

comparable truth baseline condition may be more pronounced for verbal behaviours than for nonverbal behaviours (Hypothesis 2).

Method

Participants

A total of 69 participants (49 female subjects and 20 male subjects) took part in the experiment. Age ranged from 21 to 53 years, with an average age of M = 26.88 years old (SD = 8.10).

Design

The experiment utilised a 2 (Baseline: comparable truth vs. small talk) X 2 (Veracity: truth vs. lie) between subjects design. Participants were all honest in Phase one (baseline) but they were randomly assigned to the type of baseline and veracity conditions. Regarding the type of baseline, participants either had to report truthfully a set of actions they were asked to perform which were similar to the set of actions they had to perform and were interviewed about in the actual interview (comparable truth condition) or had to tell something truthfully about their last year as a student and/or as a worker (small talk condition). Regarding the veracity condition, participants had either to tell the truth or to lie about a second set of tasks they were asked to perform. The dependent variables were the similarity scores for hands and finger movements, one arm movements, two arms movements, spatial, temporal, visual, audio and action details. We examined hands and fingers movements as meta-analyses and literature reviews show that it can discriminate between truth tellers and liars to a moderate extent (DePaulo *et al.*, 2003; Vrij, 2008). We examined one arm and two arms movements as physical rigidity may be related to attempted control, which is thought to be related to lying (Zuckerman, DePaulo & Rosenthal, 1981). Finally, we examined spatial, temporal, visual, audio and action details as they are related to the verbal veracity tool Reality Monitoring, which has a solid theoretical foundation (Johnson & Raye, 1981; Vrij, 2008). The level of motivation, measured on a 5 points likert scale, was set as a covariate.

Procedure

A list of participants was obtained during university lectures. Students who accepted to participate were offered two additional points for a university exam, but only if they could convince the interviewer that they were telling the truth. Eventually, all participants received the two points. They were also informed that the experiment lasted approximately 35-40 minutes. Upon arrival, the participants completed an informed consent form. Then, the experimenter told them that they were going to perform a set of tasks and that after completion of these tasks they would be interviewed. They were also informed that the interviews would be video-recorded. The set of tasks included two sub-sets of tasks. For the first sub-set of tasks, the participants received an envelope from the experimenter, which contained the password needed to log onto a laptop. The experimenter then left the room. Once logged onto the laptop the participants had to read the only word document file that was available on the PC desktop. This word document instructed participants to search for a backpack in the room, which contained a CD-ROM and to watch the video file from the CD-ROM. They could play, pause, stop and view the video as many times as they needed. On the video a man appeared in view who informed the participants that they had to search for a key inside the same backpack where they found the CD-ROM. That key could open a safe-deposit box that was placed near the window. The man on the video also informed the participants that they would find further instructions inside the safe-deposit box. These instructions informed participants that they had to access the Internet, send an email to a specific email address and to exit the room and wait for a person (a confederate).

The meeting with the confederate served as a landmark. Everything that happened before the meeting was part of the comparable truth baseline event, whereas everything that happened after the meeting was part of the target event. However, only participants in the comparable truth condition had to report what they had done *before* the meeting (in addition to what they had done after the meeting).

Regarding the second sub-set of actions, the confederate gave the participants a newspaper, which contained further instructions and left the room. The instructions informed the participants that they had to enter the room adjacent to where they

performed the first set of tasks and to look for a coat hook on which an USB stick was located. Participants had to take the USB stick and to look for a book inside the wardrobe in the room in which they would find another USB stick. They had to switch the two USB sticks and to keep the new one for the remaining part of the experiment. They were also asked to leave the newspaper near the book inside the wardrobe. The participants were then asked to return to the first room and to wait for the interviewer.

A 56 years old male psychology student, who was blind to both Baseline and Veracity experimental conditions, acted as the interviewer. When he arrived, he introduced himself and asked an open-ended question regarding either the first sub-set of tasks (comparable truth baseline condition) or the last year of study/work (small talk baseline condition). Then the interviewer left the room and the confederate returned to give instructions about the veracity condition: Participants allocated to the truthful condition were told to continue to answer the questions truthfully. Participants allocated to the lying condition were instructed to lie to the questions that would be asked regarding the second sub-set of tasks. They were instructed to create a detailed and credible story, which included at least 4-5 actions or events. When liars confirmed that they understood the instructions, the confederate left the room and told them that they had five minutes to prepare the lie. Truth tellers were also given five minutes to prepare themselves. After five minutes the interviewer came back into the room and asked a second open-ended question regarding what had happened after the meeting. After the interview participants completed a questionnaire investigating background characteristics and their motivation to do well in the interview (on a 5-point Likert scale ranging from [1] low to [5] high). Finally, participants were debriefed.

Coding

A coder blind to the hypotheses and the experimental conditions coded the interviews for the frequency per minute of hands and finger movements (movements of hands and fingers while arms are held still), one arm movements (movements of one arm with the other remaining still) and two arms movements (movements of both arms at the same time). The coder further coded the frequency for spatial (information about place and/or spatial arrangement of objects and people), temporal (information about when an event happens and or the description of a sequence of events), visual

(information about what the interviewee saw: “I saw a man entering the room”), audio (information about what the interviewee heard: “I heard a men shouting”) and action details (“I wrote an email). A second coder, again blind to hypotheses and experimental conditions, coded 18 interviews (25%) for the same verbal and nonverbal behaviours to evaluate the inter-rater agreement (Interclass Correlation Coefficient). The lowest level of agreement (ICC = .75) was obtained for the variable *Visual details*, whereas the highest level of agreement was obtained for the variable *Actions* and was of (ICC = .96). Therefore, the scores indicated high inter-rater agreement.

Similarity scores computation

The similarity scores were computed as follows: for each dependent variable, we divided the lowest score by the highest score, regardless of the phase of the interview. This means that if the lowest score for a variable appeared in the baseline, we divided the score of that variable in the baseline by the score of the same variable in the target phase. The same was true for the opposite situation. Then, we multiplied the result by 100. The highest the score, the more similar participants behaved between the two phases of the interview.

Results

To assess whether the level of motivation differed between conditions, an ANOVA was conducted with Baseline (comparable truth vs. small talk) and Veracity (truth tellers vs. liars) as factors and the level of motivation as the dependent variable. A main effect for Baseline appeared, $F(1, 65) = 8.40$, $p < .01$, Cohen's $d = .69$, with participants in the comparable truth condition showing higher level of motivation ($M = 4.46$, $SD = .56$) than those in the small talk condition ($M = 3.97$, $SD = .82$). The main effect for Veracity, $F(1, 65) = .80$, $p = .78$ and the Baseline X Veracity interaction effect, $F(1, 65) = .01$, $p = .91$ were not significant.

To test our hypotheses, first, a MANOVA for nonverbal cues was conducted with Baseline (comparable truth vs. small talk) and Veracity (truth tellers vs. liars) as the between-subjects factors and the similarity scores for hands and fingers, one arm and two arm movements as dependent variables. At a multivariate level, the main effect

for Baseline, $F(3, 63) = 6.15$, $p = .001$ was significant. The main effect for Veracity, $F(3, 63) = .22$, $p = .88$, and the Baseline X Veracity interaction, $F(3, 63) = 2.22$, $p = .09$ were not significant.

At a univariate level, the main effect for Baseline was significant for one arm movements only, $F(1, 65) = 18.27$, $p < .001$. Participants in the comparable truth baseline ($M = 68.16$, $SD = 26.62$, C. I. [59.29; 77.04]) showed more similarity than participants in the small talk baseline ($M = 38.45$, $SD = 30.72$, C. I. [27.37; 49.53]). On the other hand, participants in the comparable truth baseline obtained similar similarity scores for hands and finger movements ($M = 71.91$, $SD = 23.40$, C. I. [64.11; 79.71]) to participants in the small talk baseline ($M = 64.78$, $SD = 28.10$, C. I. [54.65; 74.91]), $F(1, 65) = 1.15$, $p = .29$. Likewise, those in the comparable truth baseline ($M = 53.78$, $SD = 37.00$, C. I. [41.45; 66.12]) did not differ from those in the small talk baseline ($M = 60.06$, $SD = 36.55$, C. I. [46.89; 73.24]) for what concerns two hands movements, $F(1, 65) = .48$, $p = .49$. Results, therefore, did not support Hypothesis 1 for nonverbal cues, as the interactions effect was not significant. However, the fact that participants in the comparable truth baseline obtained higher similarity scores for one arm movements than those in the small talk baseline, supports the idea that interviewees interviewed via a small talk baseline tend to change their behaviour more than those interviewed via a comparable truth baseline.

Second, a MANCOVA for verbal cues was conducted with Baseline (comparable truth vs. small talk) and Veracity (truth tellers vs. liars) as the between-subjects factors and the similarity scores for spatial, temporal, visual, audio and action details as the dependent variables. Additionally, since these details may vary with the overall number of details within each of the interview phases (Baseline and Target), the total number of details provided in the Baseline and the Target phases were introduced as covariates.

At a multivariate level, the main effect for Baseline, $F(5, 59) = 12.02$, $p < .001$, and the Baseline X Veracity interaction, $F(5, 59) = 2.52$, $p = .04$, were significant. The Veracity main effect was not significant, $F(5, 59) = .57$, $p = .72$.

At a univariate level, there was a main effect for Baseline for spatial, $F(1, 63) = 11.35$, $p = .001$, and visual, $F(1, 63) = 51.03$, $p < .001$ similarity scores. The only significant

Baseline X Veracity effect appeared for similarity scores for spatial details, $F(1, 63) = 9.24$, $p = .003$. Since the interaction effect is the most informative effect and considering that our hypotheses are related to it, only the interaction effect will be discussed. We were particularly interested in differences between truth tellers and liars in the two baseline conditions. To be as informative as possible, we report the data for all dependent variables.

Comparable truth condition

A MANOVA was conducted on the comparable truth condition data with Veracity (truth tellers vs. liars) as factor. The dependent variables were the similarity scores for hands and fingers, one arm and two arms movements. At a multivariate level, the Veracity effect was not significant, $F(3, 33) = .94$, $p = .43$. At a univariate level, no significant effect appeared (Table 1).

A MANCOVA was conducted on the same data with Veracity as factor, spatial, temporal, visual, audio, and action similarity scores as the dependent variables and the total amount of details in the two interview phases as the covariates. At a multivariate level, the veracity effect was not significant, $F(5, 29) = 1.74$, $p = .15$. At a univariate level, a significant effect for spatial detail similarity scores appeared. Truth tellers ($M = 78.67$, $SD = 15.60$, C. I. [71.14; 86.18]) showed more similarity than liars ($M = 52.25$, $SD = 26.75$, C. I. [38.95; 65.55]), $F(1, 33) = 7.44$, $p = .01$, Cohen's $d = 1.20$.

All univariate effects and their effect sizes are reported in Table 1. Three out of eight variables showed similarity scores that went in the 'wrong' direction (less similarity amongst truth tellers), but for those five who went in the predicted direction the strongest effect size was for spatial details, $d = 1.20$, followed by hands and finger movements, $d = .54$. The effect sizes for the remaining three variables were small (d-scores ranged between .10 and .25). This means that Hypothesis 2, the effect sizes would be stronger for verbal behaviours than for nonverbal behaviours, only obtained very limited support.

Enter Table 1 about here

To examine whether the similarity scores differed from perfect similarity (100) one-sample t-tests were conducted for each dependent variable and were tested against a 100 score. If the comparable truth works, effects should be significant for liars only. The results are reported in Table 2, which shows that all effects were significant, for both truth tellers and liars. The high *d*-scores imply that the responses from both liars and truth tellers in the interview differed considerably from their responses in the comparable truth baseline.

Enter Table 2 about here

Small talk condition

A MANOVA was conducted on the small talk condition data with Veracity (truth tellers vs. liars) as factor. The dependent variables were the similarity scores for hands and fingers, one arm and two arms movements. The multivariate effect for Veracity was not significant, $F(3, 28) = 1.43$, $p = .25$. None of the univariate effects was significant either (Table 3).

A MANCOVA was conducted with the same data again with Veracity (truth tellers vs. liars) as factor and spatial, temporal, visual, audio, and action similarity scores as dependent variables. The total number of details provided in the Baseline and the Target phases were set as covariates. At a multivariate level, the Veracity effect was not significant, $F(5, 24) = .57$, $p = .72$. No significant univariate effect appeared neither (Table 3).

In sum, the multivariate effects were not significant in the comparable truth nor in the small talk conditions. In the comparable truth condition truth tellers displayed greater similarity on one variable compared to no significant difference at all in the small talk condition. This means that the comparable truth baseline was more effective than the small talk baseline, supporting Hypothesis 1, but even in the comparable truth condition the findings were weak.

Similar to the comparable truth condition analyses, a one-sample t-test was conducted for each dependent variable and were tested against a 100 score. If a small talk baseline is ineffective, all results should be significant. Table 2 shows that this was indeed the case.

Enter Table 3 about here

Discussion

In this experiment, we tested the efficacy of two different types of baselines for deception detection using a similarity rating. We found that similarity scores did not differ between truth tellers and liars for participants in the small talk condition. This result is similar to that obtained by Ewens and colleagues (2014), who found that both truth tellers and liars showed differences in the amount of hard thinking and behavioural control between two phases of the interview. Our results therefore reject the efficacy of a small talk baseline procedure as proposed by Frank and colleagues (Frank *et al*., 2006) and by Inbau and colleagues (Inbau *et al*., 2013) as in such a procedure truth tellers and liars would appear equally deceptive.

We found a significant difference between truth tellers and liars in the comparable truth condition, with truth tellers showing more similarity than liars. However, it is too premature to conclude that a comparable truth baseline works for at least two reasons. First, the method only worked for one of the variables examined in the experiment: spatial details. It is possible that this result is due to the type of task. Indeed, the set of actions participants discussed in the comparable baseline phase and actual interview phase showed similarity in spatial information (moving around two rooms and to interact with several objects placed in several places). Truth tellers showed more spatial similarity in discussing these two sets of spatial activities than liars, who told the truth about one set of spatial activities but lied about the other set of activities. Thus, to create an effective baseline the type of activities discussed in the baseline and target interview need to be comparable. Second, even for the spatial detail variable the findings in the comparable truth condition were problematic because, although truth tellers showed more similarity than liars, they did not show perfect similarity. This means that in real life, judging whether or not a difference in spatial similarity between baseline and target interview occurred will not work to detect deceit. Somehow, the size of dissimilarity needs to be taken into account, which will be difficult to do.

Furthermore, we employed a procedure whereby the veracity instructions were given after the baseline and before the target phase of the interview. We opted for this

strategy as there was the risk of liars not following the instruction to be as truthful as possible during the baseline phase. That is, informing them about their veracity condition could have influenced how they answered to the baseline questions (Deeb *et al*., 2017; Strömwall & Willén, 2011). Had this been the case, we would have not had a real and truthful baseline. Perhaps participants in the present study thought about this strategy when they were given time to prepare themselves for the target phase of the interview. That is, once the interviewee was given the instruction to lie after the baseline phase, s/he may have employed consciously the strategy to try to provide responses in the target question that were as similar as possible to those given in the baseline question. Future studies could explore how the comparable truth baseline technique works when the veracity instruction are given before the baseline part of the interview.

Further research could also explore the baseline technique when no preparation time is given between the two phases. Although our results are interesting there is the risk that our interview was perceived as two separated interviews rather than one. We do not think this was the case, the time lapse we provided was enough to prepare but in our opinion not too long to make the interview look like two different interviews, but there is a small risk that at least some participants considered this to be two interviews.

Limitations

The stakes for this experiment were relatively low. However, as Ewens *et al*. (2014) noted, higher stakes are unlikely to make a small talk baseline more effective. Rather, this should further decrease the efficacy of a small talk baseline procedure as differences in stakes between baseline and investigative questions would become even more pronounced. Raising the stakes should not be an issue for the comparable truth baseline approach as one of the assumptions is to keep the stakes of the two phases constant so that they remain comparable.

Additionally, our sample size was small. Therefore, although we found some interesting results, further research should explore how the similarity scoring system performs when bigger samples are used.

Conclusion

Despite being promoted in the academic and professional literature, a baseline lie detection technique using a small talk baseline is not effective in discriminating truth tellers from liars as similar differences arise for both Veracity groups. Practitioners should stay away from using such a baseline procedure. The alternative baseline procedure, the comparable baseline procedure, was more successful than the small talk baseline procedure but still not good enough to be implemented in real life. There are two options. First, drop the baselining technique altogether. Given the paucity of support for this technique, nothing would be lost. Second, because baselining is popular amongst practitioners, researchers could spend more time designing an appropriate baseline method. This article revealed some difficulties researchers will face when designing an appropriate method.

References

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

Buller, D. B., & Burgoon, J. K. (1996). Interpersonal deception theory. *Communication Theory, 6,* 203-242. DOI: 10.1111/j.1468-2885.1996.tb00127.x

Burgoon, J. K., Buller, D. B.,White, C. H., Afifi,W., & Buslig, A. L. (1999). The role of conversational involvement in deceptive interpersonal interactions. *Personality and Social Psychology Bulletin, 25,* 669-686. DOI: 10.1177/0146167299025006003

Colwell, K., Hiscock-Anisman, C. K., Memon, A., Colwell, L. H., Taylor, L., & Woods, D. (2009). Training in Assessment Criteria Indicative of Deception (ACID) to improve credibility assessment. *Journal of Forensic Psychology Practice, 9,* 199-207.

Colwell, K., Hiscock-Anisman, C. K., Memon, A., Taylor, L., & Prewett, J. (2007). Assessment Criteria Indicative of Deception (ACID): An integrated system of investigative interviewing and detecting deception. *Journal of Investigative Psychology and Offender Profiling, 4*(3), 167-180. DOI: 10.1002/jip.73

Davis, M., & Hadiks, D. (1995). Demeanor and credibility. *Semiotica, 106,* 5–54. DOI: 10.1515/ semi.1995.106

Deeb, H., Vrij, A., Hope, L., Mann, S., Granhag, P. A., & Lancaster, G. L. (2017). Suspects' consistency in statements concerning two events when different question formats are used. *Journal of Investigative Psychology and Offender Profiling*, *14*(1), 74-87.

DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological Bulletin*, *129*(1), 74-118.

Ewens, S., Vrij, A., Jang, M., & Jo, E. (2014). Drop the small talk when establishing baseline behaviour in interviews. *Journal of Investigative Psychology and Offender Profiling*. DOI: 0.1002/jip.1414

Frank, M. G., Yarbrough, J. D., & Ekman, P. (2006). Investigative interviewing and the detection of deception. In T. Williamson (Eds.), *Investigative interviewing: Rights, research and regulation* (pp. 229–255). Cullompton, Devon: Willan.

Granhag, P. A., & Hartwig, M. (2008). A new theoretical perspective on deception detection: On the psychology of instrumental mind reading. *Psychology, Crime & Law, 14*,(3), 189-200.

Granhag, P. A., & Hartwig, M. (2015). The Strategic Use of Evidence technique: A conceptual overview. In P. A. Granhag, A. Vrij, & B. Verschuere (Ed), *Detecting deception: Current challenges and cognitive approaches* (pp. 231-252). Chichester: John Wiley and Sons.

Granhag, P. A., & Strömwall, L. A. (2002). Repeated interrogations: Verbal and non-verbal cues to deception. *Applied Cognitive Psychology, 16*, 243-257. DOI: 10.1002/acp.784

Hartwig, M., & Bond Jr., C. F. (2011). Why do lie-catchers fail? A lens model meta-analysis of human lie judgments. *Psychological Bulletin, 137*(4), 643-659. DOI: http://dx.doi.org/10.1037/a0023589

Hartwig, M., & Bond Jr,. C. F. (2014). Lie Detection from Multiple Cues: A Meta-analysis. *Applied Cognitive Psychology, 28,* 661–676. DOI: 10.1002/acp.3052

Inbau, F. E., Reid, J. E., Buckley, J. P., & Jayne, B. C. (2013). *Criminal interrogation and confessions,* 5th edition. Burlington, MA: Jones & Bartlett Learning.

Ioannou, M., & Hammond, L. (2015). The detection of deception within investigative contexts: Key challenges and core issues. *Journal of Investigative Psychology and Offender Profiling, 12*(2), 107-118. DOI: 10.1002/jip.1433

Kleinke, C. L. (1986). Gaze and eye contact: a research review. *Psychological bulletin, 100,* 78. DOI: 10.1037/0033-2909.100.1.78

Matarazzo, J. D., Wiens, A. N., Jackson, R. H., & Manaugh, T. S. (1970). Interviewee speech behavior under conditions of endogenously-present and exogenously-induced motivational states. *Journal of Clinical Psychology, 54,* 15-26. DOI: 10.1037/h0028651

Moston, S., & Engelberg, T. (1993). Police questioning techniques in tape-recorded interviews with criminal suspects. *Policing and Society: An International Journal of Research and Policy, 3*, 223-237. DOI: 10.1080/10439463.1993.9964670

Stiff, J., Corman, S., Krizek, B., & Snider, E. (1994). Individual differences and changes in nonverbal behavior unmasking the changing faces of deception. *Communication Research, 21,* 555-581. DOI: 10.1177/009365094021005001

Strömwall, L. A., & Willén, R. M. (2011). Inside criminal minds: Offenders' strategies when lying. *Journal of Investigative Psychology and Offender Profiling*, *8*(3), 271-281.

Vrij, A. (2006). Challenging interviewees during interviews: The potential effect on lie detection. *Psychology, Crime & Law, 12,* 193-206.

Vrij, A. (2008). *Detecting lies and deceit: Pitfalls and opportunities*. Chichester: John Wiley and Sons.

Vrij, A. (2014). Interviewing to detect deception. *European Psychologist, 19,* 184-194. DOI: 10.1027/1016-9040/a000201

Vrij, A. (2015). A cognitive approach to lie detection. In P. A. Granhag, A. Vrij, & B. Verschuere (Ed), *Detecting deception: Current challenges and cognitive approaches* (pp. 205-230). Chichester: John Wiley and Sons.

Vrij, A., Fisher, R. P., Blank, H., Leal, S., & Mann, S. (2016). A cognitive approach to elicit verbal and nonverbal cues to deceit. In J-W. van Prooijen, & P. A. M. van Lange (Eds.), *Cheating, corruption, and concealment: the roots of dishonesty*. Cambridge: Cambridge University Press.

Vrij, A., & Granhag, P. A. (2012). Eliciting cues to deception and truth: What matters are the questions asked. *Journal of Applied Research in Memory and Cognition, 1*(2), 110-117. DOI: 10.1016/j.jarmac.2012.02.004

Vrij, A., & Winkel, F. W. (1991). Cultural patterns in Dutch and Surinam nonverbal behavior: An analysis of simulated police/citizen encounters. *Journal of Nonverbal Behavior, 15,* 169–184. DOI: 10.1007/BF01672219

Table 1. *Similarity scores as a function of Veracity in the comparable truth condition*

	Truth Tellers		Liars				
Similarity scores	*M*	*SD*	*M*	*SD*	*F*	*p*	*d*
Hands and fingers movements	77.95	23.53	65.54	22.12	2.72	.10	.54
One arm movements	69.52	27.82	66.73	26.02	.09	.75	.10
Two arms movements	52.87	39.42	54.74	35.38	.02	.88	- .05
Spatial details	78.67	15.60	52.25	26.75	7.44	.01	1.20
Temporal details	43.51	33.14	55.79	30.18	1.09	.30	- .38
Visual details	70.56	17.95	65.53	21.12	.53	.47	.25
Audio details	14.47	32.61	30.46	39.78	.51	.48	- .44
Action details	66.09	18.21	61.96	22.89	.00	.98	.20

Table 2. *Statistics examining whether the similarity scores differ from perfect similarity*

	Comparable Truth				Small Talk			
	Truth Tellers		Liars		Truth Tellers		Liars	
	t	*d*	*t*	*d*	*t*	*d*	*t*	*d*
Spatial details	-5.961	-1.40	-7.574	-1.83	-11.425	-2.85	-6.285	-1.68
Temporal details	-7.429	-1.75	-6.213	-1.50	-6.309	-1.57	-6.081	-1.62
Visual details	-7.147	-1.68	-6.922	-1.69	-28.682	-7.17	-12.890	-3.44
Audio details	-11.431	-2.69	-7.415	-1.80	-4.781	-1.19	-5.909	-1.58
Action details	-8.115	-1.91	-7.048	-1.71	-10.115	-2.53	-6.501	-1.74
Hands and fingers movements	-4.084	-.96	-6.609	-1.60	-6.265	-1.56	-4.027	-1.11
One arm movements	-4.775	-1.12	-5.424	-1.31	-7.497	-1.87	-10.607	-2.94
Two arms movements	-5.210	-1.23	-5.427	-1.31	-4.368	-1.09	-4.436	-1.23

Table 3. *Similarity scores as a function of Veracity in the small talk condition*

	Truth Tellers		Liars				
Similarity scores	*M*	*SD*	*M*	*SD*	*F*	*p*	*d*
Hands and fingers movements	57.94	27.68	70.58	27.33	2.23	.14	-.46
One arm movements	41.52	32.16	30.33	24.57	.35	.55	.39
Two arms movements	59.77	37.98	57.58	35.78	.00	.96	.06
Spatial details	14.69	30.79	33.66	42.93	.09	.76	-.50
Temporal details	40.98	30.57	54.60	30.23	1.518	.23	-.45
Visual details	10.36	12.88	13.20	23.68	.21	.65	-.14
Audio details	41.17	50.73	33.93	45.58	.42	.52	.15
Action details	45.50	22.21	55.45	27.87	.78	.39	-.39

Appendix 4: Caso, L., Palena, N., Vrij, A., & Gnisci, A. (2019). Observers' performance at evaluating truthfulness when provided with Comparable Truth or Small Talk Baselines. *Psychiatry, Psychology and Law*. doi:10.1080/13218719.2018.1553471. [PA3]

Abstract

Research has shown that the Comparable Truth Baseline Technique outperforms the Small Talk with respect to the elicitation of cues to deception. However, their impact on observers' accuracy has not been evaluated yet. In this experiment, participants (*N* = 74) watched ten interviews where senders either lied or told the truth about a set of tasks. Half of the interviews were conducted with a Comparable Truth Baseline, the other half with a Small Talk Baseline. As predicted, results showed that observers in the Comparable Truth Baseline condition outperformed the participants in the Small Talk Baseline condition in terms of total accuracy rates. The paper sheds light on the impact of the two baseline techniques in distinguishing truth tellers from liars and discourages the use of a small talk baseline. It also provides insights for future studies.

Keywords: veracity assessment, baseline technique, observers' accuracy, comparable truth, small talk, deception detection, interviewing techniques

Observers' performance at evaluating truthfulness when provided with Comparable Truth or Small Talk Baselines

Introduction

Research has shown that people's accuracy in assessing truthfulness and deceit is low (Bond & DePaulo, 2006; Vrij, 2008). Bond and DePaulo (2006) meta-analysis including almost 25,000 observers found a 54% accuracy rate, whereby 50% could be achieved by chance. In addition, accuracy of deception judgments are unrelated to confidence in one's own judgements (DePaulo, Charlton, Cooper, Lindsay, & Muhlenbruck, 1997), suggesting that people have no insight into their own competence. Additionally, people from whom one should expect a higher accuracy, such as police officers, are not more accurate than laypeople (Bond & DePaulo, 2006).

To explain the low accuracy rates, it has been argued that cues to deception are faint and unreliable (Vrij, 2008). DePaulo and colleagues (2003), in one of the most comprehensive meta-analyses on cues to deception, have found that effect sizes for behavioural and verbal differences between truth and lie telling are small, with an average effect size of Cohen's d = .25 for the most diagnostic cues (Vrij & Granhag, 2012). Consequently, it should not be surprising that observers' accuracy is low, as having weak cues to rely on when making judgments makes the judgment itself difficult (Hartwig & Bond, 2011). It is for this reason that academics are now focusing on interviewing techniques aiming at enhancing differences between truth and lie telling (Granhag & Vrij, 2010; Vrij, 2014; Vrij & Granhag, 2012). The rationale behind this is that if truth tellers and liars do not differ much, it is desirable to find a way to make such differences more evident. Academics have explored several approaches that indeed augmented differences between truth tellers and liars, including the Strategic Use of Evidence (Granhag & Hartwig, 2015) and cognitive credibility assessment (Vrij, 2015; Vrij, Fisher, & Blank, 2017).

There is an alternative interviewing strategy often proposed within police forces: The baseline technique (Ewens, Vrij, Jang, & Jo, 2014; Vrij, 2016). The underlying idea is that since there are interpersonal differences in behaviour, an interviewer should start by obtaining a behavioural baseline of the interviewee through observing the responses in chitchat while answering trivial questions. Then, while the interview

progresses, the interviewer should compare this baseline behaviour with interviewees' responses to target questions (questions related to the topic under investigation). If any difference arises, then one may conclude that the interviewee is lying (Frank, Yarbrough, & Ekman, 2006).

The use of the baseline technique in this manner is problematic. People's nonverbal behaviour is related to stakes and topic of conversation (Vrij, 2008), and in terms of stakes and topic of conversation the baseline response and the target response are not comparable to each other. The result is that both liars and truth tellers are likely to change their behaviour when baseline and target responses are compared (Caso, Maricchiolo, Bonaiuto, Vrij, & Mann, 2006; Gnisci, Caso e Vrij, 2010; Moston & Engelberg, 1993). Ewens, Vrij, Jang and Jo (2014) empirically tested the efficacy of this type of baselining, which they labelled "Small Talk". When interviewing their participants, they started by asking an initial small talk baseline question, followed by several target questions. It was found that both truth tellers and liars changed their behaviour during the interview. Therefore, they concluded that the small talk baseline in not an effective strategy to detect deception.

Ewens and colleagues (2014) also noted that there is a different type of baseline which may work better, the comparable truth baseline. They described it as follows: "*Comparable means that the baseline the investigator uses must be similar in content, context, stakes, and cognitive and emotional involvement to investigative questions*" (Vrij, 2016, p. 1114). Vrij and Mann (2001) gave a real life example of this type of baseline. They compared several behaviours displayed by a suspect in a murder case in different phases of the police interrogation. The suspects was interrogated about his activities during the day of the murder and provided a detailed answer covering the entire day. Police investigated his whereabouts and could only verify his morning activities. Eventually, it became known that he had met the victim in the afternoon and killed her later. Vrij and Mann (2001) analysed the videotaped interrogation and found that the suspect did show a difference in his behaviour when he discussed his activities in the morning (truth) compared to the afternoon and evening (lie). Their results therefore gave credit to the potential effectiveness of the comparable truth baseline.

There is to date only one experimental study in which comparable truth and small talk baselines were compared (Palena, Caso, Vrij, & Orthey, 2018). In the comparable truth baseline condition, the baseline question referred to three tasks that the participants performed. All participants had to answer the baseline question truthfully. Then, in the target phase of the interview, half of the participants told the truth and the other half lied about three additional, but similar, tasks. In the small talk baseline condition, the baseline question referred to personal information about the interviewee (e.g.: the last year spent as a student). Again, all participants had to answer the baseline question truthfully. The target questions referred to the same three additional tasks as for participants in the comparable truth baseline condition. Again, half of the participants responded to this target question truthfully, whereas the other half lied. The authors compared the baseline and target responses in terms of similarity for both nonverbal and verbal behaviour. Truth tellers' levels of similarity did not differ from those of liars in the small talk baseline condition, but the truth tellers' levels of similarity were higher than those of liars in the comparable truth baseline condition, but only for spatial details. This study thus replicated Ewens et al.'s (2014) study that a small talk baseline does not work. Second, it showed that nonverbal baselining is problematic, but that verbal baselining may be more effective.

However, verbal baselining is not straightforward either as speech content is affected by the topic of conversation (Vrij, 2008). For example, in Ewens' et al (2014) study, the baseline concerned the informed consent form, whereas the target phase concerned the actual/pretended job. Consequently, the speech content of the two phases is expected to be different regardless of veracity, as the interviewee talks about two different topics.

Palena and colleagues (2018) did not test whether the differences in speech content would be clear to lay observers. Bond and DePaulo's meta-analysis (2006) underlined that when observers were previously exposed to senders' truthful baselines, their accuracy improved. However, such previous exposure cannot be considered a baseline obtained through strategic questioning. Rather, since observers had simply the opportunity to become familiar with the sender, it was possibly the result of "*baseline familiarity*" (Bond & DePaulo, 2006; Brandt, Miller, & Hocking, 1980, 1982). Feeley, deTurck and Young (1995) provided their participants with zero,

one, two or four truthful baseline exposures, and found that there was a positive linear relationship between the amount of familiarity with the sender and observers' accuracy. Results about the positive effects of baseline familiarity are important in personal and intimate relationships but baseline familiarity may be difficult to achieve during police investigations, as suspects and investigators are often strangers to each other and there is often the need to interview the suspect as soon as possible. Hence, it is more likely that police officers will try to obtain baselines with specific questioning rather than through increased familiarity.

Based on the literature presented above and expecting better efficacy of the comparable truth over the small talk baseline, we made the following predictions.

H1: Observers in the comparable truth baseline condition will reach higher overall accuracy rates than those in the small talk baseline condition.

H2: Observers in the comparable truth baseline condition will also reach higher accuracy rates for truth (Hypothesis 2a) and lie detection (Hypothesis 2b).

Material and methods

Participants

Seventy-four participants (56 women and 18 men), between 19 and 51 years of age (M = 25.67; SD = 6.14), took part in the experiment. One of them was excluded from the analyses as the person did not followed the instructions. This left us with a total sample of seventy-three participants. Thirty-seven participants were allocated to the comparable truth condition, whereas the remaining thirty-six were allocated to the small talk condition.

Design

The experiment was based on a two-group comparison. The type of baseline (Comparable Truth vs. Small Talk) was the between-subjects factor. Total, truth and lie accuracy rates, and d-prime and β values were the dependent variables.

Stimulus material

Twenty video stimuli were used for this experiment. All senders portrayed in the videos performed a mission that consisted of a series of tasks. The mission started with the participant receiving an envelop from the experimenter, which contained a PC password. Then, the experimenter left the room. As soon as the participant logged onto the PC only one file, a word document named "Read Me", appeared on the desktop. Such file informed the participant that s/he had now to look for CD-ROM that was placed inside the only backpack available in the room. The word document ended informing the participant to watch the video file that was recorded in such CD-ROM. This showed a mean telling the participant to look again into the backpack to search for a key. The man also explained that the key served to open a safe-deposit box that was in the room. Once the participant opened the box, s/he found additional instruction that asked them to send an e-mail to a specific address and to exit the room and wait for a person to come. This person was actually a confederate and the meeting served to split the mission into two subsets. Everything that happened before the meeting was the comparable truth baseline section. Everything that happened after the meeting was the target section of the interview. The confederate gave a newspaper to the participant and informed him/her that it contained further instruction. Then, the confederate left the room. The instruction guided the participant to a room adjacent to the first one and informed them that they had to look for a pen-drive, which was attached on a coat hook. Once found, the participant had to take it and switch it with another pen-drive, which was hidden inside a book at the bottom of the wardrobe located in the room. The instruction also told the participant to leave the newspaper with the instruction near that book and to keep the first pen-drive with them for the rest of the experiment. The participant then came back to the first room waited for the interviewer. Of the twenty videos, ten were used for the comparable truth condition. Here, for the baseline questioning, the participants reported everything that happened before the meeting with the confederate. For the target questioning, the participant reported everything that happened after the meeting with the confederate. The remaining ten videos were used for the small talk condition. Here, the baseline questioning consisted of participants providing personal information (e.g.: describe their last year as a worker), whereas the target questioning was identical to that for participants in the comparable truth condition. That is, also participants in the small

talk condition had to report everything that happened after the meeting with the confederate for the target questioning. All senders were honest in the baseline phase, whereas half of them were instructed to tell the truth and the other half to lie when responding to the target questioning.

Each video sequence was produced as follows: a black screen with a white text indicating "Baseline" appeared and lasted for three seconds. Then, the first sender appeared on the screen and started answering to the baseline question. Once s/he finished, a second black screen with the text "Target" appeared, also lasting three seconds. Then the first sender started answering the target question. Once s/he finished answering, another black screen appeared, and lasted for 30 seconds. In this time-window, participants had to make their veracity decisions. When the 30 seconds expired, a high frequency sound warned the participant that the time to evaluate the first sender was over and that the second sender was going to appear on the screen. This sequence was repeated until the end of the 10 videos. Participants expressed their veracity decisions answering the following question: "*Do you believe the interviewee was…"* . The answer alternatives were "*Telling the truth"* and *"Lying"*. Senders' veracity status was counterbalanced, and each ten targets tape consisted of five truth tellers and five liars.

Procedure

Upon the arrival, participants met the experimenter and were briefed about the aim of the study. They were informed that they were going to watch some interviews that were divided into two sections: A baseline section and a target section. Observers where informed that the senders were always honest in the baseline section whereas they may have been either telling the truth or lying in the target section. Observers were asked to decide whether each sender was telling the truth or lying in the target section. They were told that they needed to pay attention to (non)verbal deviations from the baseline to make their decision. No training was offered. Assignment to the two baseline conditions was random. The participants were also informed that there were more answer sheets (twenty) than the actual number of stimuli and also that truth and lie telling may be balanced in different amounts. Such instructions were given to prevent participants making decisions based on balancing expectations rather than lie

detection task decisions. The participant also read and signed a consent form and were offered an additional point for a university exam if they performed well. Eventually, all participants received the point, regardless. Once they confirmed they understood the instructions, they were left alone in a room with a computer and the answer sheets. They were also instructed to exit the room once they completed their task. They then started watching the stimuli. Each participant was shown the tapes individually.

Results

A *t*-test with type of Baseline as factor and overall accuracy rates as the dependent variable revealed that participants in the comparable truth condition were more accurate [*M* = 56.49; *SD* = 18.74, 95% CI (50.24, 62.73)] than those in the small talk condition [*M* = 47.41; *SD* = 18.62, 95% CI (41.11, 53.71)], $t(71) = 2.076$, $p = .042$, Cohen's $d = .49$, supporting Hypothesis 1. One *t*-test, again with type of Baseline as factor, showed no difference between conditions for truth accuracy, $t(71) = 1.200$, $p = .22$, Cohen's $d = .28$ but showed a difference for lie accuracy, $t(71) = 1.990$, $p = .050$, Cohen's $d = .47$, with the Comparable Truth baseline condition resulting in a higher lie accuracy rate than the Small Talk baseline condition (Table 1). Therefore, Hypothesis 2a was not supported but Hypothesis 2b received support.

Enter Table 1 about here

Further analyses (see Table 2) showed that in the *comparable truth baseline* condition the total accuracy rate was significantly above chance. No other accuracy rate differed from chance and only lie accuracy rate differed from the accuracy rates found in Bond and DePaulo's (2006) meta-analysis. In the *small talk baseline* condition, none of the accuracy rates differed from chance and the total and truth accuracy rates were significantly lower than those found in Bond and DePaulo's (2006) meta-analysis.

Enter Table 2 about here

Signal Detection Analyses

It has been suggested to use Signal Detection Theory to analyse the accuracy of deception judgments in more detail (Jupe, Akehurst, Vernham, & Allen, 2016; Meissner & Kassin, 2002). Therefore, participants' performance was assessed via

discrimination accuracy, using *d*-prime values, and responding bias, using β values. The former is a measure of sensitivity expressed in standard deviations units (Stanislaw & Todorov, 1999). Values of 0 indicate an inability to distinguish between the signal and noise, in our case liars from truth tellers. Values greater than 0 indicate that the observers are indeed able to distinguish truthful from lying senders.

The β value is a measure of response bias (Stanislaw & Todorov, 1999) whereby values of 1 indicate no response bias. On the other hand, values greater than 1 indicate a truth bias and values below 1indicate a lie bias.

A first *t*-test with baseline condition as factor and *d*-prime values as the dependent variable showed that observers in the comparable truth baseline [*M* = .35; *SD* = 1.00, 95% CI (.02, .68)] were better than those in the small talk baseline [*M* = -.13; *SD* = .99, 95% CI (-.47, .20)] at discriminating truth tellers from liars, $t(71) = 2.079$, $p = .040$, Cohen's $d = .49$. This, again, supports Hypothesis 1.

A between-subjects *t*-test on β values showed that participants in the comparable truth condition [*M* = 1.27; *SD* = .44, 95% CI (1.12, 1.42)] did not differ from those in the small talk condition [*M* = 1.21; *SD* = .48, 95% CI (1.04, 1.37)] in their response bias, $t(71) = .59$, $p = .55$, Cohen's $d = .14$. Additionally, β values for participants in both baseline conditions were significantly greater than 1, indicating that all participants were truth biased, regardless of the type of Baseline exposure (Table 3).

Enter Table 3 about here

Discussion

In this experiment, we found that observers in the comparable truth condition were more accurate in distinguishing truth tellers from liars than observers in the small talk condition. In addition, observers in the comparable truth condition performed significantly better than chance levels whereas observers in the small talk condition did not. This shows the benefit of using a comparable truth baseline compared to a small talk baseline. Our results discourage the use of a small talk baseline technique, the technique used by practitioners (Inbau, Reid, Buckley, & Jayne, 2013) and advocated by some academics (Frank et al., 2006).

It should be noted that, although better than the small talk baseline, the comparable truth baseline is still not ready to be implemented in real life. It needs to be improved and should focus on verbal content only (Vrij, 2016). Verbal content is more diagnostic than nonverbal behaviour (Bond & DePaulo, 2006) and it may be easier to establish verbal than nonverbal baselines as Palena et al's (2018) study suggests. Future efforts can improve the comparable truth baseline technique integrating it with the techniques used for strategic questioning. For example, the interviewer can start the interview with the baseline approach and then employ strategic questioning using the already available within-subjects comparisons techniques reported in Vrij (2016), such as the reverse order technique (Ewens, Vrij, Mann, & Leal, 2016; Vrij, Leal, Mann, & Fisher, 2012) and the verifiability approach (Nahari, Vrij, & Fisher, 2014; Vrij, Nahari, Isitt, & Leal, 2016). In the reverse order technique, the interviewee is first asked to report the story in chronological order. Then, s/he is asked to report the same story from the end to the beginning, which typically results in truth tellers to report more reminiscences but fewer contradictions than liars (Vrij, 2016). According to the verifiability approach liars try not to report details which can be verified by an investigator. Consequently, the proportion of verifiable details (verifiable detail/(verifiable and unverifiable details combined) is higher for truth tellers than for liars (Nahari et al., 2014; Vrij, 2016).

Another possibility is to create a (semi)structured interview protocol which incorporates a baseline technique. A good example is the Assessment Criteria Indicative of Deception (ACID, Colwell et al., 2007). This protocol starts with a baseline question coupled with rapport building and is based on the concept of Differential Recall Enhancement (DRE), whereby it is thought that the use of mnemonics techniques and forced choice questions will result in truth tellers reporting more new details than liars. Colwell et al. (2007, 2013) used a small talk baseline (the last meal someone had or the first day of a semester). Future studies may explore how the ACID technique performs when a comparable truth baseline is used.

Future studies can also account for some of the limitations present in this study. First, the target event represented here is of low stakes. Increase in stakes makes a small talk baseline even less effective, as emotions experienced by the interviewee during the target response such as fear (of being caught, for the liar and of not being

believed, for the truth teller), may result in more pronounced differences between the two phases of the interview, regardless of sender's veracity (Ewens et al., 2014; Palena et al., 2018). Second, our participants did not receive any form of training in truth/lie detection. Therefore, their accuracy may benefit from training, as they would focus on more effective cues to truth/deception. For example, a recent meta-analysis (Hauch, Sporer, Michael, & Meissner, 2016) explored to what degree different types of training (e.g.: in verbal content, nonverbal behaviour, paralinguistic cues, etc.) affected credibility assessments. Their results showed that training did improve credibility assessments, particularly training focusing on speech content. Third, our sample consisted of university students. It may be worthwhile to explore how professionals (e.g. police officers) perform when they are exposed to a comparable truth baseline.

The authors have declared no conflicts of interest

References

Bond Jr, C. F., & DePaulo, B. M. (2006). Accuracy of deception judgments. *Personality and social psychology Review*, *10*(3), 214-234. DOI: 10.1207/s15327957pspr1003_2.

Brandt, D. R., Miller, G. R., & Hocking, J. E. (1980). The truth-deception attribution: Effects of familiarity on the ability of observers to detect deception. *Human Communication Research*, *6*(2), 99-110. DOI: 10.1111/j.1468-2958.1980.tb00130.x.

Brandt, D. R., Miller, G. R., & Hocking, J. E. (1982). Familiarity and lie detection: A replication and extension. *Western Journal of Communication (includes Communication Reports)*, *46*(3), 276-290. DOI: 10.1080/10570318209374086.

Caso L., Maricchiolo F., Bonaiuto M., Vrij A., Mann S. (2006). The impact of deception and suspicion on different hand movements. *Journal of Nonverbal behavior*, *30*(1), 1-19. DOI: 10.1007/s10919-005-0001-z.

Colwell, K., Hiscock-Anisman, C., & Fede, J. (2013). Assessment criteria indicative of deception: An example of the new paradigm of differential recall enhancement. In B. S. Cooper, D. Griesel, & M. Ternes (Eds.), *Applied issues in investigative interviewing, eyewitness memory, and credibility assessment* (pp. 259-291). New York: Springer.

Colwell, K., Hiscock-Anisman, C. K., Memon, A., Taylor, L., & Prewett, J. (2007). Assessment

Criteria Indicative of Deception (ACID): An integrated system of investigative interviewing

and detecting deception. *Journal of Investigative Psychology and Offender Profiling, 4*, 167–180. DOI: 10.1002/jip.73.

DePaulo, B. M., Charlton, K., Cooper, H., Lindsay, J. J., & Muhlenbruck, L. (1997). The accuracy-confidence correlation in the detection of deception. *Personality and Social Psychology Review, 1*(4), 346-357. DOI: 10.1207/s15327957pspr0104_5.

DePaulo, B.M., Lindsay, J. L., Malone, B.E., Muhlenbruck, L., Charlton, K, & Cooper, H. (2003). Cues to deception. *Psychological Bulletin, 129*, 74-118. DOI: 10.1037/0033-2909.129.1.74.

Ewens, S., Vrij, A., Jang, M., & Jo, E. (2014). Drop the small talk when establishing baseline behaviour in interviews. *Journal of Investigative Psychology and Offender Profiling, 11*(3), 244-252. DOI: 10.1002/jip.1414.

Ewens, S., Vrij, A., Mann, S., & Leal, S. (2016). Using the reverse order technique with non-native speakers or through an interpreter. *Applied Cognitive Psychology, 30*, 242–249. DOI:10.1002/acp.3196.

Feeley, T. H., deTurck, M. A., & Young, M. J. (1995). Baseline familiarity in lie detection. *Communication Research Reports*, *12*(2), 160-169. DOI: 10.1080/08824099509362052.

Frank, M. G., Yarbrough, J. D., & Ekman, P. (2006). Investigative Interviewing and the Detection of Deception. In T. Williamson (Ed.), Investigative Interviewing: Rights, Research and Regulation (pp. 229 – 255). Cullompton, Devon: Willan Publishing.

Gnisci A., Caso L., Vrij A. (2010). Have you made up your story? The effect of suspicion and liars' strategies on reality monitoring. *Applied cognitive psychology*, *24*(6), 762-773. DOI: 10.1002/acp.1584.

Granhag, P. A., & Hartwig, M. (2015). The Strategic Use of Evidence technique: A conceptual overview. In P. A. Granhag, A. Vrij, & B. Verschuere (Eds.), *Detecting*

Deception: Current Challenges and Cognitive Approaches (pp. 231-251). Chichester: John Wiley & Sons.

Granhag, P. A., & Vrij, A. (2010). Interviewing to Detect Deception. In P. A. Granhag (Eds.), *Forensic psychology in context: Nordic and international approaches* (pp. 75-93). Collumpton: Willian.

Hartwig, M., & Bond Jr, C. F. (2011). Why do lie-catchers fail? A lens model meta-analysis of human lie judgments. *Psychological Bullettin, 137*(4), 643-459. DOI: 10.1037/a0023589.

Hartwig, M., & Bond, C. F. (2014). Lie detection from multiple cues: A meta-analysis. *Applied Cognitive Psychology*, *28*(5), 661-676. DOI: 10.1002/acp.3052.

Hauch, V., Sporer, S. L., Michael, S. W., & Meissner, C. A. (2016). Does training improve the detection of deception? A meta-analysis. *Communication Research, 43*(3), 283-343. DOI: 10.1177/0093650214534974.

Inbau, F. E., Reid, J. E., Buckley, J. P., & Jayne, B. C. (2013). *Criminal interrogation and confessions* (5th edn). Burlington, MA: Jones & Bartlett Learning.

Jupe, L., Akehurst, L., Vernham, Z., & Allen, J. (2016). Teenage offenders' ability to detect deception in their peers. *Applied Cognitive Psychology*, *30*(3), 401-408. DOI: 10.1002/acp.3214.

Meissner, C. A., & Kassin, S. M. (2002). "He's guilty!": Investigator bias in judgments of truth and deception. *Law and human behavior*, *26*(5), 469.

Moston, S., & Engelberg, T. (1993). Police questioning techniques in tape recorded interviews with criminal suspects. *Policing and Society: An International Journal of Research and Policy, 3*, 223–237. DOI: 10.1080/10439463.1993.9964670.

Nahari, G., Vrij, A., & Fisher, R. P. (2014). Exploiting liars' verbal strategies by examining the verifiability of details. *Legal and Criminological Psychology, 19*, 227–239. DOI:10.1111/j.2044-8333.2012.02069.x.

Palena, N., Caso, L., Vrij, A., & Orthey, R. (2018). Deception detection through small talk and comparable truth baselines. *Journal of Investigative Psychology and Offender Profiling, 1-9*. DOI: 10.1002/jip.1495

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior research methods, instruments, & computers*, *31*(1), 137-149. DOI: 10.3758/BF03207704.

Vrij, A. (2008). *Detecting lies and deceit: Pitfalls and opportunities*. Chichester: John Wiley & Sons.

Vrij, A. (2014). Interviewing to detect deception. *European Psychologist*, *19*(3), 184-194. DOI: 10.1027/1016-9040/a000201.

Vrij, A. (2015). A cognitive approach to lie detection. In P. A. Granhag, A. Vrij, & B. Verschuere (Eds.), *Detecting Deception: Current Challenges and Cognitive Approaches* (pp. 205-230). Chichester: John Wiley & Sons.

Vrij, A. (2016). Baselining as a Lie Detection Method. *Applied Cognitive Psychology*, *30*(6), 1112-1119. DOI: 10.1002/acp.3288.

Vrij, A., Fisher, R. P., & Blank, H. (2017). A cognitive approach to lie detection: A meta-analysis. *Legal and Criminological Psychology*, *22*(1), 1-21. DOI: 10.1111/lcrp.12088.

Vrij, A., & Granhag, P. A. (2012). Eliciting cues to deception and truth: What matters are the questions asked. *Journal of Applied Research in Memory and Cognition*, *1*(2), 110-117. DOI: 10.1016/j.jarmac.2012.02.004.

Vrij, A., Leal, S., Mann, S., & Fisher, R. (2012). Imposing cognitive load to elicit cues to deceit: Inducing the reverse order technique naturally. *Psychology, Crime & Law*, *18*, 579–594. DOI:10.1080/1068316X2010.515987

Vrij, A., & Mann, S. (2001). Telling and detecting lies in a high-stake situation: the case of a convicted murderer. *Applied Cognitive Psychology*, *15*, 187–203. doi: 10.1002/1099-0720(200103/04)15:2<187::AID-ACP696>3.0.CO;2-A.

Vrij, A., Nahari, G., Isitt, R., & Leal, S. (2016). Using the verifiability lie detection approach in an insurance claim setting. *Journal of Investigative Psychology and Offender Profiling*, *13*(3), 183-197. DOI:10.1002/jip.1458.

Table 1. *Means, Standard Deviation and Confidence Intervals for Truth and Lie Accuracy according to the Baseline Condition*

	Comparable Truth			Small Talk		
	M	SD	CI	M	SD	CI
Truth accuracy	57.84	30.10	47.80, 67.88	49.44	29.66	39.41, 59.48
Lie accuracy	55.14	20.22	48.49, 61.88	45.42	21.49	38.14, 52.69

Table 2. *One Sample t-tests Comparing Observers' Accuracy Rates with a Meta Analysis Average Scores*[1] *and with Chance in each Baseline Condition.*

Comparable Truth Baseline						
	Test value: meta-analyses average scores			Test value: chance (50%)		
Accuracy	t	p	d	t	p	d
Total	.80	.42	.13	2.105	.04	.34
Truth	.64	.52	-.10	1.584	.12	.25
Lie	2.447	.02	.39	1.544	.13	.25
Small Talk Baseline						
	Test value: meta-analyses average scores			Test value: chance (50%)		
Accuracy	t	p	d	t	p	d
Total	-2.124	.04	-.34	-.83	.40	-.13
Truth	-2.338	.02	-.38	-.11	.91	-.01
Lie	-.44	.66	-.07	-1.279	.20	-.21

[1]Bond & DePaulo (2006). Total accuracy, 54%. Truth accuracy, 61%. Lie accuracy, 47%.

Table 3. *One Sample t-tests Exploring Observers' Sensitivity and Response Bias in each Baseline Condition*

	Comparable Truth			Small Talk		
	t	p	d	t	p	d
d'	2.129	.04	.34	-.817	.42	-.13
β	3.737	.001	.60	2.629	.01	.43

Appendix 5: Caso, L., Palena, N., Carlessi, E., & Vrij, A. (2019). Police accuracy in truth/lie detection when judging baseline interviews. *Psychiatry, Psychology, and Law.* doi: 10.1080/13218719.2019.1642258. [PA4]

Abstract

Research has shown that a Comparable Truth Baseline approach elicits more cues to deception and results in higher accuracy rates than a Small Talk Baseline. Past research focused on accuracy rates obtained by laypeople. In the current experiment we examined whether the Comparable Truth Baseline also has a positive effect on law enforcement personnel accuracy. In this study, ninety-five police officers judged ten interviews, whereby half of the senders told the truth and the other half lied about a mock undercover mission. Half of the interviews included only questioning about the event under investigation, whereas the other half also included questioning aimed at creating a Comparable Truth Baseline. Observers did not differ in their total and truth accuracy, but those who watched interviews with a Comparable Truth Baseline obtained higher lie detection accuracy rates than those who watched interviews without the Baseline questioning. Signal detection analyses showed that this effect could be attributed to a decreased response bias in the Comparable Truth Baseline condition.

Keywords: Comparable Truth Baseline; Baseline interviewing; Deception Detection; Police officers' accuracy; Interviewing techniques; Credibility assessment; Interrogation

Police accuracy in truth/lie detection when judging baseline interviews

Introduction

Cues to Deception and Interviewing Techniques

Detecting deception is a flourishing research area, from both a theoretical (Bond, Levine, & Hartwig, 2015; Caso, Maricchiolo, Livi, Vrij, & Palena, 2018; Vrij, Hartwig, & Granhag, 2018; Walczyk, Harris, Duck, & Mulay, 2014) and applied perspective (Vrij & Fisher, 2016). Research initially focused on searching for reliable cues to deceit that liars display spontaneously (see Vrij [2008] for a comprehensive review of this research). This research has proven to be largely unsuccessful: Cues to deception in this setting are typically faint and unreliable (DePaulo et al., 2003; Vrij, 2008).

Due the paucity of support for the search for reliable cues to deception that liars display spontaneously, scholars have started to examine whether such cues can be elicited or enhanced through specific interviewing protocols (Vrij & Granhag, 2012; Vrij, 2014). Several interview protocols have emerged. In the Strategic Use of Evidence technique, the available evidence is presented to interviewees in such a way that it elicits within-statement and between-statements inconsistencies in liars and elicits admissions from them (Granhag & Hartwig, 2008, 2015; May, Granhag, & Tekin, 2017). The Cognitive Credibility Assessment approach builds on the research findings that lying is usually more mentally taxing than telling the truth (Vrij, Fisher, & Blank 2017). This approach is based on the idea that questions can be asked, or instructions can be given, that elicit different mental processes in truth tellers and liars (Vrij, 2015; Vrij, Fisher, Blank, Leal, & Mann, 2016). Another approach, the Assessment Criteria Indicative of Deception technique, is a (semi)structured interview which builds on the memory-enhancement techniques (mnemonics) which are part of the Cognitive Interview (Fisher & Geiselman, 1992). Such memory-enhancing techniques have a different impact on truth tellers and liars when considering cues such as response length, amount of details and coherence (Colwell et al., 2009; Colwell, Hiscock-Anisman, Memon, Taylor, & Prewett, 2007). This so-called interviewing to detect deception approach has proven to be successful and

several of the techniques proposed and tested in this area are ready to be implemented in real life (Vrij & Fisher, 2016; Vrij, 2018).

Observers' Lie Detection Accuracy

Another line of research examines observers' accuracy at evaluating interviewees credibility. The most comprehensive meta-analysis published to date showed that people –laypersons and professionals alike- are poor at this task. The average accuracy, obtained by over 24,000 observers, was 54%, which is only just above the level of chance (50%) (Bond & DePaulo, 2006). However, the studies presented in the meta-analysis did not account for the effect of strategic interviewing, which is worth exploring. Luke et al. (2016) found that training law enforcements in the SUE technique increased their accuracy rates up to 65%. Similarly, another study found that police officers trained in the Cognitive Credibility Assessment approach asked more effective questions and improved their accuracy rates up to 74% (Vrij, Leal, Mann, Vernham, & Brankaert, 2015). Finally, Colwell et al. (2009) found that observers trained in the Assessment Criteria Indicative of Deception technique obtained accuracy rates up to 77%. These pictures show strong potential for training practitioners into strategic interviewing.

The Baseline Approach

Another interview technique is based on the rationale that if the observer has previous knowledge of sender's truthful behavior, this can be used as a truthful baseline to inform observer's decisions (see Vrij, 2008). Indeed, research has supported this claim, as it was found that being familiar with the sender increases observer's performance (Brandt, Miller, & Hocking, 1980, 1982). Feeley, de Turk, and Young (1995) found a positive linear relationship between the level of familiarity and observers' performance; that is, the more the observer was familiar with senders' truthful behavior, the more s/he was accurate in detecting deception. However, research also shows that familiarity only works when the baseline is truthful. For example, Garrido and Masip (2001) provided observers with a baseline behavior of the sender, which could be either truthful or deceitful. The authors found that observers benefited from the baseline exposure only when this was truthful.

The type of baselining reported above is different from the one that one can expect in investigative interviewing. The suspect and the interviewer are often strangers; therefore, such *familiarity baseline* is difficult to obtain (if possible at all) during investigations. Instead, investigators must obtain the baseline during the first phases of the interview. There have been some suggestions on how to create a baseline *on the spot*.

Initially, it was suggested to create such a baseline by asking neutral, non-threatening questions (Frank, Yarbrough, & Ekman, 2006) but both theoretical reasons and experimental results have revealed that this method is ineffective (Ewens, Vrij, Yang, & Yo, 2014; Moston & Engelberg, 1993; Palena, Caso, Vrij, & Orthey, 2018). This approach is thought to be a striking misuse of psychological theory (Moston & Engelberg, 1993; Vrij, Leal, Jupe, & Harvey, 2018). The problem is that the baseline and target (which concerns the event under investigation) sections are fundamentally different in such an approach. For example, the stakes are higher in the target than in the baseline section and interviewee's engagement is often different between the two sections. As a result, not only liars but also truth tellers change their behaviour when the baseline and target responses are compared (Ewens et al., 2014; Palena et al., 2018; Vrij, 2016). Therefore, both theoretical explanations (Moston & Engelberg, 1993; Vrij, 2016; Vrij, Leal, Jupe, & Harvey, 2018) and experimental findings (Caso, Palena, Vrij, & Gnisci, 2019; Ewens, Vrij, Yang, & Yo, 2014) support the assumption that a baseline created through small talk (i.e., Small Talk Baseline) is ineffective.

Vrij (2016) states that a different baselining approach, the creation of a comparable truth baseline (CTB), may be more fruitful. In the CTB approach factors such as engagement, emotion, cognitive load and context are kept comparable between the baseline and the target sections of the interview. To keep such factors comparable, baseline questioning should pertain to the same topic of that discussed when posing question concerning the event under investigation. Similarly, questions should not be posed in a way that creates different emotional arousal. Two recent studies found that using a CTB, liars changed their behaviour and speech more than truth tellers (Palena et al., 2018) and that observers provided with a CTB obtained high accuracy rates than those provided with a small talk baseline (Caso et al., 2019).

The main aim of the present study was to explore whether a CTB would also be effective with law enforcement personnel as observers. Involving law enforcement personnel as participants in studies is important as it may make it more likely that they will endorse the findings and, if the findings are positive, start to use the techniques themselves (Vrij & Granhag, 2012).

We decided to compare the CTB approach to a "no baseline" rather than to a Small Talk Baseline for two main reasons. First, we wanted to test the hypothesis that having a comparable baseline behaviour of the interviewee makes the observers more accurate than not having such reference. This builds on the idea that accuracy is increased because a comparable baseline reduces the effect of interpersonal differences (Vrij, 2016). Second, we did not want to compare the CTB with the small talk baseline because: i) research has already shown that such an approach does not work because truth tellers and liars appear equally deceptive (Ewens et al., 2014; Palena et al., 2018); and ii) research with laypersons has shown that observers provided with a CTB outperform those provided with a Small Talk baseline (Caso et al., 2019) This is unlikely to change with practitioners, as the inefficacy of the Small Talk baseline approach depends on its underpinnings rather than on the observer: Both truth tellers and liars display differences between baseline and target periods with this approach (Ewens et al., 2014). Therefore, no reliable cue to deception appears and neither laypersons nor practitioners can benefit from the small talk baseline approach.

Building on previous results on objective cues to deception elicited with a CTB (Palena et al., 2018) and on the rationale presented above, we expected that practitioners in the CTB condition would achieve higher total (Hypothesis 1), truth (Hypothesis 2) and lie (Hypothesis 3) accuracy rates than practitioners in the no-baseline condition.

Materials and Method

Participants

A total of 95 practitioners (88 men and seven women) took part in the experiment. Of them, 42 belonged to the state police, 28 to the financial and economic crimes police, and 25 to the Italian Military police (Carabineri). Age ranged from 28 to 58, with a

mean of $M = 45.39$ ($SD = 6.71$). Professional experience ranged from five to 38 years, with a mean of $M = 23.84$ ($SD = 7.82$). One participant was excluded from the analyses because he did not follow the instructions.

Design

The experiment utilised a 2 (Baseline: no baseline vs. CTB, between-subjects) by 2 (Veracity: Truth tellers vs. Liars, within-subjects) mixed design. For the factor Baseline, observers in the no-baseline condition just saw suspects being questioned about the event under investigation. On the other hand, observers in the CTB condition saw interviews where suspects were also questioned about an event other than -but comparable to- that under investigation. For the factor Veracity, half of the senders told the truth, whereas the other half lied. For the CTB condition, all senders truthfully reported the additional event, which served as the CTB. The dependent variables were the three accuracy rates obtained by the observers: total accuracy, truth accuracy and lie accuracy. Accuracy rates were obtained using the formula $\frac{number\ of\ correct\ answers}{total\ number\ of\ possible\ answers} \cdot 100$.

Procedure

The first and the second author made an appointment with the prosecutor (who is the person in charge for recruiting participants for any research project with practitioners) and presented him an outline of the study. The importance of conducting research with practitioners was stressed. The prosecutor was interested in the project and proceeded to contact high rank officials belonging to the three organisations (State Police, Economic and Financial Police, and Military Police), who were informed about the research goals. High rank officers then provided the authors with a list of participants from the three police organisations. The participants were then contacted to take part in the study. Data were collected in three different places, one for each police organisation. Each participant took part in the experiment individually. Upon arrival, s/he was welcomed and briefed about the aim of the study. They were informed that they were going to watch a series of interviews with different people and then decide for each interviewee whether s/he was lying or telling the truth. They were not informed about the number of interviews they would see and about the

truth/lie telling ratio in the interviews to avoid these aspects affecting their decisions. In total, each judge evaluated every sender (ten in total).

Participants in the no baseline condition were not given further information. Participants in the CTB condition were informed that interviews were split in two sections. The first section was the baseline, the second was the target section of the interview. They were also informed that senders were always honest in the baseline section and therefore they had to take a decision about sender's (dis)honesty only regarding the target section. They were invited to examine deviations from the baseline to inform their decisions on the rationale that the more a sender changed his/her behaviour and speech between the two phases, the more likely it was that s/he was lying. Information about which behaviours or speech patterns might be indicative of deceit was not provided. The participants were not informed about the aims of the study (comparing accuracy of practitioners in the two experimental conditions).

In the experiment, veracity decisions were made answering the dichotomous question "Do you think the interviewee was…"Lying" or "Telling the truth". After the experimenter felt confident that the participant understood the instruction, s/he was left alone to carry out the lie detection test, which lasted on average about 30 minutes. Once the participant had finished the test, they were thanked and debriefed.

Stimulus material

Senders depicted in the videoclips had performed a mock undercover mission. The videoclips were obtained from a previous study (reference not reported for blind review). The mock crime started with participants receiving an envelope from the experimenter that contained the password for accessing a laptop. Once logged in, the sender read a Word document file that informed him/her to look for a CD-ROM in a backpack that was placed in the room. The CD-ROM depicted a man informing the sender to look for a key in the same backpack, which would open a safe deposit box placed near the window in the same room. The safe deposit box contained further written instructions, asking the participant to send an email to a specific address and to wait for a person to collect them from the room. Everything until that moment was part of the CTB, whereas everything happened that after this point was part of the target event, the event under investigation.

After a short while the person (a confederate) arrived and gave a newspaper to the sender, informing him/her to read it for further instructions. These instructions informed the sender that s/he had to go to an adjacent room and to look for a USB stick hidden behind a coat hook. The sender had to take that USB stick and put it in place of a second USB stick that was hidden in a book placed in a wardrobe. The sender then had to leave the newspaper next to the book and keep the second USB (the one found in the book) until the end of the experiment. Once these tasks were completed, the sender had to come back to the first room and wait for the interviewer.

We used ten senders in total. For the CTB, the observer watched both the baseline and target sections, which was composed by the sender answering a free recall baseline question and a free recall target question. All ten senders answered the baseline question truthfully, whereas for the target question, five of them told the truth and five lied. The veracity status was counterbalanced. For the condition without the baseline, the observers only saw the target question and answers.

The ten clips in the "no baseline" condition lasted 76.70 seconds on average (*SD* = 17.36), those in the CTB lasted 170.10 seconds on average (*SD* = 20.51). This difference is due to the presence of the baseline.

Results

Total Accuracy

To test Hypothesis 1, an ANOVA was carried out with Baseline (no baseline vs. CTB) as factor, and the total accuracy rate as dependent variable. No significant effect appeared for the Baseline factor, $F(1, 92) = 1.554$, $p = .21$, $d = -0.26$ [-0.66, 0.15], post-hoc achieved power .24. Mean accuracy for participants in the no baseline condition (*M* = 49.58; *DS* = 13.67; 95% CI [45.61, 53.55]) was similar to that of participants in the CTB (*M* = 53.26; *DS* = 14.91; 95% CI [48.83, 57.69]). Hypothesis 1 was thus rejected.

Truth Accuracy

Preliminary tests assessing ANOVA assumptions showed that homoscedasticity was not respected, $F(1, 92) = 7.110$, $p = .01$. Consequently, we run a Mann-Whitney *U* test to test Hypothesis 2. The truth accuracy for participants in the no-baseline

condition (M = 60.00; SD = 17.01; 95% CI [55.06; 64.93]) did not differ from that of participants in the CTB condition (M = 51.73; SD = 22.93; 95% CI [44.93, 58.55]), U = 863.50, p = .058, post-hoc achieved power .48. Therefore, Hypothesis 2 was also rejected.

Lie Accuracy

To test Hypothesis 3, an ANOVA was carried out with Baseline as factor and lie accuracy as the dependent variable. The Baseline effect was significant, $F(1, 92)$ = 17.16, $p < .001$, d = -0.85 [-1.27, -.43], post-hoc achieved power .98. Supporting Hypothesis 3, observers in the CTB condition (M = 54.78; SD = 20.41, 95% CI [48.72, 60.84]) outperformed those in the no baseline condition (M = 39.16; SD = 15.95, 95% CI [34.53, 43.80]). In addition, more observers (63%) in the CTB condition than in the control condition (25%) obtained an accuracy rate of at least 60%, χ^2 (1, N = 94) = 13.82, $p < .001$, Cramer's V = .38. Figure 1 depicts a density plot of such values.

enter Figure 1 about here

Signal detection analyses.

Our results supported Hypothesis 3. However, lie detection accuracy for participants in the CTB condition did not differ from chance, $t(45) = 1.589$, p = .12, BF_{10} = 0.513, d = 0.23 [-0.06, 0.52], post-hoc achieved power .33. This makes an alternative explanation possible. Since with a baseline approach (including the CTB) both truth tellers and liars appear to behave and speak differently between the two phases of the interview (Palena et al., 2018; Vrij, 2016), the increased lie accuracy may be partly due to a lie bias. We tested this possibility by exploring participants' response bias. Although historically the β value was the preferred measure for bias, some scholars have suggested to use c instead of β (see Stanislaw & Todorov, 1999) as the former is less biased than the latter. c is described as deviations, in standard deviation units, from the neutral point (where neither answer is preferred, i.e., there is no bias), which is set at 0. If the c value is greater than 0, there is a bias toward responding "no" (in our context, "truth teller"); if the value is less than 0 there is a bias toward responding "yes" (in our context, "liar"). A one-sample t-test with c as the dependent variable and 0 as test score, was significant for participants in the no baseline condition, $t(47)$

= 7.803, $p < .001$, $d = 1.13$ [0.76, 1.49], post-hoc achieved power 1, but not significant for participants in the CTB condition, $t(45) = -.555$, $p = .58$, $d = -0.08$ [-0.37, 0.21], $BF_{01} = 5.404$, post-hoc achieved power .08. An independent sample *t*-test with Baseline (no baseline vs. CTB) as the factor and *c* as the dependent variable showed that participants in the "no baseline" condition obtained higher scores ($M = 0.29$; $SD = 0.26$, 95% CI [0.22, 0.37]) than those in the CTB condition ($M = -0.05$; $SD = 0.61$, 95% CI [-0.23, 0.13]), $t(60.62) = 3.547$, $p = .001$, $d = 0.73$ [0.31, 1.14], post-hoc achieved power 0.94. Results for the *c* score therefore suggest that participants in the "no baseline" condition were truth biased, whereas those in the CTB displayed no bias. Table 1 displays the rates of truth and lie judgments in each condition.

enter Table 1 about here

Given the response bias results, we also analysed *d*' scores, which is a measure of sensitivity reported in standard deviation units (Stanislaw & Todorov, 1999). A *d'* score of 0 indicates an inability to distinguish between the two stimuli (truth telling vs. lying), whereas scores greater than 0 indicate that participants were able to make such a distinction. Neither participants in the "no baseline" condition ($M = -0.002$; $SD = 0.82$, 95% CI [-0.24, 0.23]), $t(47) = -0.021$, $p = .98$, $BF_{01} = 6.376$, post-hoc achieved power .05, nor those in the CTB condition ($M = 0.24$; $SD = 0.99$, 95% CI [-0.05, 0.53]), $t(45) = 1.620$, $p = .11$, $BF_{01} = 1.864$, post-hoc achieved power .36, were able to discriminate truth tellers from liars. Considering the analyses on response bias and sensitivity, we can conclude that the difference in accuracy between the two conditions was driven by the difference in response bias.

Discussion

In this experiment, we tested Italian police officers' ability to evaluate credibility when provided with a CTB compared to when no baseline was presented. We found support for Hypothesis 3, as officers in the CTB condition (54.78%) outperformed those in the no-baseline condition (39.16%) in terms of lie accuracy. However, further analyses showed that this was the result of officers in the CTB condition being less biased rather than being more accurate.

Our results are in alignment with the Truth Default Theory (Levine, 2014), which predicts that observers usually tend to believe others, unless "deceptive triggers" appear and make the observer considering the possibility of deception. In our context, such "deceptive triggers" may originate from behavioural differences that appear- for both truth tellers and liars- between the baseline and the target phase of the interview. However, the idea of humans having a cognitive default (truth) bias has been questioned (see Street, 2015). Future studies should try to disentangle this issue.

The baseline approach comes with several issues. First, truth tellers may be classified as liars because they also often change their behaviour throughout the interview. Second, the problems with the CTB are also shown in Palena et al. (2018), who found that it only worked for one cue (spatial details) when comparisons between baseline and target periods were made. The authors noted that their result was likely due to the task performed by the interviewees, which was mainly spatial. Indeed, participants in Palena's et al. (2018) study had to commit a mock crime which requested them to explore different rooms and interact with several objects, making the statements rich in spatial details. This makes clear that the cues to be used with the CTB approach are tightly connected to the content of the story itself.

Third, obtaining a CTB in laboratory settings is easy as the experimenters exert full control over baseline veracity. However, in real life it may be difficult to obtain a CTB which is really truthful and comparable, as a ground truth is often missing. And, in case the baseline is a lie, it loses its efficacy (Garrido & Masip, 2001).

In sum, the positive result for CTB concerning lie accuracy was the result of reduced probability of guessing truth, and no differences were found for total and truth accuracy. This shows no real positive effect for using a CTB. Previous research on the effectiveness of various interview techniques reached better results than those we found in the current experiment. This may be partly due to the fact that such techniques are more active approaches -the interviewer conducts the interview actively- whereas the baseline approach is more passive. The only thing the interviewer has to do is to create the baseline. It has already been suggested that active strategic interviewing is more effective for deception detection (Vrij & Granhag, 2012; Vrij, 2014).

Limitations

There were some limitations in the present study. First, our stimulus material was relatively low-stakes. This may have affected the results, although it is not straightforward to predict how. However, baseline research relies in part on behavioural patterns and such patterns are affected by stakes (Hartwig & Bond, 2014). It is therefore important to start examining the baseline technique also in higher-stakes situations. Second, observers only watched ten senders, which may not represent variations in senders' performance adequately. Third, it appeared that police officers participating in this study do not use Small Talk baselining or any other type of baselining in their daily practice. Given the problems associated with baselining, this cannot be considered a bad thing. Rather, they reported that when they interview real suspects, they tend to base their decision on available evidence Yet, we did not provide our participants with evidence.

The authors have declared no conflict of interests

The authors declare that all measures and conditions have been reported, that one participant was excluded from the analyses because he did not follow the instructions, and that the sample size was determined in order to have at least 30 participants per cell, accepting any further available participant.

References

Bond, C. F. Jr., & DePaulo, B. M. (2006). Accuracy of deception judgments. *Personality and Social Psychology Review, 10*(3), 214–234. doi:10.1207/s15327957pspr1003_2.

Bond, C. F., Jr., Levine, T. R., & Hartwig, M. (2015). New findings in non-verbal lie detection. In P. A. Granhag, A. Vrij, & B. Vershuere (Eds.), *Detecting deception: Current challenges and cognitive approaches* (pp. 37-58). Chichester, UK: John Wiley.

Brandt, D. R., Miller, G. R., & Hocking, J. E. (1980). The truth-deception attribution: Effects of familiarity on the ability of observers to detect deception. *Human Communication Research, 6*(2), 99-110. doi: 10.1111/j.1468-2958.1980.tb00130.x.

Brandt, D. R., Miller, G. R., & Hocking, J. E. (1982). Familiarity and lie detection: A replication and extension. *Western Journal of Communication (includes Communication Reports), 46*(3), 276-290. doi: 10.1080/10570318209374086.

Caso, L., Maricchiolo, F., Livi, S., Vrij, A., & Palena, N. (2018). Factors affecting observers' accuracy when assessing credibility: The effect of the interaction between media, senders' competence and veracity. *The Spanish Journal of Psychology*.

Caso, L., Palena, N., Vrij, A., & Gnisci, A. (2019). Observers' performance at evaluating truthfulness when provided with Comparable Truth or Small Talk Baselines. *Psychiatry, Psychology, and Law.*

Colwell, K., Hiscock-Anisman, C. K., Memon, A., Colwell, L. H., Taylor, L., & Woods, D. (2009). Training in assessment criteria indicative of deception to improve credibility judgments. *Journal of Forensic Psychology Practice, 9*(3), 199–207. doi:10.1080/15228930902810078.

Colwell, K., Hiscock-Anisman, C. K., Memon, A., Taylor, L., & Prewett, J. (2007). Assessment criteria indicative of deception (ACID): An integrated system of investigative interviewing and detecting deception. *Journal of Investigative Psychology and Offender Profiling, 4*(3), 167–180. doi:10.1002/jip.73.

DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological Bulletin, 129*(1), 74–118. doi:10.1037/0033-2909.129.1.74.

Ewens, S., Vrij, A., Jang, M., & Jo, E. (2014). Drop the small talk when establishing baseline behaviour in interviews. *Journal of Investigative Psychology and Offender Profiling, 11*(3), 244–252. doi:10.1002/jip.1414.

Feeley, T. H., deTurck, M. A., & Young, M. J. (1995). Baseline familiarity in lie detection. *Communication Research Reports, 12*(2), 160-169. doi: 10.1080/08824099509362052.

Fisher, R. P., & Geiselman, R. E. (1992). *Memory enhancing techniques for investigative interviewing: The cognitive interview*. Charles C. Thomas Publisher.

Frank, M. G., Yarbrough, J. D., & Ekman, P. (2006). Investigative interviewing and the detection of deception. In T. Williamson (Ed.), *Investigative interviewing: Rights, research and regulation* (pp. 229–255). Cullompton, Devon: Willan.

Garrido, E., & Masip, J. (2001). Previous exposure to the sender's behavior and accuracy at judging credibility. In R. Roesch, R. R. Corrado, & R. J. Dempster (Eds.), *Psychology in the courts: International advances in knowledge*, (pp. 271-287). London: Routledge.

Granhag, P. A., & Hartwig, M. (2008). A new theoretical perspective on deception detection: On the psychology of instrumental mind reading. *Psychology, Crime & Law, 14*(3), 189–200. doi:10.1080/10683160701645181.

Granhag, P. A., & Hartwig, M. (2015). The Strategic Use of Evidence technique: A conceptual overview. In P. A. Granhag, A. Vrij, & B. Verschuere (Eds.), *Detecting deception: Current challenges and cognitive approaches* (pp. 231–252). Chichester: John Wiley and Sons.

Hartwig, M., & Bond, C. F. Jr. (2014). Lie detection from multiple cues: A meta-analysis. *Applied Cognitive Psychology, 28*, 661–676. doi:10.1002/acp.3052.

Levine, T. R. (2014). Truth-Default Theory (TDT): A Theory of Human Deception and Deception Detection. *Journal of Language and Social Psychology*, *33*(4), 378–392. doi:10.1177/0261927X14535916.

Luke, T. J., Hartwig, M., Joseph, E., Brimbal, L., Chan, G., Dawson, E., ... & Granhag, P. A. (2016). Training in the Strategic Use of Evidence technique: Improving deception detection accuracy of American law enforcement officers. *Journal of Police and Criminal Psychology*, *31*(4), 270-278.

May, L., Granhag, P. A., & Tekin, S. (2017). Interviewing Suspects in Denial: On How Different Evidence Disclosure Modes Affect the Elicitation of New Critical Information. *Frontiers in psychology*, *8*, 1154. doi: 10.3389/fpsyg.2017.01154.

Meissner, C. A., & Kassin, S. M. (2002). "He's guilty!": Investigator Bias in Judgments of Truth and Deception. *Law and Human Behavior*, *26*(5), 469-480. doi:10.1023/A:1020278620751.

Moston, S. J., & Engelberg, T. (1993). Police questioning techniques in tape recorded interviews with criminal suspects. *Policing and Society*, *6*, 61–75. doi:10.1080/10439463.1993.9964670.

Palena, N., Caso, L., Vrij, A., & Orthey, R. (2018). Detecting deception through small talk and comparable truth baselines. *Journal of Investigative Psychology and Offender Profiling*, *15*(2), 124-132. doi:10.1002/jip.1495.

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior research methods, instruments, & computers*, *31*(1), 137-149.

Street, C. N. (2015). ALIED: Humans as adaptive lie detectors. *Journal of Applied Research in Memory and Cognition*, *4*(4), 335-343. doi: 10.1016/j.jarmac.2015.06.002.

Vrij, A. (2008). *Detecting lies and deceit: Pitfalls and opportunities*. Chichester: John Wiley and Sons.

Vrij, A. (2014). Interviewing to detect deception. *European Psychologist*, *19*, 184–194. doi:10.1027/1016-9040/a000201.

Vrij, A. (2015). A cognitive approach to lie detection. In P. A. Granhag, A. Vrij, & B. Verschuere (Eds.), *Detecting deception: Current challenges and cognitive approaches* (pp. 205–230). Chichester: John Wiley and Sons.

Vrij, A. (2016). Baselining as a lie detection method. *Applied Cognitive Psychology, 30*, 1112–1119. doi:10.1002/acp.3288.

Vrij, A. (2018). Verbal Lie Detection Tools From an Applied Perspective, In J. P. Rosenfeld (Eds.), *Detecting Concealed Information and Deception* (pp. 297-327). London, UK: Elsevier Academic Press.

Vrij, A., Fisher, R. P., Blank, H., Leal, S., & Mann, S. (2016). A cognitive approach to elicit verbal and nonverbal cues to deceit. In J.W. Van Prooijen, & P. A. M. Van Lange (Eds.), *Cheating, corruption, and concealment: The roots of dishonesty* (pp. 284-302). Cambridge: Cambridge University press.

Vrij, A., & Fisher, R. P. (2016). Which lie detection tools are ready for use in the criminal justice system?. *Journal of Applied Research in Memory and Cognition, 5*(3), 302-307. doi: 10.1016/j.jarmac.2016.06.014.

Vrij, A., Fisher, R. P., & Blank, H. (2017). A cognitive approach to lie detection: A meta-analysis. *Legal and Criminological Psychology, 22*(1), 1-21. doi:doi:10.1111/lcrp.12088.

Vrij, A., & Granhag, P. A. (2012). Eliciting cues to deception and truth: What matters are the questions asked. *Journal of Applied Research in Memory and Cognition, 1*(2), 110–117. doi:10.1016/j.jarmac.2012.02.004.

Vrij, A., Hartwig, M., & Granhag, P. A. (2018). Reading lies: Nonverbal communication and deception. *Annual Review of Psychology*.

Vrij, A., Leal, S., Jupe, L., & Harvey, A. (2018). Within-subjects verbal lie detection measures: A comparison between total detail and proportion of complications. *Legal and Criminological Psychology, 23*(2), 265-279. doi: 10.1111/lcrp.12126.

Vrij, A., Leal, S., Mann, S., Vernham, Z., & Brankaert, F. (2015). Translating theory into practice: Evaluating a cognitive lie detection training workshop. *Journal of*

Applied Research in Memory and Cognition, *4*(2), 110-120. doi:10.1016/j.jarmac.2015.02.002.

Wagenmakers, E.-J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., . . . Morey, R. D. (2018). Bayesian inference for psychology. Part II: Example applications with JASP. *Psychonomic Bulletin & Review*, *25*(1), 58-76. doi:10.3758/s13423-017-1323-7.

Walczyk, J. J., Harris, L. L., Duck, T. K., & Mulay, D. (2014). A social-cognitive framework for understanding serious lies: Activation-decision-construction-action theory. *New Ideas in Psychology*, *34*, 22-36. doi: 10.1016/j.newideapsych.2014.03.001.

Figure 1. *Density plot for Lie Accuracy according to the Baseline condition.*

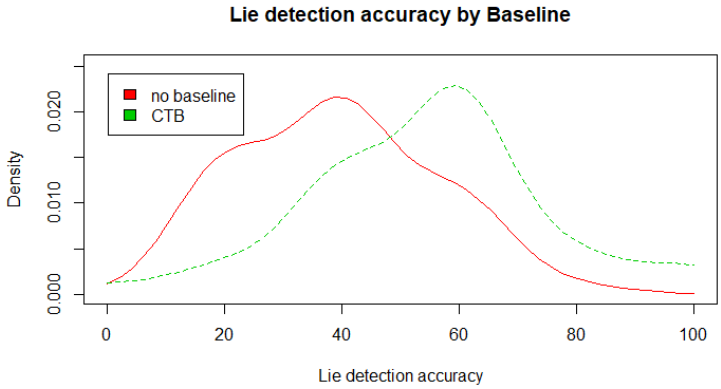


Table 1. *Means and Standard Deviations for rates of truth and lie judgments in each condition.*

	Baseline	Decides Truth	Decides Lie
Mean	No Baseline	60.42	39.58
	Comparable truth baseline	48.48	51.52
Standard deviation	No Baseline	9.22	9.22
	Comparable truth baseline	15.77	15.77

Appendix 6: Caso, L., Palena, N., Vrij, A., & Melocchi, L. (in preparation). Implementing the use of a checklist with comparable truth and small talk baselines: The effect on laypersons' lie detection accuracy. [IP3].

Abstract

Training programmes, checklists, and interviewing techniques are being developed to help lie detection. In this experiment, we combined the baseline technique (either comparable truth- where the baseline section is comparable to the section of the interview dealing with the event under investigation, the target section- or small talk- where the baseline is created through chit-chat) with the use of a checklist. One-hundred and twenty participants took part in the experiment and rated ten interviewees, who either told the truth or lied, using a checklist. Observers made their veracity decisions according to how much different they thought the senders appeared between the two phases of the interview (the baseline section and the target section). Accuracy rates did not differ between the two baseline conditions, but observers in the comparable truth condition noticed more substantial differences within liars' statements than within truth tellers' statements (as predicted by the baseline technique), whereas participants in the small talk condition showed the opposite pattern (contradicting the baseline technique). Our results support the assumption that a small talk baseline is ineffective and suggest that future studies focusing on ways to refine the checklist are needed.

Keywords: baseline technique, interviewing to detect deception, checklists, lie detection accuracy

Implementing the use of a checklist with comparable truth and small talk baselines: The effect on laypersons' lie detection accuracy

Research has shown that people's ability to detect lies is modest (Bond Jr. & DePaulo, 2006, 2008; Vrij, 2008), due to the finding that cues to deception are faint and unreliable (DePaulo et al., 2003; Hartwig & Bond Jr, 2011; Vrij & Granhag, 2012). Further, both senders' and observers' characteristics play an important role in lie detection accuracy (Caso, Maricchiolo, Livi, Vrij, & Palena, 2018). Interviewing to detect deception research aims at magnifying differences between truth tellers and liars through specific interviewing strategies (Vrij, 2014; Vrij & Granhag, 2012). Several techniques have been introduced, including the Assessment Criteria Indicative of Deception (ACID) procedure (Colwell, Hiscock-Anisman, & Fede, 2013; Colwell, Hiscock-Anisman, Memon, Taylor, & Prewett, 2007); the Strategic Use of Evidence (SUE) procedure (Granhag & Hartwig, 2015) and the Cognitive Credibility Assessment (CCA) procedure (Vrij, Fisher, & Blank, 2017). Some of these procedures appear to be ready for use in the criminal justice system (Vrij & Fisher, 2016).

Another procedure that received some attention in research is the use of a baseline, which aims at controlling for individual differences between interviewees. It assumes that the interviewer should start the interview asking questions for which s/he knows the interviewee will tell the truth. The (non)verbal behaviour displayed by the interviewee serves as a baseline. Then, the interviewer will start asking questions related to the event under investigation and will look for differences in (non)verbal behaviour between the baseline and target phases. Differences will be then interpreted as signs of deceit (Frank, Yarbrough, & Ekman, 2006; Vrij, 2016).

There are three important points to be noted about the baseline technique. First, practitioners (Inbau, Reid, Buckley, & Jayne, 2013) and some academics (Frank et al., 2006) claimed that such baseline questions can be general, non-threatening questions (*small talk baseline*). However, research has shown that such an approach does not work as both truth tellers and liars change their (non)verbal behaviour between the baseline and target phases when such a comparison is made (Caso, Maricchiolo, Bonaiuto, Vrij, & Mann, 2006; Ewens, Vrij, Jang, & Jo, 2014; Gnisci,

Caso, & Vrij, 2010). Second, if the baseline and the target phases of the interview are similar (in content, context, emotional and cognitive involvement, see Vrij, 2008, 2016), we have a *comparable truth baseline* (see also Vrij & Mann 2001, for a real-life case), which may be less affected by the problems reported above. Indeed, recent studies found that when a comparable truth baseline was introduced, liars' responses in the baseline and target phases were more different from each other than those of truth tellers (Palena, Caso, Carlotto, De Mizio, & Marciali, 2017; Palena, Caso, Vrij, & Orthey, 2018). Palena et al. (2017) explored the efficacy of such a comparable truth baseline. They asked their participants to commit a mock crime. While performing the mock crime, the participants met a confederate. Everything that happened before such a meeting served as the comparable truth baseline (about which all participants told the truth), whereas everything that happened after the meeting was the actual event under investigation (about which half of the sample told the truth and the other half lied) for which veracity had to be assessed and served as the target section. Palena et al. (2017) found that liars reported more contextual (spatial) and sensory (visual) details in the baseline section than in the target section, whereas truth tellers' differences did not reach statistical significance.

Third, a recent study found that the comparable truth baseline had a positive effect on untrained observers' accuracy compared to the small talk baseline. Caso, Palena, Vrij, and Gnisci (2019) showed ten videos to their participants (observers)- who either saw interviews conducted via a comparable truth baseline or conducted via a small talk baseline- and the observers had to decide whether the senders were telling the truth or lying. All the senders depicted in the 20 videos (10 for observers in the comparable truth baseline condition and 10 for observers in the small talk baseline condition) had performed the same undercover mission, as in Palena et al. (2017). However, there was a difference in the two conditions. Senders in the comparable truth baseline had to report the entire undercover mission (all senders answered the baseline question truthfully, whereas half told the truth and the other half lied when answering the target question). By contrast, senders in the small talk baseline condition answered non-threatening questions about themselves, as suggested by Inbau et al. (2013) and by Frank et al. (2006) as a baseline question (all senders answered truthfully to this question) but they described the second half of the mock crime they performed (what

happened after they met a confederate) when answering the target question (again, half of the sample answered truthfully whereas the other half lied). Caso et al. (2019) found that observers in the comparable truth condition reached higher total and lie accuracy rates than observers in the small talk condition. Summarising, the comparable truth baseline may have potential - particularly when it is related to speech content- but it should be further studied and developed (Vrij, 2016).

Ideally, before introducing interviewing techniques into the real world, professionals should be taught how to use them. A recent meta-analysis found that training people to detect deception results in higher accuracy levels, particularly when the training is based on verbal content rather than on nonverbal behaviour (Hauch, Sporer, Michael, & Meissner, 2016). Similarly, training in ACID (Colwell et al., 2009), SUE (Hartwig, Granhag, Strömwall, & Kronkvist, 2006) and CCA (Vrij, Leal, Mann, Vernham, & Brankaert, 2015) improved lie detection. Unfortunately, it is not always possible to train professionals in interviewing techniques, due to economic and time constraints. Therefore, there is the need to find alternative ways to improve veracity assessments when training is not available.

One possible way is to provide interviewers with a checklist that can help them in the veracity assessment process. Evans, Michael, Meissner, and Brandon (2013) have developed an eleven-items checklist starting from previous research on the most effective cues to truthfulness and deception. Their checklist included cues derived from the verbal veracity assessment tools Criteria Based Content Analysis and the Reality Monitoring, as well as from research on impression management theory (Köhnken, 1996). In their first study, Evans et al. (2013) found that three cues (admitted lack of memory, spontaneous corrections and rate of speech) were rated differently by interviewers when assessing truth tellers compared to when assessing liars (eight senders in total were showed). Additionally, an interaction effect emerged whereby the number of cues rated differently between truth and lie telling rose to nine in a 'high cognitive load' condition[28]. Evans' and colleagues' cognitive load

[28] "High cognitive load" refers to situations where the interviewee finds the interview more mentally taxing, i.e.: it requires higher cognitive resources. For example, in Evans et al. (2013) cognitive load was increased by asking to the interviewee to recall the event in a reverse chronological order (study 1).

manipulation concerned the order with which the statement was provided (forward chronological order for the low cognitive load condition vs. reverse chronological order for the high cognitive load condition). In their second study, the authors found that participants provided with a checklist containing information about details present in the story (e.g. spatial, temporal), memory related criteria (e.g. lack of memory) and others (e.g., cognitive load, rate of speech, etc.), obtained higher accuracy rates than participants who were not provided with it (again, eight videos where shown in total). This was a promising result given that the training on how to use the checklist only took a few minutes.

The novelty of the current study is the examination of the use of a checklist in the context of the baseline technique and the exploration of how the two different types of baseline (comparable truth vs. small talk) affect the effectiveness of the checklist. We developed a checklist similar to the one used by Evans et al. (2013) but included fewer cues as we eliminated some of the cues that were not effective in Evans et al.'s (2013) study. Of course, the shorter the checklist, the easier to teach it to practitioners and to use it.

In accordance with Caso et al. (2019), we predicted that observers who watched interviews that were conducted with a comparable truth approach will obtain higher accuracy rates in distinguishing truth tellers from liars than observers who watched interviews that were conducted with a small talk approach (Hypothesis 1).

We also predicted an interaction effect between type of baseline and veracity of the sender (Hypothesis 2) on the checklist scores. Participants in the comparable truth condition were expected to rate liars as more different from truth tellers than participants in the small talk condition.

Method

Participants

A list of participants interested in participating was obtained during university lectures. Participants were also informed that they would obtain one additional credit for a university course if they performed well in the task. All participants received that credit. Eventually, one hundred and twenty participants (27 male subjects and 93

female subjects) took part in the experiment. Their age ranged from 18 to 57 years, with an average age of M = 23.60 (SD = 5.75). Participant quantity was chosen to ensure at least 60 participants in each of the baseline conditions (see below).

Design

The study employed a 2 (Baseline: comparable truth vs. small talk) x 2 (Veracity: honest vs. liar) mixed design with Baseline as a between-subjects and Veracity as a within-subjects factor. Participants either saw interviews conducted with a comparable truth baseline or with a small talk baseline. In the comparable truth condition, the baseline section of the interview was similar to the target section in term of stakes, cognitive load, emotional involvement and content, see below. In the small talk condition, the baseline concerned personal information of the interviewees, whereas the target section concerned the same mock crime as that of the comparable truth condition, see below. In total, each observer judged ten interviewees, of which half (five) told the truth and half (five) lied when answering the target question. The dependent variables were the accuracy of observers in percentage (total accuracy, truth accuracy and lie accuracy) and the checklist ratings.

Checklist development

We derived our checklist (Appendix 1) from Evans et al. (2013) (Appendix 2). However, we wanted to obtain a shorter checklist, easy to use by observers. First, we merged spatial information (information about the location of objects: "*The car was on my left"*) and temporal information (information about the timeline of the event: *"He entered the room after me"*) into one criterion Contextual Embedding, similar to what happens in CBCA research (Vrij, 2008). Second, rather than just including the cue auditory details (Evans et al., 2013), we included the cue Sensory Details, derived from RM research. The RM approach focuses on all human senses (Johnson & Raye, 1981; Vrij, 2018). Therefore, eliminating all senses but auditory details from the checklist may reduce its efficacy.

Third, we entered the cue "Verifiable Details", as recent research has found that this is a diagnostic cue to distinguish truth tellers from liars (Nahari, Vrij, & Fisher, 2014a, 2014b). A verifiable detail is everything said by the sender that can be checked

by the investigator: Liars typically include fewer verifiable details in their stories than truth tellers. Fourth, as crimes involve actions and considering the literature about liars' strategies, who tend to withhold information (Hartwig, Granhag, Strömwall, & Doering, 2010), we entered the cue Actions into the checklist. Liars may be willing to conceal what they have done, which could result in liars reporting fewer action details than truth tellers.

All these cues where scored on three-point scales where 0 indicated that the criterion was absent, 1 that it was occasionally present, and 2 that it was frequently present. Furthermore, we added two cues to be rated on a continuum scale ranging from 0 to 4: Cognitive load and vagueness. Research has shown that lying is more mentally taxing than telling the truth (Vrij et al., 2017). In addition, liars often tend to withhold and manipulate information (DePaulo et al., 2003; Hartwig et al., 2010), which may result in a vaguer account (in particular, the quantity maxim by Grice, 1991).

The baseline approach is based on the idea that liars change their behaviour more than truth tellers without specifying the direction (Ewens et al., 2014; Palena et al., 2017; Vrij, 2016). However, since research has shown that cues to deception tend to follow a specific pattern (DePaulo et al., 2003) (e.g. liars reporting fewer details than truth tellers) it may be more fruitful to look at the direction of the differences between the two phases of the interview. For this reason, we scored the checklist as follows. For the first four cues (contextual, sensory, verifiable and action details) we used the formula *ratings in the target phase - ratings in the baseline phase*. Therefore, a negative score should be indicative of deception as negative scores imply a *decrease* in the amount of details in the target phase compared to the baseline. Cognitive load and vagueness tend to *increase* during lying. Therefore, to continue with the principle that a negative score indicates deception, we used the formula *ratings in the baseline phase – ratings in the target phase.*

Eventually, we summed up the scores for each item of the checklist, and the score could vary between -16 (highest possible decrease in details and increase in vagueness and cognitive load when lying) to 0 (no difference at all) to +16 (highest possible increase in details and decrease in vagueness and cognitive load when lying).

Stimulus material

We randomly selected twenty videotaped stimuli from a recent study (reference not reported for blind review). Of these, ten were used for the comparable truth condition and ten videos depicted senders who were interviewed via a small talk baseline.

All participants performed the following set of tasks. In first instance, they were given a password to log into a laptop. They then read a word document which instructed them to look for a CD-ROM, which was placed inside a backpack. Once the participants played the CD-ROM, a video depicting a man appeared. The man instructed the participants to look for a key inside the same backpack from where they collected the CD-ROM. Such a key could open a safe-deposit box which contained additional information. Such information instructed the participants to send an email to a specific email address and, once done, to wait for a person to come. Such a person was a confederate, who delivered a newspaper to the participants, in which additional written instruction were hidden. Such instruction informed the participants that they had to go to another room (adjacent to that where the laptop was located) and to look for a USB stick hidden behind a coat-hook. They had to take it and swap it with another USB stick which was hidden inside a book located in a wardrobe. Further, the participants had to leave the newspaper near the book and keep with them the new USB (the one initially placed inside the book) for the remaining of the experiment.

Everything that happened before the meeting between the participants and the confederate was part of the "Baseline section" (for the Comparable Truth Baseline only, see below). Everything that happened after the meeting was instead part of the "Target section" (the event under investigation).

All participants performed the full set of tasks (both those for the baseline section and for the target section), but there was an important difference between the Comparable Truth Baseline and the Small Talk Baseline conditions. Since the Comparable Truth Baseline approach implies that the context, stakes, emotional and cognitive involvement of the baseline and the target questioning should be similar (Vrij, 2016), participants in the Comparable Truth Baseline were asked to report what they had done before they met the confederate as a baseline question. Conversely, the Small

Talk Baseline build on the idea that the baseline questioning can be conducted asking personal, non-threatening questions to the interviewee (Ewens et al., 2014; Frank et al., 2006). For this reason, the participants in the Small Talk baseline condition were asked personal information (e.g.: last year spent as a student). All participants answered to the baseline question truthfully.

Concerning the target section, all participants reported what they had done after they met the confederate. Yet, half of them were asked to tell the truth about it, whereas the other half were asked to lie, by providing a credible story which included at least 4-5 events or actions. The order of lie-truth telling, within each baseline condition, was counterbalanced. Furthermore, having a ground truth available, it was possible to exclude any interviewee who did not follow the veracity instructions (all the 20 senders followed the instructions).

To increase participants' motivation, all participants were told that they would receive two additional points for a university examination only if they were believed by the interviewer. However, all of them received the bonus.

The comparable truth baseline answers duration was of M = 64.00 seconds (SD = 17.71). The small talk baseline answers duration was of M = 91.11 seconds (SD = 53.11). The target answers duration for participants in the comparable truth baseline was of M = 58.90 seconds (SD = 17.13). The target answers duration for participants in the small talk baseline was of M = 70.00 seconds (SD = 36.90). The duration of the baseline answer and the target answer did not differ for participants in the comparable truth baseline condition, $t(9) = .619$, $p = .55$, nor it did for those in the small talk condition, $t(8) = .858$, $p = .41$. It is worthwhile to note here that it is not the length of the baseline and the target phases of the interview to make them comparable or not. It is their content.

The baseline and the target answer for participants in the comparable truth baseline were comparable as: a) both concerned discussing a mock crime (in particular, two sections of the *same* mock crime), b) both are embedded in the same context (a mock crime), c) the two sections included similar sets of actions.

Additionally, the two sections of the interview for the comparable truth baseline condition: a) are in accordance with Vrij's (2016) definition of a comparable truth baseline, b) are similar to Vrij's and Mann's (2001) study where a real murderer discussed what he had done in the morning (comparable truth baseline section) and after the morning (target section) of the day he committed the crime, c) have been previously used successfully (Caso et al., 2019; Palena et al., 2018).

By contrast, participants in the small talk baseline had to report personal information when answering the baseline question. Being this different in content and context to the target question (participants discuss two completely unrelated topics: personal information in one case and a mock crime in another case) they do not respect Vrij's (2016) definition of a comparable truth baseline and, rather, respect the definition of a small talk baseline (Ewens et al., 2014; Frank et al., 2006; Palena et al., 2018).

Procedure

Upon arrival, each participant was welcomed and briefed. S/he was then asked to sign the consent form, which indicated what the experiment was about, and that the participant was free to leave the experiment at any point. Participants were also informed that there were more answer sheets (20) than the real number of senders they would judge (ten) and that the number of truth tellers and liars may have been unequally balanced. This to avoid that the participant would try to balance the number of "truth" and "lie" answers in the first ten answer sheets.

Participants were then introduced to the checklist. This included a ten minutes presentation with background information of the cues, a sheet of paper with the description of each cue to be rated and two practice videos. Participants were free to look at the description during the experiment. Participant's answers to the two practice videos were compared to the answers previously given by an experienced coder. This was done to verify that the participant understood how to use the checklist. If any difference arose, a discussion with the experimenter took place to solve disagreements. After this, when the experimenter felt confident that the participant understood how to use the checklist, the experiment started. Participants first completed the checklist (Appendix 1) twice for each interviewee: Once after the baseline section and once after the target section. After this they answered the

following question: *"Do you believe the interviewee was: "Telling the truth"* or *"Lying"*.

In each condition, each video was presented as follows. A text indicating "Baseline" appeared on the screen and lasted for three seconds. After that, the first sender appeared, and s/he answered the baseline question. Once finished, a text indicating "evaluation" appeared and lasted for fifteen seconds. Here, participants had to complete the checklist related to the baseline question. Once the fifteen seconds had expired, a high frequency sound appeared, together with a text indicating "target" to warn the participant that the video was going to continue. Then, the sender started answering the target question and, when s/he finished, a text indicating "evaluation" appeared again. At this point, the observer had to complete the checklist related to the target question in a fifteen seconds time-window. Then, another high frequency sound signalled that it was now the time to make the decision about the sender's (dis)honesty. To complete the veracity assessment the observers were given thirty seconds. Participants were informed that to make this veracity decision, they had to evaluate how differently they rated the two phases of the interview, bearing in mind that larger differences are more likely to indicate lying than small differences. When the thirty seconds expired, a high frequency sound together with a text indicating "baseline" (lasting three seconds) appeared, warning the participant that the following interview was going to start. The pattern was repeated until the end of the tenth video. After the last video, the experiment was finished, and the participants were thanked and debriefed.

Most importantly, participants were instructed to compare baseline and target answers within each sender. That is, they were instructed to pay attention to difference showed by the same subjects, and not differences showed across subjects.

Results

Accuracy

Three different *t*-tests with Baseline (comparable truth vs. small talk) as the factor and total, truth and lie accuracy as the dependent variable showed that there was no significant effect (Table 1). Furthermore, signal detection analyses showed that

participants in the comparable truth condition (M = .28, SD = .97, 95% C. I. [.02, .53]) did not discriminate truth tellers from liars to a higher extent than participants in the small talk condition (M = .23, SD = .87, 95% C. I. [.00, .45]), $t(117) = .273$, $p = .78$, $d = .05$. Therefore, Hypothesis 1 was not supported. Also, participants in the comparable truth condition did not differ from those in the small talk condition in response bias, $t(117) = .412$, $p = .681$, $d = .08$.

Two one-sample *t*-tests showed that participants' *d*' values differed from zero for both the comparable truth $t(57) = 2.181$, $p = .03$, $d = .28$, and small talk conditions, $t(60) = 2.083$, $p = .04$, $d = .26$. Since a *d*' of 0 means that observers are not able to discriminate the stimuli, these one-sample *t*-tests indicate that participants in both baseline conditions were able to discriminate truth tellers from liars.

The β values did not differ from 1 for participants in the comparable truth condition, (M = 1.09, SD = .42), $t(57) = 1.700$, $p = .09$, $d = .21$, nor for those in the small talk condition, (M = 1.06, SD = .37), $t(60) = 1.330$, $p = .18$, $d = .17$. Since a β of 1 indicates no response bias, these results indicate that it was not possible to conclude that participants showed a clear response bias. Nonetheless, values were greater than 1 for all participants, indicating a tendency toward a truth bias.

enter Table 1 here about here

Checklist scorings

A mixed-design ANOVA was carried out with Baseline (comparable truth vs. small talk) as the between-subject factor, Veracity (honest vs. liar) as the within-subjects factor, and the checklist score as the dependent variable. The main effects for Baseline, $F(1, 90) = 50.083$, $p < .001$, $\eta_p^2 = .36$, and Veracity, $F(1, 90) = 71.829$, $p < .001$, $\eta_p^2 = .44$ were both significant. Participants in the comparable truth condition (Est. M = -.99, SE = .14, C. I. [-1.27, -.70]) rated senders as more deceptive than participants in the small talk condition (Est. M = .48, SE = .15, C. I. [.18, .78]). Additionally, observers rated liars (M = -.92, SD = 1.44, C. I. [-1.22, -.62]) as more deceptive than truth tellers (M = .35, SD = 1.41, C. I. [.05, .64]). Furthermore, the Baseline X Veracity interaction, $F(1, 90) = 1.831$, $p = .18$, $\eta_p^2 = .02$ was not significant. Since the interaction was not significant, we did not find support for

Hypothesis 2. Although the interaction effect was not significant, one-sample *t*-tests examining difference scores for liars and truth tellers separately for each of the two baseline conditions showed interesting findings. Means and standard deviations are reported in Table 2.

For the comparable truth condition, both truth tellers, $t(47) = -3.339$, $p = .002$, $d = -.48$ [-.78, -.18], and liars, $t(47) = -9.136$, $p < .001$, $d = -1.32$ [-1.70, -.92], differed significantly from zero with the effect size being larger for liars than for truth tellers. This larger difference for liars is consistent with the baseline assumptions. For the small talk condition, truth tellers differed significantly from 0, $t(43) = 6.052$, $p < .001$, $d = .91$ [.55, 1.25], whereas liars did not, $t(43) = -1.201$, $p = .236$, $d = -.18$ [-.47, .12]. This means that truth tellers appeared to behave differently between the two phases of the interview, whereas liars did not. This is inconsistent with the baseline assumptions.

enter Table 2 about here

Discussion

In this experiment, we tested two hypotheses building on previous research regarding the baseline technique (Caso et al., 2019; Ewens et al., 2014; Palena et al., 2017; Palena et al., 2018) and the use of a checklist for lie detection purposes (Akehurst et al., 2017; Evans et al., 2013).

We were not able to support Hypothesis 1, as no difference appeared between the two baseline conditions in terms of lie detection accuracy. Participants obtained accuracy rates in the 50-58% range, typical for lie detection research (Bond Jr. & DePaulo, 2006). One explanation for the absence of an effect is that participants failed to comply with the request to base their veracity judgements on the checklist and decided according to their preferences (such as gut feeling). Future studies should therefore explore how checklist scores predict observers' decisions about interviewee's (dis)honesty, as well as focus on ways to make the checklist more effective (e.g.: via a longer training).

The possibility of observers failing to base their veracity judgments on the checklist scores becomes plausible when we consider the checklist results. For the comparable

truth condition, the checklist scores revealed that liars changed their behaviour between the baseline phase and target phase to a larger extent than truth tellers, which is consistent with the baseline assumptions. Furthermore, liars' changes went in the predicted direction. In contrast, for the small talk condition truth tellers changed their behaviour from baseline to target phase more than liars, which is inconsistent with the baseline assumptions. However, since in the comparable truth condition, truth tellers also changed their behaviour between the two phases (not just liars), employing the comparable truth baseline technique remains difficult, as observers need to consider the relative change in behaviour between the baseline and target phases rather than an absolute change (Palena et al., 2018; Vrij, 2016).

The traditional small talk baseline approach states that only liars will show differences between the baseline phase and the other section of the interview (Ewens et al., 2014; Frank et al., 2006). We found no support for this assumption, as truth tellers rather than liars changed their behaviour in that baseline condition. This result is not surprising. Our checklist focused mainly on details, cognitive load, and vagueness. Senders in the small talk baseline condition discussed their last year as students (or worker), which in all likelihood did not contain many verifiable details, contextual details or actions and which sounded a bit vague as people usually find it difficult to understand how many details are expected from them when they are asked an open-ended question to 'tell it all' (Leal, Vrij, Warmelink, Vernham, & Fisher, 2015). In contrast, the mock crime they reported during the target questioning was rich in details (especially spatial) and truth tellers in particular may have been motivated to report many details. Consequently, the difference in the amount of details between the two phases became more distinct for truth tellers than for liars.

The traditional small talk baseline approach further refers to changes in (non)verbal behaviour without specifying the specific (non)verbal behaviours involved and the direction of these changes (Ewens et al., 2014; Frank et al., 2006). This is problematic. The baseline approach should be developed so that specific cues are examined, and predictions are made regarding how these cues will change between baseline and target phases. This should be based on (non)verbal cues to deception research. For example, rather than observing whether changes occur in the number of details someone reports when describing the baseline and target events,

someone should examine whether a decrease in details occurs in reporting the target event compared to the baseline event, as research has shown that liars typically report fewer details than truth tellers (DePaulo et al., 2003; Palena et al., 2017; Vrij, 2008), as well as that when people tell a mixture of truths and lies, they tend to report fewer details in the section of the statement about which they lied than in the section of the statement about which they told the truth (Palena, Caso, & Vrij, 2019).

In sum, the comparable truth condition revealed more pronounced differences in responses between truth tellers and liars than the small talk baseline condition, but observers were equally accurate in identifying truth tellers and liars in both baseline conditions. There may be several reasons for this. First, the differences between truth tellers and liars may have not been substantial enough in the comparable truth condition to become more accurate, or, second, observers may have paid attention to irrelevant cues. Third, the instructions about how to use the checklist were either too short or not easy to understand and, fourth, interactions between senders' and observers' characteristics may have affected the outcome (Caso et al., 2018). This makes it important to explore what happens between attending a stimulus (sender) and the output decision and how to guide observers in this process.

References

Akehurst, L., Easton, S., Fuller, E., Drane, G., Kuzmin, K., & Litchfield, S. (2017). An evaluation of a new tool to aid judgements of credibility in the medico-legal setting. *Legal and Criminological Psychology, 22*(1), 22-46. doi:10.1111/lcrp.12079

Bond Jr., C. F., & DePaulo, B. M. (2006). Accuracy of deception judgments. *Personality and Social Psychology Review, 10*(3), 214-234. doi:10.1207/s15327957pspr1003_2.

Bond Jr., C. F., & DePaulo, B. M. (2008). Individual differences in judging deception: Accuracy and bias. *Psychological bulletin, 134*(4), 477. doi:10.1037/0033-2909.134.4.477.

Caso, L., Maricchiolo, F., Bonaiuto, M., Vrij, A., & Mann, S. (2006). The impact of deception and suspicion on different hand movements. *Journal of Nonverbal behavior, 30*(1), 1-19. doi:10.1007/s10919-005-0001-z.

Caso, L., Maricchiolo, F., Livi, S., Vrij, A., & Palena, N. (2018). Factors affecting Observers' Accuracy when Assessing Credibility: The Effect of the Interaction between Media, Senders' Competence and Veracity. *The Spanish Journal of Psychology, 21*, 1-10. doi:10.1017/sjp.2018.54.

Caso, L., Palena, N., Vrij, A., & Gnisci, A. (2019). Observers' performance at evaluating truthfulness when provided with Comparable Truth or Small Talk Baselines. *Psychiatry, Psychology and Law*. doi:10.1080/13218719.2018.1553471.

Colwell, K., Hiscock-Anisman, C., & Fede, J. (2013). Assessment Criteria Indicative of Deception: An Example of the New Paradigm of Differential Recall Enhancement. In B. S. Cooper, D. Griesel, & M. Ternes (Eds.), *Applied Issues in Investigative Interviewing, Eyewitness Memory, and Credibility Assessment* (pp. 259-291). New York, NY: Springer New York.

Colwell, K., Hiscock-Anisman, C., Memon, A., Colwell, L. H., Taylor, L., & Woods, D. (2009). Training in Assessment Criteria Indicative of Deception to Improve Credibility Judgments. *Journal of Forensic Psychology Practice, 9*(3), 199-207. doi:10.1080/15228930902810078

Colwell, K., Hiscock-Anisman, C. K., Memon, A., Taylor, L., & Prewett, J. (2007). Assessment Criteria Indicative of Deception (ACID): An integrated system of investigative interviewing and detecting deception. *Journal of Investigative Psychology and Offender Profiling, 4*(3), 167-180.

DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological bulletin, 129*(1), 74-118. doi:10.1037/0033-2909.129.1.74.

Evans, J. R., Michael, S. W., Meissner, C. A., & Brandon, S. E. (2013). Validating a new assessment method for deception detection: Introducing a Psychologically Based Credibility Assessment Tool. *Journal of Applied Research in Memory and Cognition, 2*(1), 33-41. doi:https://doi.org/10.1016/j.jarmac.2013.02.002

Ewens, S., Vrij, A., Jang, M., & Jo, E. (2014). Drop the small talk when establishing baseline behaviour in interviews. *Journal of Investigative Psychology and Offender Profiling, 11*(3), 244-252. doi:doi.org/10.1002/jip.1414.

Frank, M. G., Yarbrough, J. D., & Ekman, P. (2006). Investigative interviewing and the detection of deception. In T. Williamson (Ed.), *Investigative interviewing: Rights, research and regulation* (pp. 229-255). Cullompton, Devon: Willian.

Gnisci, A., Caso, L., & Vrij, A. (2010). Have you made up your story? The effect of suspicion and liars' strategies on reality monitoring. *Applied Cognitive Psychology, 24*(6), 762-773. doi:doi:10.1002/acp.1584

Granhag, P. A., & Hartwig, M. (2015). The strategic use of evidence technique: A conceptual overview. In P. A. Granhag, V. Aldert, & B. Verschuere (Eds.), *Detecting Deception: Current Challenges and Cognitive Approaches* (pp. 231-251). Chichester, UK: Wiley.

Grice, H. P. (1991). *Studies in the Way of Words*: Harvard University Press.

Hartwig, M., & Bond Jr, C. F. (2011). Why do lie-catchers fail? A lens model meta-analysis of human lie judgments. *Psychological bulletin, 137*(4), 643. doi:10.1037/a0023589

Hartwig, M., Granhag, P. A., Strömwall, L. A., & Doering, N. (2010). Impression and information management: On the strategic self-regulation of innocent and guilty suspects. *The Open Criminology Journal, 3*(1), 10-16. doi:10.2174/1874917801003010010

Hartwig, M., Granhag, P. A., Strömwall, L. A., & Kronkvist, O. (2006). Strategic use of evidence during police interviews: When training to detect deception works. *Law and human behavior, 30*(5), 603-619.

Hauch, V., Sporer, S. L., Michael, S. W., & Meissner, C. A. (2016). Does training improve the detection of deception? A meta-analysis. *Communication Research, 43*(3), 283-343.

Inbau, F. E., Reid, J. E., Buckley, J. P., & Jayne, B. C. (2013). *Criminal interrogation and confessions (5th ed.)*. Burlington, MA: Jones & Bartlett Learning.

Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological Review, 88*(1), 67-85. doi:10.1037/0033-295X.88.1.67

Köhnken, G. (1996). Social psychology and the law. In G. R. Semin & K. Fiedler (Eds.), *Applied social psychology* (pp. 257-282). Thousand Oaks, CA, USA: Sage Publications, Inc.

Leal, S., Vrij, A., Warmelink, L., Vernham, Z., & Fisher, R. P. (2015). You cannot hide your telephone lies: Providing a model statement as an aid to detect deception in insurance telephone calls. *Legal and Criminological Psychology, 20*(1), 129-146. doi:doi:10.1111/lcrp.12017

Nahari, G., Vrij, A., & Fisher, R. P. (2014a). Exploiting liars' verbal strategies by examining the verifiability of details. *Legal and Criminological Psychology, 19*(2), 227-239. doi:10.1111/j.2044-8333.2012.02069.x.

Nahari, G., Vrij, A., & Fisher, R. P. (2014b). The Verifiability Approach: Countermeasures Facilitate its Ability to Discriminate Between Truths and Lies. *Applied Cognitive Psychology, 28*(1), 122-128. doi:10.1002/acp.2974.

Palena, N., Caso, L., Carlotto, G., De Mizio, L., & Marciali, M. (2017). Efficacia della tecnica di baselining nella valutazione della credibilità (Efficacy of the baseline

technique when assessing credibility). *Giornale italiano di psicologia, 44*(4), 905-916. doi:10.1421/88773

Palena, N., Caso, L., & Vrij, A. (2019). Detecting Lies via a Theme-Selection Strategy. *Frontiers in Psychology, 9*(2775). doi:10.3389/fpsyg.2018.02775

Palena, N., Caso, L., Vrij, A., & Orthey, R. (2018). Detecting deception through small talk and comparable truth baselines. *Journal of Investigative Psychology and Offender Profiling, 15*(2), 124-132. doi:doi:10.1002/jip.1495

Vrij, A. (2008). *Detecting lies and deceit: Pitfalls and opportunities, second edition*. Chichester: John Wiley and Sons.

Vrij, A. (2014). Interviewing to detect deception. *European Psychologist, 19*, 184-194. doi:10.1027/1016-9040/a000201.

Vrij, A. (2016). Baselining as a lie detection method. *Applied Cognitive Psychology, 30*(6), 1112-1119. doi:10.1002/acp.3288.

Vrij, A. (2018). Verbal Lie Detection Tools From an Applied Perspective. In J. P. Rosenfeld (Ed.), *Detecting Concealed Information and Deception: Recent Developments* (pp. 297-327). San Diego, USA: Academic Press.

Vrij, A., & Fisher, R. P. (2016). Which lie detection tools are ready for use in the criminal justice system? *Journal of Applied Research in Memory and Cognition, 5*(3), 302-307.

Vrij, A., Fisher, R. P., & Blank, H. (2017). A cognitive approach to lie detection: A meta-analysis. *Legal and Criminological Psychology, 22*(1), 1-21. doi:doi:10.1111/lcrp.12088

Vrij, A., & Granhag, P. A. (2012). Eliciting cues to deception and truth: What matters are the questions asked. *Journal of Applied Research in Memory and Cognition, 1*(2), 110-117. doi:10.1016/j.jarmac.2012.02.004.

Vrij, A., Leal, S., Mann, S., Vernham, Z., & Brankaert, F. (2015). Translating theory into practice: Evaluating a cognitive lie detection training workshop. *Journal of*

Applied Research in Memory and Cognition, 4(2), 110-120. doi:https://doi.org/10.1016/j.jarmac.2015.02.002

Vrij, A., & Mann, S. (2001). Telling and detecting lies in a high-stake situation: the case of a convicted murderer. *Applied Cognitive Psychology, 15*(2), 187-203. doi:10.1002/1099-0720(200103/04)15:2<187::aid-acp696>3.0.co;2-a

Table 1. *Accuracy rates according to the baseline condition*

	Comparable Truth	Small Talk			
	M(SD)	*M(SD)*	*t*	*p*	*Cohen's d*
Total accuracy	55.18 (17.89)	54.31 (16.37)	.277	.78	.05
Truth accuracy	57.88 (24.76)	58.44 (24.08)	-.126	.90	-.02
Lie accuracy	52.37 (24.40)	50.21 (24.69)	.481	.63	.09

Table 2. *Difference scores between the Baseline and the Target phase of the interview.*

	Comparable truth *M (SD)*	Small Talk *M (SD)*
Truth tellers	-.45 (.93)	1.22 (1.34)
Liars	-1.52 (1.15)	-.26 (1.45)

Appendix 1: *checklist used by the observers*

BASELINE (The interviewee is telling the truth)

	Absent	**Occasional**	**Frequent**
Contextual details	□	□	□
Sensory details	□	□	□
Verfiable details	□	□	□
Actions	□	□	□

Cognitive load

Very low **Uncertain** **Very high**

|----------|----------|----------|----------|----------|----------|----------|----------|

Vagueness

Not at all **Uncertain** **Very vague**

|----------|----------|----------|----------|----------|----------|----------|----------|

TARGET

(The interviewee can either be telling the truth or lying)

	Absent	**Occasional**	**Frequent**
Contextual details	□	□	□
Sensory details	□	□	□
Verifiable details	□	□	□
Actions	□	□	□

Cognitive load

Very low **Uncertain** **Very high**

|----------|----------|----------|----------|----------|----------|----------|----------|

Vagueness

Not at all **Uncertain** **Very vague**

|----------|----------|----------|----------|----------|----------|----------|----------|

DECISION

Do you believe the interviewee was

□ Telling the truth □ Lying

Appendix 2: *Evans' and colleagues' (2013) PBCAT Checklist*

Sensory details	**Not present**	**Occasional**	**Frequent**
Auditory details	□	□	□
Spatial details	□	□	□
Temporal details	□	□	□

	Not present	**Present (1 or 2 times)**	**Present (3 or more times)**
Admitted lack of memory	□	□	□
Spontaneous corrections	□	□	□

Overall quantity of details/Talking time	Very few details/Very vague	Unsure	Numerous details/complete
	|---------|---------|---------|---------|---------|---------|---------|---------|		
Does it make sense? (contradictions? Plausible?)	Not at all	Unsure	Completely
	|---------|---------|---------|---------|---------|---------|---------|---------|		

Thought hard

Did not think hard Unsure Thought extremely hard

|---------|---------|---------|---------|---------|---------|---------|---------|

Nervousness (tense? Fidgeting?)

Extremely relaxed/Comfortable Unsure Extremely tense/nervous

|---------|---------|---------|---------|---------|---------|---------|---------|

Negativity/Complaints

Extremely positive Neutral/Unsure Extremely negative

|---------|---------|---------|---------|---------|---------|---------|---------|

Rate of speech

Extremely fast Unsure Estremely slow

|---------|---------|---------|---------|---------|---------|---------|---------|

Appendix 7: Palena, N., Caso, L., & Vrij, A. (2019). Detecting Lies via a Theme-Selection Strategy. *Frontiers in Psychology, 9*(2775). doi:10.3389/fpsyg.2018.02775. [PA6].

Abstract

Most of deception research has focused on past events that were either completely truthful or a complete fabrication. However, people often tell a mixture of truths and lies. This could enable investigators to make within-subjects comparisons between different themes discussed in one interview, which we examined in the current experiment. Seventy-three participants took part in the experiment and were asked to either tell the truth about two themes, or to tell the truth about one theme and lie about the second theme in a HUMINT setting. Results showed that examining the differences in the amount of detail provided by the interviewees for each theme- obtained through a Theme-Selection strategy (a within-subjects measure)- yielded stronger results than examining differences between truth tellers and liars based on the entire interview without accounting for themes (between-subjects measure). The present study therefore highlighted the effectiveness of within-subjects measurements to both discriminate truth tellers from liars and to discover which section of a statement is false.

Keywords: Theme-Selection strategy, within-subjects comparisons, lie detection, HUMINT interviewing, Investigative Interviewing, Strategic Interviewing.

Introduction

Research has shown that cues to deception are faint and unreliable (DePaulo et al., 2003) and that people's ability to detect lies is low (Bond & DePaulo, 2006, 2008). Consequently, scholars switched their attention to the development of interviewing techniques aimed at enhancing the amount of information revealed by the interviewee and to elicit cues to deception (Vrij, 2014; Vrij & Granhag, 2012). Most of them are based on memory research and cognition and have shown potential in discriminating truth tellers and liars (Ost, Scoboria, Grant, & Pankhurst, 2015; Rosenfeld, 2018; Vrij, 2008, 2015, 2018).

One of the new developments is to focus on differences within an individual rather than between individuals. So-called within-subjects comparisons can reduce problems caused by interpersonal differences (Nahari & Pazuelo, 2015; Vrij, 2016) and are preferred by both practitioners and scholars (Nahari, 2016; Nahari & Vrij, 2014, 2015; Vrij, 2016). Comparisons within an individual can be made in different ways (for a detailed and recent review of within-subjects comparisons, see Vrij, 2016, 2018). An example is to compare specific variables within an interviewee statement. For example, research has shown that the proportion of verifiable details compared to non-verifiable details (Nahari, Vrij, & Fisher, 2014) was higher for truth tellers than for liars. Similarly, truth tellers reported a higher proportion of complications (operationalized as complications / (complications + common knowledge details + self-handicapping strategies) than liars (Vrij, Leal, Jupe, & Harvey, 2018).

Another example is the use of the reverse order technique, where the interviewee is firstly asked for a free recall and then to report the same event starting from the end and going back towards the beginning (Fisher & Geiselman, 1992; Fisher, Schreiber Compo, Rivard, & Hirn, 2014). Research has shown that truth tellers report more reminiscences than liars when asked to recall the event in reverse order (see Vrij, 2016).

We examined another within-subjects measure, described below, based on the finding that interviewees often tell a mixture of truths and lies (Leins, Fisher, & Ross, 2013; Maguire & John, 1995). That is, interviewees are honest in one section of their statement in which they describe one event (or topic) but lie in another section of the statement where they discuss a second event (or topic) (Palena, Caso, Vrij, & Orthey, 2018). Discussing various topics in one interview is not uncommon in intelligence interviews (Deeb et al., 2017).

In a recent study introducing an intelligence type interview setting, truth tellers honestly reported two events, whereas liars lied about one event and told the truth about the other event (Deeb et al., 2017). All participants were interviewed twice, always with a free recall in the first interview, and with either a free recall or a set of specific questions in the second interview. It was found that liars' accounts included

less repetitions than truth tellers' accounts for both events, particularly when the second interview was conducted via specific questioning.

The present study was similar to Deeb et al. (2017) in that our lying participants also told the truth about one event (theme 1, noncritical event) but lied about another event (theme 2, the critical event). However, our study differs from Deeb et al. (2017) in at least two ways. First, we only interviewed the participants once. Second, rather than focusing on consistency, we focused on the amount of information revealed by the interviewees., based on research showing that liars are typically less detailed and forthcoming than truth tellers (DePaulo et al., 2003; Hartwig, Granhag, & Luke, 2014; Vrij, 2008),

We postulate that differences between truth tellers and liars should become more evident when the interviewer examines differences between the specific themes (Theme-Selection Strategy) than when the interviewer considers the statement as a whole without accounting for specific themes. The differences in detail provided between the themes within each individual's statement can then be used to decide whether someone is lying, and which part of the statement is the lie. As such, the comparison is not between truth tellers' and liars' entire statement, but between the interviewee's answers regarding the different themes of the interview. It thus becomes a within-subjects comparison. We predict that truth tellers, who will tell the truth about both themes, will show no difference in the amount of information provided when talking about the two themes, whereas liars, who tell the truth about the noncritical theme but lie about the critical theme, will report more information when talking about the noncritical theme (truth) than when talking about the critical theme (lie) (Hypothesis 1). We further predict that comparing interviewees' answers to the two themes (within-subjects measure) is more efficient for lie detection purposes than comparing truth tellers' and liars' answers taken as a whole (without accounting for themes, between-subjects measure) (Hypothesis 2). The reason for Hypothesis 2 is two-fold. First, within-subjects measures are typically more diagnostic than between-subjects measures (Vrij, 2016). Second, only the within-subjects measure is a true comparison between truths and lies, the between-subjects measure is a comparison between a total truth (truth tellers) and a mixture of truths and lies (liars).

Materials and Methods

Participants

All participants were university students. An announcement was made at the beginning of lectures and a list of names and email addresses was obtained. The students were informed that in case of a convincing performance during the interview they would be offered one additional credit for their exam (all participants received the credit). Seventy-three participants took part in the experiment (61 females and 12 males). Age of this sample ranged from 20 to 45 years, $M = 22.06$ ($SD = 2.95$), median = 22. However, data screening showed that five participants were outliers[29] in terms of details reported. The new sample, on which all analyses were conducted, therefore consisted of 68 participants (56 females and 12 males). The age ranged from 20 to 25, $M = 21.71$ ($SD = .99$), median = 21.50.

Design

We employed a 2 (Theme: noncritical vs. critical, within-subjects) by 2 (Veracity: truth telling vs. lie, between-subjects) mixed design. For the factor Theme, the participant had to report information about the structure and activities of the criminal organization (noncritical theme) and information about the hideout of the boss, his routines etc. (critical theme). For the factor Veracity, the participants either told the truth or lied about the critical theme (all participants told the truth about the noncritical theme). Additionally, since the amount information remembered (see below) can influence the amount of revealed information, the number of pieces of information for the two themes remembered by the participants were entered as covariates. The amount of revealed information was the dependent variable.

Procedure

Participants were informed that they would participate in an experiment mirroring a HUMINT interview. Upon arrival, each participant was welcomed and asked to read and sign the consent form if s/he decided to participate. S/he was then told that s/he

[29] Usually, any participant whose score is beyond the third quartile plus 1.5 times the interquartile range (IQR) or any participant whose score is below the first quartile minus 1.5 times the IQR is labelled as an outlier. However, the 1.5 by IQR formula has been criticised and a 2.2 multiplier is suggested instead (Hoaglin & Iglewicz, 1987). Therefore, to calculate our outliers we used the second formula.

had to play the role of a secret agent whose agency was trying to dismantle a criminal organization. All participants were informed that there was a spy working against their agency, whose goal was to protect the criminal organization. However, justifications for the following interview differed between the conditions (as in Deeb et al., 2017).

Truth tellers were informed that the interviewer could be trusted. They were asked to report all information honestly so that the interviewer would be fully informed about the interviewee's experiences and could conclude that the interviewee was not hiding anything. Liars were told that there was a risk that the interviewer was a spy which they needed to fool. Therefore, they were asked to adhere to the following instructions: To make an honest impression on the interviewer they had to report honestly everything relating to the structure of the criminal organization, its activities and components (noncritical section, 18 pieces of information, the truth). However, they were asked to lie about the boss' hideout, activities and routines (critical event, 18 pieces of information, the lie).

The assignment to the Veracity condition was alternated, meaning that the first participant was assigned to the honest condition, the second to the lying condition, the third to the honest condition and so on.

The experimenter then gave each participant (both truth tellers and liars) a file containing all the information about the criminal organization that the agency possessed. The participant was asked to study it in detail and to remember it for the following interview. The participant was then left alone to study the file and was asked to inform the experimenter when s/he had memorized the file information.

After fifteen minutes, the experimenter returned to the room and made sure that the participant understood the role and instructions. The participant was then asked to complete a memory-check questionnaire, where open ended question regarding each of the 36 pieces of information where asked (*e.g.: "At what time does the boss leaves his hideout?"*). After that, s/he was given ten minutes to prepare for the interview. Then, the participant was informed that the interview would start and was reminded that s/he would receive the additional study credit only when performing well during the interview. Eventually, all participants received the credit. All interviews were

video-recorded. When the interview was finished, the participant was told that the experiment was concluded, and s/he was asked to complete another memory-check. This contained two additional questions compared to the first memory-check. First, the participant was asked how motivated s/he was to convince the interview that s/he was telling the truth on a 7-point Likert scale ranging from one (not at all) to seven (totally). Second, the participant was asked whether s/he believed that s/he appeared credible to the interviewer on a 7-point Likert scale ranging from one (not at all) to seven (totally). Eventually, the participant was thanked, debriefed about the aims of the experiment, and told that s/he received the additional study credit.

Interview

Three people acted as interviewers, and each of them interviewed about one-third of the sample. All interviewers were blind to the study hypotheses and experimental conditions. The interviewers were trained and carried out simulated interviews before interviewing actual participants, similar to Oleszkiewicz, Granhag, and Kleinman (2014). This to make sure that: a) they followed the structure of the protocol; b) would not improvise or make changes during the interview; c) and kept a constant demeanor during the interviews (Mann et al., 2013). The interviewer started introducing herself, then asked the participant to briefly tell something about themselves, such as their hobbies and interest. This question was asked to put the interviewee at ease and to avoid that the first moment of the interview were influenced by the context (such as the presence of the camera). The interviewer then asked the following free recall question: *"As you know, our agency is investigating a criminal organization led by the Passatante clan. Tell me everything you know about this criminal organization in as much detail as possible".* This was followed by the question: *"Is there anything you would like to add*? The interviewer then asked two follow up questions to elaborate on the two themes. *"Ok, now tell me everything you know about the structure of the organization, such as its components, roles, and activities"* (non-critical theme) and *"Ok, now tell me everything you know about the Boss of the criminal organization"* (critical theme).

The order of the two follow-up questions was counterbalanced. Additionally, the second follow-up question was followed by another open-ended question: *"Is there*

anything you would like to add?". After this, the participant was thanked, and the interview ended.

Coding

First, all video interviews were transcribed. Two experienced coders, both blind to the experimental conditions and the aims of the study, coded 22 (about 30%) of the interviews for the presence of information regarding the noncritical and the critical themes revealed by the interviewee throughout the interview. Each piece of information was counted only once. The coding took place using a checklist that included all the 36 pieces of information provided to the interviewee, similar to Oleszkiewicz et al. (2014). For both the noncritical and critical themes, the scores on this checklist could range from 0 to 18. The total score accounting for both themes together could thus range from 0 to 36. Yet, liars may opt for the strategy to be as detailed as possible to appear credible. Therefore, they could report pieces of information, not present in the story they were initially given. For this reason, the coders also counted the number of pieces of information not initially given.

We calculated inter-rater reliability on 30% of the transcripts[30] using the two-way random, single measure, model: ICC (2,1) (Landers, 2015; Shrout & Fleiss, 1979). The absolute agreement for the noncritical theme was of ICC =.99, and for the critical theme, ICC = .93, showing high agreement. At this point, any disagreement was discussed by the two coders and resolved. Then, one coder coded the remaining 70% of the transcripts. Pieces of information were divided into "true" information and "false" information for manipulation checks (see below). However, for hypothesis testing, the information was separated for information concerning the noncritical and critical themes but not for veracity. One aim of the present study was to mirror a situation where the interviewers did not have previous knowledge that would give them the opportunity to detect any statement-evidence inconsistency.

Results

[30] Reliability coding on around 25% of the sample is usual practice in verbal lie detection research (Deeb et al., 2017; Nahari & Vrij, 2015). Although performing the ICCs only on a proportion of stimuli may have limitations, our high agreement should make this less significant.

Manipulation check

Participants reported high levels of motivation to perform well during the interview (M = 6.09, SD = .91) and thought that they appeared credible (M = 4.85, SD = .95). Motivation did not differ between truth tellers (M = 6.08, SD = 1.02) and liars (M = 6.09, SD = .78), $t(64.47) = -.048$, $p = .96$, $d = -0.01$ [-0.49, 0.47]. Perceived credibility differed between truth tellers (M = 5.08, SD = .94) and liars (M = 4.59, SD = .91), $t(66) = 2.179$, $p = .03$, $d = 0.53$ [0.04, 1.01]. Additionally, liars revealed more pieces of information that were false (M = 11.44, SD = 4.14) than truth tellers (M = 1.05, SD = 1.09), $t(34.845) = 13.758$, $p < .001$, Cohens' $d = -3.53$ [-4.24, -2.73]. This means that the Veracity manipulation was successful[31].

Previous research found that interviewers can have an influence on the interviewees' answers (Mann et al., 2013). A linear mixed model analysis was conducted. The mean intercept and the interviewer condition (3 levels: interviewer 1, vs. interviewer 2, interviewer 3, between-subjects) were the fixed factors. Intercepts were the random factor. The number of reported pieces of information was the dependent variable. The model was not significant, $F(2, 65) = .17$, $p = .84$.

Hypothesis testing

A linear mixed-model was conducted with Theme (noncritical vs. critical, within-subjects) and Veracity (truth telling vs. lying, between-subjects) as fixed factors and the total amount of *reported* information (information included in the original story plus information added by the participants) throughout the interview as dependent variable. The amount of information for the noncritical and critical themes *remembered* by the interviewees before the beginning of the interview were the covariates. The intercepts were the random effect. Effect sizes are reported in Table 1 and Figure 1 for comparisons purposes.

[31] The fact that perceived credibility differed between veracity conditions further supports the assumption that the veracity manipulation worked. Indeed, liars believed to be less credible than truth tellers, which would be expected because of the illusion of transparency (Gilovich, Savitsky, & Medvec, 1998). We decided not to enter the *believability* score as a covariate because we asked how believable the participants thought they were *after* the interview was conducted and thought it to be unlikely that such score would influence the conclusion drawn from the analyses. Re-running the mixed model with the inclusion of *believability* covariate supported this assumption as the conclusions remained the same.

The Theme main effect was significant, $F(1, 62) = 47.11$, $p < .001$. Participants reported more pieces of information when talking about the noncritical theme (M = 16.26, DS = 1.90) than when talking about the critical theme (M = 13.56, DS = 4.03) (Table 2).

The main effect for Veracity was also significant, $F(1, 62) = 10.08$, $p = .002$. Truth tellers reported more pieces of information overall (M = 31.86, DS = 4.26) than liars (M = 27.53, DS = 4.76) (Table 3).

The Theme by Veracity interaction was also significant, $F(1, 62) = 17.63$, $p < .001$ (Figure 2). Simple effect analyses (Table 4) showed that truth tellers reported a similar amount of information for the noncritical (M = 16.25, DS = 1.90) and critical (M = 15.61, DS = 2.87) themes, $F(1, 62) = 3.47$, $p = .07$, $\text{LogBF}_{(10)} = -0.523$[32]. In contrast, liars reported more pieces of information when talking about the noncritical theme (M = 16.28, DS = 1.92) than when talking about the critical theme (M = 11.25, DS = 3.93), $F(1, 62) = 62.66$, $p < .001$, $\text{LogBF}_{(10)} = 12.70$. These findings support Hypothesis 1.

We conducted the same analyses without entering the covariates into the model and obtained similar results. The Theme main effect, $F(1, 66) = 52.92$, $p < .001$, Veracity main effect, $F(1, 66) = 15.64$, $p < .001$, and Theme by Veracity interaction, $F(1, 66) = 31.76$, $p < .001$, were again all significant. Simple effect analyses were again not significant for truth tellers, $F(1, 66) = 1.43$, $p = .23$, but significant for liars, $F(1, 66) = 78.70$, $p < .001$.

We conducted the same analyses without entering the covariates into the model and without excluding outliers. We obtained similar results except that for the Veracity main effect. The Theme main effect, $F(1, 71) = 25.97$, $p < .001$, and the Theme by Veracity interaction, $F(1, 71) = 11.54$, $p = .001$, were again significant. Simple effect for truth tellers were again not significant, $F(1, 71) = 1.46$, $p = .23$, whereas simple effect for liars were again significant, $F(1, 71) = 35.58$, $p < .001$. The Veracity main effect, however, was no longer significant, $F(1, 71) = 1.42$, $p = .23$.

[32] The Log(BF)s do not include the covariates as they are based on Bayesian Paired *t*-tests.

In Hypothesis 2 we predicted that the within-subjects measure would be more effective to discriminate truth tellers from liars than the between-subjects measure. An appropriate way to test this hypothesis is to compare the effect sizes of the two methods. The effect sizes are a measure of the magnitude of differences, where larger effect sizes imply larger differences (see for overviews about the importance of effect sizes and its comparison with significance testing, du Prel, Hommel, Röhring, and Blettner [2009] and Fritz, Morris, and Richler [2012]). Such an approach has already been used in previous research (Deeb et al., 2017).

Cohen (1988, 1992) states that an effect of $d > .80$ is large and noticeable by observers. For the between-subjects measure focused on the entire interview, we obtained a Cohens' $d = 0.96$ [0.45, 1.45]. For the within-subjects measure, we obtained a Cohens' $d = 0.26$ [-0.04, 0.57] for truth tellers and a Cohens' $d = 1.63$ [1.04, 2.21] for liars (Table 1 and Figure 1).

It is also important to compare truth tellers and liars when focusing on the critical theme only, as this is the only theme about which the participants were asked to either lie or tell the truth.

An ANCOVA with Veracity (truth tellers vs. liars) as the factor, the amount of remembered information for the critical theme as the covariate, and the amount of revealed information for the critical theme as the dependent variable showed that the effect for Veracity was significant, $F(1, 64) = 17.75$, $p < .001$, $d = 1.28$ [0.74, 1.79]. Truth tellers reported more pieces of information than liars (Table 4).

Hypothesis 2 can be only partially supported for the following two reasons. First, although the effect size for the within-subjects measure, when looking at liars, was larger than the effect size obtained for the between-subjects measure focusing on the entire interview, both were large. Second, there is an overlap between the confidence intervals of the two effect sizes. However, in favor of the within-subjects measure, truth tellers showed only a small difference when talking about the two themes, Cohens' $d = 0.26$ [-0.04, 0.57] and an investigator would probably not notice a difference (Cohen, 1988). In contrast, the effect size for liars was large, Cohens' $d = 1.63$ [1.04, 2.21] and an investigator would arguably notice a difference (Cohen,

1988). Furthermore, there was no overlap between the confidence intervals of the two within-subjects measures.

Discussion

In this experiment, we compared the efficacy of a within-subjects measure to that of a between-subjects measure to detect deception and tested the efficacy of a Theme-Selection approach to detect which part of the statement included a lie. Truth tellers reported the same amount of information about both themes, whereas liars reported less information for the theme they lied about than for the theme they told the truth about. Furthermore, larger differences between truth tellers and liars were found when focusing on within-subjects than on between-subjects comparisons focusing on the entire interview. This supports the idea that within-subjects measures are preferable to between-subjects measures.

The between-subjects comparisons, similarly to the within-subjects measure, also yielded strong effect sizes (especially when focusing on the critical theme only), yet this result has little applied value. To apply a between-subjects comparison an investigator should first determine a cut off score: what is the minimum amount of information that should be provided to consider a statement as truthful? This is an impossible task, for example, due to substantial individual differences between interviewees in how much information they volunteer in interviews (Nahari & Pazuelo, 2015; Vrij, 2016) and situational differences (some events are richer in detail than other events).

The fact that the effect size for the between-subjects comparison concerning the critical theme only was larger than that for the comparison accounting for the entire interview is due to the fact that liars' statements concerning the critical theme only were entirely deceptive. Therefore, the advantage of our within-subjects measure over the between-subjects measure was reduced when examining the critical theme only. Yet, such a comparison is only possible when the statements have been split into the two themes.

Our results strengthen the idea that within-subjects measures are better than between-subjects measures, but they must be taken with caution as there are some limitations.

For example, in the present experiment the deceptive part of the statement was entirely false[33], which often would not mirror real life, as liars typically tell a mixture of truths and lies (Leins et al., 2013). Future research should explore the present approach when the false theme is itself a mixture a truth and lies.

In addition, the Theme-Selection approach also has a cut off score problem: Which difference in reported information between the two themes is required to decide that someone is lying? Although a within-subjects comparison controls for individual differences, the issue of situational differences is still relevant (some events are richer in detail than other events) and a difference in reporting details between the two events could appear also for truth tellers. The same applies when an interviewee has better memory for one theme than for the other theme. Hence, a difference in detail between themes does not automatically imply lying.

Finally, a liar may lie about both events in which case liars may report an equal amount of details for both themes. Therefore, a lack of difference does not automatically imply truth telling. It is therefore important that future research explores the effectiveness of the Theme-Selection Strategy when the two (or more) themes are intrinsically different and/or include a different amount of detail. We expect the approach to be less effective in those situations than in the current experiment. Future research could also explore in a lie detection experiment how the Theme-Selection strategy affects observers' accuracy in discriminating between truth tellers and liars. In addition, in the present experiment we demarcated the two subthemes a priori and future research could explore how skilled interviewers are in separating subthemes in a story. Indeed, research is needed to explore if two (or more) interviewers split interviewees' statement in the same way. For this to happen, a theme needs to be defined. In our view a theme is a cluster of pieces of information or events that are more related to each other than other pieces of information or events.

Furthermore, we did not apply any strategic questioning, but combining the Theme-Selection approach with strategic questioning may prove effective for separating truth

[33] Twenty-one (65.6%) of lying participants did not report any true detail about the critical event, whereas the remaining eleven liars (35.4%) reported on average 12.6% of true details about the critical event.

tellers from liars and to understand what specific part(s) of the story is false. For example, the interviewer could ask unexpected questions (Lancaster et al., 2013; Vrij, 2018) for each theme and explore if the interviewee's answers to such unexpected questions for one theme differ from those of another theme. Similarly, other measures such as ratios between verifiable and non-verifiable details (Nahari et al., 2014) or between complication and other types of details (Vrij et al., 2018) may be employed.

There was a methodological limitation in our study worth mentioning: We did not counterbalance the theme about which the interviewees lied. Although the rationale on which we built our experiment (different cognitive processes and strategies between truth telling and lying) is not affected by the theme about which the interviewees lies, the content of the various subthemes could have had an influence on the outcome. Therefore, counterbalancing should take place in future research. Lastly, our experiment was based on a role-playing situation. This has to be taken into account when considering the ecological validity of our results. A recent meta-analysis aiming to shed light on the issue concluded that "*[…]the findings from deception research are not laboratory artifacts- the detectability of deception remains stable across a variety of situational variables.*" (Hartwig & Bond, 2014, p. 667). Thus, although there are differences between real life and lab settings, lab research is still informative. In addition, we realise that there are difficulties in generalizing lab findings to real life when 'stakes' play an important role in the lab study and interpretation of the lab research findings. However, in our experiment we focused on interviewees' strategies. There is no reason to believe that truth tellers and liars in laboratory settings use different strategies to appear credible than truth tellers and liars in real life.

Data Availability

The raw data supporting the conclusions of this manuscript will be made available by the authors, without undue reservation, to any qualified researcher.

Ethics statement

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with APA regulations.

Author contributions

NP conceived the idea for the study, designed the experiment, conducted the data analysis, interpreted the results, and wrote the manuscript. LC contributed to the design of the experiment, to the data analysis and interpretation of the results, to writing up the manuscript, and provided feedback. AV contributed to the data analysis, to the interpretation of the results, to writing up the manuscript and provided feedback. All authors agreed on the final version of the manuscript.

Conflict of Interests Statement

The authors declare no conflict of interest.

References

Bond, C. F. Jr., & DePaulo, B. M. (2006). Accuracy of deception judgments. *Personality and Social Psychology Review*, *10*(3), 214‑234. doi:10.1207/s15327957pspr1003_2.

Bond Jr, C. F., & DePaulo, B. M. (2008). Individual differences in judging deception: accuracy and bias. *Psychological Bulletin*, *134*(4), 477-492. doi:10.1037/0033-2909.134.4.477.

Christ, S. E., Van Essen, D. C., Watson, J. M., Brubaker, L. E., & McDermott, K. B. (2008). The contributions of prefrontal cortex and executive control to deception: evidence from activation likelihood estimate meta-analyses. *Cerebral cortex*, *19*(7), 1557-1566. doi: 10.1093/cercor/bhn189.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (second edition)*. Hillsdale, NJ: Erlbaum.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155‑159. doi:10.1037/0033-2909.112.1.155.

Debey, E., Verschuere, B., & Crombez, G. (2012). Lying and executive control: An experimental investigation using ego depletion and goal neglect. *Acta Psychologica*, *140*(2), 133-141. doi: 10.1016/j.actpsy.2012.03.004.

Deeb, H., Vrij, A., Hope, L., Mann, S., Granhag, P. A., & Lancaster, G. L. (2017). Suspects' consistency in statements concerning two events when different question formats are used. *Journal of Investigative Psychology and Offender Profiling*, *14*(1), 74-87. doi: 10.1002/jip.1464.

DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. Psychological Bulletin, 129(1), 74‑118. doi:10.1037/0033-2909.129.1.74.

du Prel, J. B., Hommel, G., Röhrig, B., & Blettner, M. (2009). Confidence interval or p‑value?: Part 4 of a series on evaluation of scientific publications. *Deutsches Ärzteblatt International*, *106*, 335‑339. doi:10.3238/arztebl.2009.0335.

Evans, J. R., Michael, S. W., Meissner, C. A., & Brandon, S. E. (2013). Validating a new assessment method for deception detection: Introducing a Psychologically Based Credibility Assessment Tool. *Journal of Applied Research in Memory and Cognition*, *2*(1), 33-41. doi: 10.1016/j.jarmac.2013.02.002.

Ewens S., Vrij A., Jang M., & Jo E., (2014). Drop the Small Talk When Establishing Baseline Behaviour in Interviews. *Journal of Investigative Psychology and Offender Profiling*, *11*, 244–252. doi: 10.1002/jip.1414.

Fisher, R. P., & Geiselman, R. E. (1992). *Memory enhancing techniques for investigative interviewing: The cognitive interview*. Charles C. Thomas Publisher.

Fisher, R. P., Schreiber Compo, N., Rivard, J., & Hirn, D. (2014). Interviewing witnesses. *The SAGE handbook of applied memory*, 559-578..

Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: Current use, calculations, and interpretation. *Journal of Experimental Psychology: General*, *141*, 2‑18. doi:10.1037/a0024338.

Gilovich, T., Savitsky, K., & Medvec, V. H. (1998). The illusion of transparency: Biased assessments of others' ability to read one's emotional states. *Journal of Personality and Social Psychology, 75*, 332–346. doi: 10.1037/0022-3514.75.2.332.

Granhag, P. A., & Luke, T. J. (2018). How to Interview to Elicit Concealed Information: Introducing the Shift-of-Strategy (SoS) Approach, In J. P. Rosenfeld (Eds.), *Detecting Concealed Information and Deception* (pp. 271-295). London, UK: Elsevier Academic Press.

Hartwig, M., & Bond Jr, C. F. (2014). Lie detection from multiple cues: A meta-analysis. *Applied Cognitive Psychology*, *28*(5), 661-676. doi: 10.1002/acp.3052.

Hartwig, M., Granhag, P. A., & Luke, T. (2014). Strategic Use of Evidence during investigative interviews: The state of the science. In D. C. Raskin, C. R. Honts, & J. C. Kircher (Eds.), *Credibility assessment: scientific research an applications* (pp. 1-36). Academic Press.

Hoaglin, D. C., & Iglewicz, B. (1987). Fine-tuning some resistant rules for outlier labeling. *Journal of the American Statistical Association*, *82*(400), 1147-1149.

Lancaster, G. L., Vrij, A., Hope, L., & Waller, B. (2013). Sorting the liars from the truth tellers: The benefits of asking unanticipated questions on lie detection. *Applied Cognitive Psychology, 27*(1), 107-114. doi: 10.1002/acp.2879.

Landers, R. (2015). Computing Intraclass Correlations (ICC) as Estimates of Interrater Reliability in SPSS, The Winnower 2: e143518. 81744, 2015, DOI: 10.15200/winn.

Leins, D. A., Fisher, R. P., & Ross, S. J. (2013). Exploring liars' strategies for creating deceptive reports. *Legal and Criminological Psychology, 18*(1), 141-151. doi: 10.1111/j.2044-8333.2011.02041.x.

Mann, S., Vrij, A., Shaw, D., Leal, S., Ewens, S., Hillman, J., ... Fisher, R. (2013). Two heads are better than one? how to effectively use two interviewers to elicit cues to deception. *Legal and Criminological Psychology, 18*(2), 324-340. DOI: 10.1111/j.2044-8333.2012.02055.x

Maguire, M., & John, T. (1995). *Intelligence, surveillance and informants: Integrated approaches.* (Police Research Group Paper No. 64). London: Home Office.

Nahari, G. (2016). Advances in lie detection: Limitations and potential for investigating allegations of abuse. In R. Burnett (Ed.), *Wrongful allegations of sexual and child abuse* (pp. 242–252). Oxford, UK: Oxford University Press.

Nahari, G., & Pazuelo, M. (2015). Telling a convincing story: Richness in detail as a function of gender and priming. *Journal of Applied Research in Memory and Cognition, 4*, 363–367. doi: 10.1016/j.jarmac.2015.08.005.

Nahari, G., & Vrij, A. (2014). Are you as good as me at telling a story? Individual differences in interpersonal reality-monitoring. *Psychology, Crime & Law, 20*, 573–583. doi: 10.1080/1068316x.2013.793771.

Nahari, G., & Vrij, A. (2015). Systematic errors (biases) in applying verbal lie detection tools: Richness in detail as a test case. *Crime Psychology Review, 1*, 98–107. doi: 10.1080/ 23744006.2016.1158509.

Nahari, G., Vrij, A., & Fisher, R. P. (2014). Exploiting liars' verbal strategies by examining the verifiability of details. *Legal and Criminological Psychology, 19,* 227–239. doi: 10.1111/j.2044-8333.2012.02069.x.

Oleszkiewicz, S., Granhag, P. A., & Kleinman, S. M. (2014). On eliciting intelligence from human sources: Contextualizing the Scharff-technique. *Applied Cognitive Psychology, 28*(6), 898-907. doi: 10.1002/acp.3073.

Ost, J., Scoboria, A., Grant, T., & Pankhurst, G. (2015). Recall, verbatim memory and remembered narratives. In G. Oxburgh, T. Myklebust, T. Grant, & R. Milne (Eds.), *Communication in investigative and legal contexts: Integrated approaches from forensic psychology, linguistics and law enforcement* (39-54). Chichester: Wiley.

Palena, N., Caso, L., Vrij, A., & Orthey, R. (2018). Detecting deception through small talk and comparable truth baselines. *Journal of Investigative Psychology and Offender Profiling, 15*(2), 124-132. doi:10.1002/jip.1495.

Rosenfeld, J. P. (Ed). (2018). *Detecting Concealed Information and Deception. Recent Developments*. Academic Press.

Shrout, P., & Fleiss, J. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86,* 420-428. DOI: 10.1037/0033-2909.86.2.420.

Vrij, A. (2008). *Detecting lies and deceit: Pitfalls and opportunities*. Chichester: John Wiley and Sons.

Vrij, A. (2014). Interviewing to detect deception. *European Psychologist, 19,* 184‑194. doi:10.1027/1016-9040/a000201.

Vrij, A. (2015). A cognitive approach to lie detection. In P. A. Granhag, A. Vrij, & B. Verschuere (Eds.), *Detecting deception: Current challenges and cognitive approaches* (pp. 205‑230). Chichester: John Wiley and Sons.

Vrij, A. (2016). Baselining as a lie detection method. *Applied Cognitive Psychology, 30,* 1112–1119. doi: 10.1002/acp.3288.

Vrij, A. (2018). Verbal Lie Detection Tools From an Applied Perspective, In J. P. Rosenfeld (Eds.), *Detecting Concealed Information and Deception* (pp. 297-327). London, UK: Elsevier Academic Press.

Vrij, A., Fisher, R. P. and Blank, H. (2017), A cognitive approach to lie detection: A meta-analysis. *Legal and Criminological Psychology, 22*, 1-21. doi:10.1111/lcrp.12088.

Vrij, A., & Granhag, P. A. (2012). Eliciting cues to deception and truth: What matters are the questions asked. *Journal of Applied Research in Memory and Cognition*, *1*(2), 110-117. doi: 10.1016/j.jarmac.2012.02.004.

Vrij, A., Leal, S., Jupe, L., & Harvey, A. (2018). Within-subjects verbal lie detection measures: A comparison between total detail and proportion of complications. *Legal and Criminological Psychology*. doi: 10.1111/lcrp.12126.

Table 1. *Between-subjects vs. Within-subjects effect sizes and parameter estimates comparisons. Effects are obtained from the linear mixed model used for hypothesis testing*

		95% C.I.			95% C. I.	
Between-subject effects	*Cohen's d*[a]	LL	UL	*Parameter estimates*[b] (*SE*)	LL	UL
Veracity [both themes]	0.96	0.45	1.45	-0.83 (0.26)	-1.34	-0.32
Veracity [critical theme]	1.28	0.74	1.79	-1.65 (0.39)	-2.43	-0.87
Within-subject effects						
Theme [both veracity conditions]	0.86	0.54	1.17	-1.42 (0.21)	-1.83	-1.02
Theme [truth tellers]	0.26	-0.04	0.57	-0.55 (0.30)	-1.14	0.04
Theme [liars]	1.63	1.04	2.21	-2.29 (0.29)	-2.87	-1.71

[a]*Cohen's d are computed on original metrics and are not controlled for the covariates;* [b]*Parameter estimates are obtained from the linear mixed model used for hypothesis testing*.

Figure 1. *Comparison of between-subjects and within-subjects effect sizes, with 95% CIs. Cohen's d are computed on original metrics and are not controlled for the covariates.*

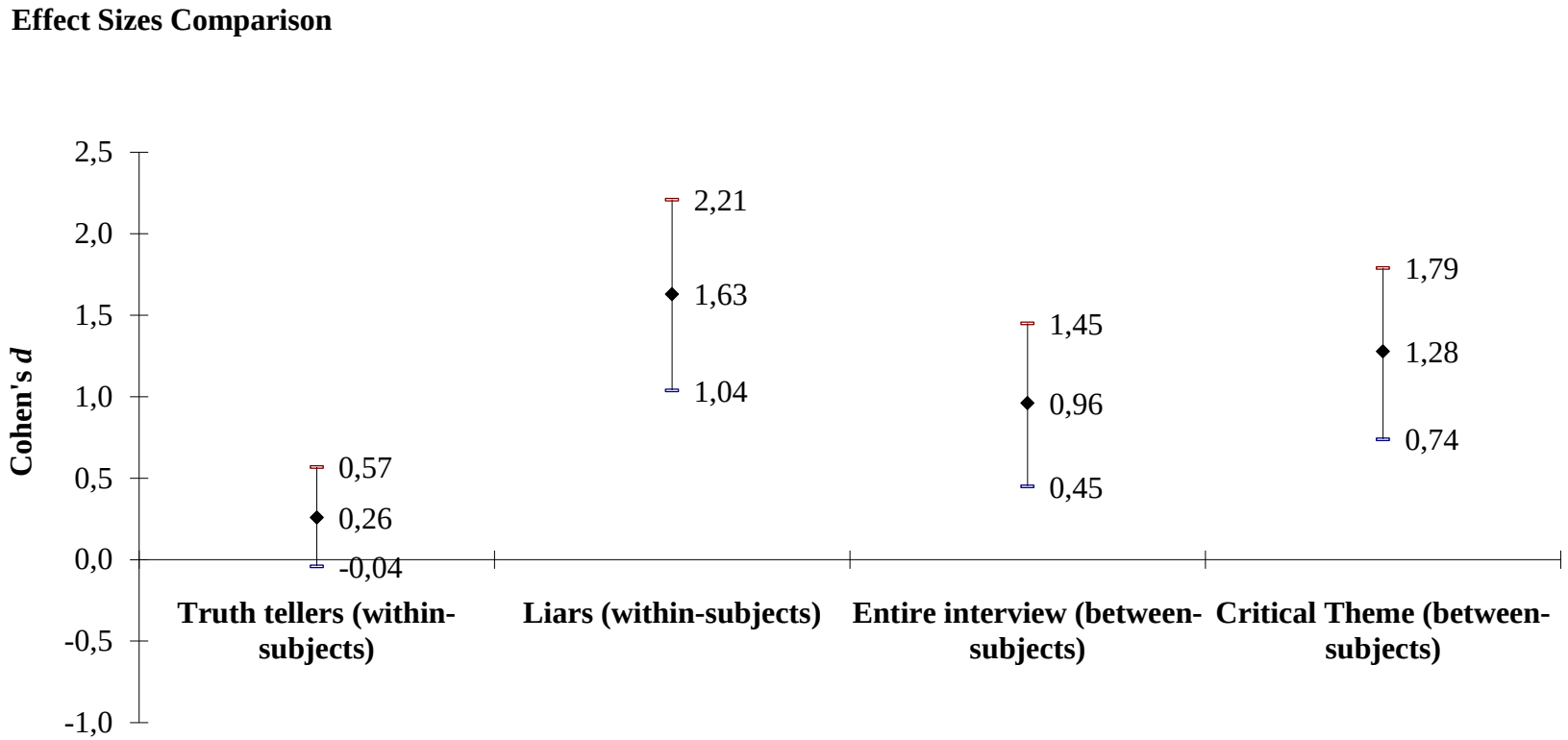


Figure 2. *Theme by Veracity interaction. Y-axis shows the amount of revealed information.*

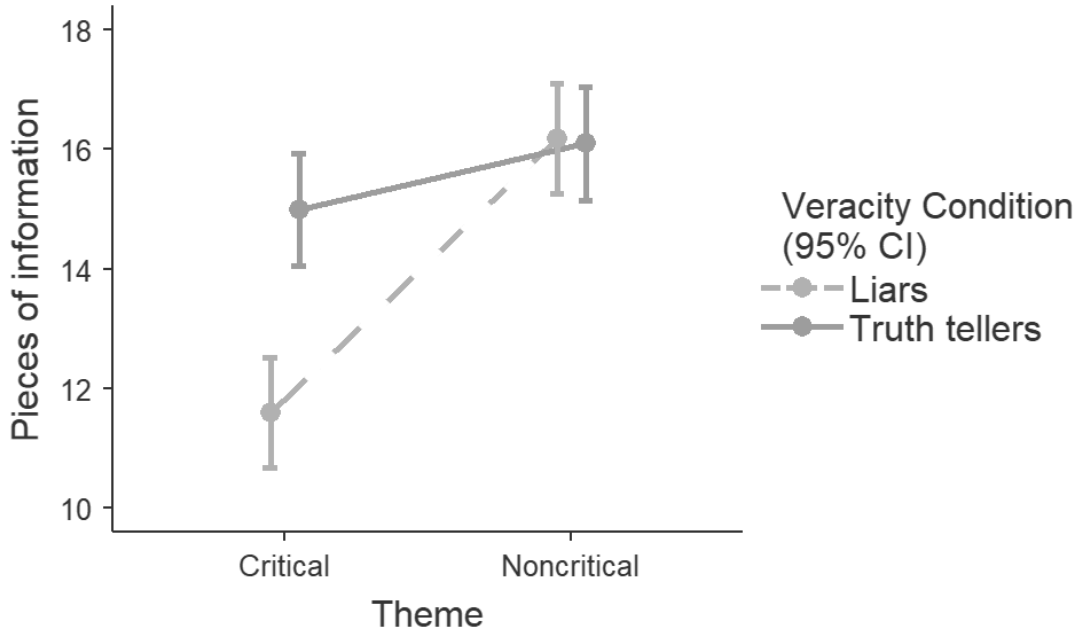


Table 2. *Descriptive for the factor "Theme" obtained from the linear mixed model used for hypothesis testing. Means and SD are computed without adjusting for the covariates and for the factor "Veracity". Estimated marginal means show the mean effect for the factor Theme, adjusting for the effect of the covariates and factor Veracity.*

Theme	*M*	*DS*	C.I.		Est. Marginal Mean	*SE*	C.I.	
			LL	UL			LL	UL
Noncritical	16.26	1.90	15.80	16.72	16.14	.33	15.48	16.80
Critical	13.56	4.03	12.58	14.53	13.29	.33	12.63	13.95

Table 3. *Descriptive for the factor "Veracity" obtained from the linear mixed model used for hypothesis testing. Means and SD are computed without adjusting for the covariates and for the factor "Theme". Estimated marginal means show the mean effect for the factor Veracity, adjusting for the effect of the covariates and factor Theme.*

	M	*DS*	C.I.		Est. Marginal Mean	*SE*	C.I.	
			LL	UL			LL	UL
Truth tellers	31.86	4.26	30.42	33.30	15.54	.37	14.80	16.29
Liars	27.53	4.76	25.81	29.25	13.89	.36	13.16	14.62

Table 4. *Simple effects analyses Descriptive. All statistics are reported without adjusting for the covariates*.

	Noncritical				**Critical**			
	M	*DS*	C.I.		*M*	*DS*	C.I.	
			LL	UL			LL	UL
Truth tellers	16.25	1.90	15.60	16.89	15.61	2.87	14.64	16.58
Liars	16.28	1.92	15.59	16.97	11.25	3.93	9.83	12.67

Appendix 8: Caso, L., & Palena, N. (2018). Tecniche di interrogatorio con soggetti adulti: una rassegna internazionale. *Rassegna Italiana di Criminologia, 2*, 130-140. [PA5]

Abstract Italiano

La credibilità della persona sottoposta ad interrogatorio è un tema centrale per l'ambito giudiziario e il cui studio richiede sinergie con altre discipline, in particolare le scienze psicologiche. Infatti, in letteratura, soprattutto internazionale, sono disponibili diverse tecniche sviluppate a partire da collaborazioni tra mondo della giustizia e psicologia. Lo scopo di questo lavoro è presentare le diverse strategie ad oggi presenti. A tal fine, sono stati passati in rassegna i lavori più recenti, cercando di sottolineare i pro e i contro di ogni tecnica presentata. In conclusione, risultano chiari l'efficacia di alcune di queste ed il beneficio che il loro utilizzo da parte delle forze di polizia potrebbe portare.

Interrogatorio, testimonianza, credibilità, interviste strutturate

Abstract Inglese

Credibility assessment is a central theme in juridical settings, whose study needs synergies with other disciplines, in particular with psychological sciences. Indeed, several techniques, which have been developed through the collaboration between the criminal and the psychological settings, are available in the literature, especially in the international scenario. The main goal of this work is to present the strategies which are available today. To attain this aim, most recent work on the topic has been reviewed, and we underlined the strengths and the weaknesses of each of the reported techniques. In conclusion, the efficacy of some of these techniques appears clear, as well as the fact that police forces may benefit from their use.

Interrogation, testimony, credibility assessment, structured interviews

Introduzione

In Italia, sul tema degli interrogatori dei sospettati, assistiamo ad una diversità di prassi operative. Infatti, le nostre forze dell'ordine raramente frequentano specifici corsi sulle tecniche di interrogatorio. Piuttosto, apprendono come condurre un'intervista investigativa sulla base dell'esperienza e degli insegnamenti dei colleghi più anziani. Inoltre, nel caso di interrogatorio a soggetti adulti presunti rei, sono presenti anche delle limitazioni di carattere normativo; non è infatti prevista la presenza di uno psicologo che coadiuvi tale fase se non in particolari condizioni (articoli da 220 a 232 e 508 c.p.p.). Tuttavia, sappiamo che in ogni iter investigativo l'interrogatorio è una parte importantissima: le prime rilevazioni testimoniali possono indirizzare in un senso rispetto ad un altro le indagini successive, per cui comprendere velocemente se la persona che viene interrogata è credibile e se dice o meno il vero è una questione molto importante. Infine la giustizia, oltre che a trattare casi cosiddetti "ordinari", deve confrontarsi sempre più spesso con situazioni complesse, come ad esempio quelle dove ad essere interrogati sono i collaboratori di giustizia ed i soggetti coinvolti in atti terroristici. Ad oggi non troviamo in Italia ricerche che discutano degli elementi psicologici a supporto di tecniche vantaggiose per chi interroga, mentre la letteratura internazionale si sta connotando di ricerche interessanti che stanno cercando di convergere verso protocolli scientificamente validi e condivisi anche dalla giustizia.

Scopo principale di questo lavoro è quindi analizzare lo sviluppo delle tecniche di intervista utilizzate con gli adulti, soprattutto in un'ottica di valutazione della credibilità, e mettere in luce pro e contro di ognuna di esse. Verranno presi in considerazione sia gli approcci già strutturati che quelli in via di sviluppo. Nel primo caso, si parla di tecniche basate su una metodologia prestabilita e sottoposte ad analisi sperimentale. Nel secondo caso invece vengono incluse quelle tecniche che, pur avendo ricevuto supporto empirico, non siano ancora state sviluppate in un protocollo strutturato. I presupposti teorici sui quali poggiano tali tecniche sono sia di origine cognitiva (codifica e recupero del ricordo, presa di decisione, ruolo della memoria di lavoro) che sociale (ruolo dell'interazione intervistatore-intervistato, *rapport*, ecc.). Al fine di raggiungere questo obiettivo, è stata fatta una ricerca degli articoli scientifici che trattano le varie tecniche di interrogatorio ed intervista attraverso i più

comuni database (es.: PsychInfo) con le parole chiave "*Investigative Interviewing*", "*Interrogation*" e "*Deception Detection*". L'analisi ha prodotto oltre 300 articoli, i quali sono stati scremati sulla base degli obiettivi di questa revisione ed escludendo articoli che hanno utilizzato le stesse metodologie o quelli troppo specifici. Allo stesso modo, sono stati esclusi quegli articoli che non erano inerenti al focus di questo lavoro e quelli di taglio puramente teorico. Alla fine di questa scrematura, sono stati selezionati 74 lavori.

Tecniche di intervista (semi)strutturate

Intervista Cognitiva

Una delle più note tecniche di intervista è l'Intervista Cognitiva (Fisher & Geiselman, 1992; Fisher, Geiselman, Raymond, Jurkevich, & Warhaftig, 1987; Sooniste, Granhag, Strömwall, & Vrij, 2015), comunemente utilizzata dalle forze di polizia di alcuni paesi come Germania e Regno Unito (Milne & Bull, 1999). Essa ha come obiettivo principale quello di aumentare il numero delle informazioni riportate dall'intervistato, senza che ne venga influenzata negativamente l'attendibilità. Questa tecnica di intervista è stata sviluppata partendo da presupposti psicologici che si basano su aspetti sociali, mnemonici e comunicativi. Da una parte, viene prestata particolare attenzione alle relazioni interpersonali tra l'intervistatore e l'intervistato, con lo scopo principale di gestire le interviste di soggetti ansiosi o in difficoltà. Dall'altra viene riconosciuta l'importanza di una tecnica pulita e strutturata al fine di migliorare il recupero dei ricordi, attraverso l'utilizzo di tecniche specifiche che derivano direttamente dalla ricerca empirica sulla memoria, soprattutto tenendo in considerazione che la memoria può essere facilmente influenzata.

La persona intervistata viene inoltre invitata a mantenere un ruolo attivo, sia al fine di metterla a proprio agio, sia perché in questo modo l'intervista stessa risulta più efficace. È infatti riconosciuta l'importanza del creare una relazione positiva tra intervistatore ed intervistato (in inglese *rapport*) al fine di produrre un ambiente privo di *stressors* negativi, nei limiti del possibile.

Per quanto riguarda invece le specifiche mnemotecniche l'Intervista Cognitiva ne prevede quattro (cfr. De Leo, Scali, & Caso, 2005; Dodson, Powers, & Lytell, 2015):

Reintegrare il contesto: viene chiesto al soggetto di rievocare e ricreare il contesto all'interno del quale è successo l'evento. Si chiede quindi di ricordare posti, suoni, odori, emozioni e pensieri che egli stava vivendo in quel momento. Tali tecniche permettono un migliore recupero del ricordo, e sono legati alla teoria della specificità della codifica (Bui, 2015; Tulving & Thomson, 1973)

Riferire ogni cosa: spesso le persone intervistate credono che gli unici dettagli di interesse siano quelli strettamente legati all'evento accaduto. In realtà, in questo modo le forze di polizia potrebbero perdere informazioni molto importanti. Per questo motivo, la persona intervistata è invitata a riportare ogni cosa ricordi, anche ciò che crede essere irrilevante.

Ricordare e riportare gli eventi in ordine differente: viene fatta richiesta di riportare l'evento anche con ordine non puramente cronologico, partendo ad esempio dalla fine o dalla metà dell'evento. Sebbene questa tecnica possa essere utilizzata solo a partire dagli 8 anni, essa risulta particolarmente importante in quanto l'utilizzo di ordini differenti può fungere da *cuing* (suggerimento) e far ricordare informazioni che altrimenti andrebbero dimenticate.

Cambiare prospettiva: infine, il teste viene incoraggiato a riportare l'evento in una prospettiva differente da quella in prima persona, ovvero come se l'evento venisse visto da un terzo soggetto o da una specifica posizione (ad esempio come se una vittima si trovasse dietro all'abusante invece che dinanzi a lui). Questa tecnica però può essere utilizzata solo a partire dai 10 anni.

Considerando quanto detto sopra, appare evidente come l'Intervista Cognitiva sia applicabile solo a partire dagli 8 anni di età. Oltre alle tecniche appena elencate, la procedura si basa su 5 specifiche fasi:

Viene costruito *rapport* con il testimone, al fine di accogliere la persona intervistata e creare un'atmosfera tranquilla.

Racconto libero: si chiede di riportare tutto quanto viene ricordato e non si interrompe il teste.

Domande: in questa fase vengono poste domande specifiche, che però dovrebbero essere aperte, non suggestive, e riferite solo a ciò che è stato detto fino a quel momento.

Secondo racconto con specifiche modalità: si chiede nuovamente un racconto libero, facendo però richiesta di riportare l'evento secondo una delle tecniche prima esposte, come il raccontare l'evento da una prospettiva differente.

Fase di chiusura: in questo momento l'intervistatore porterà al termine l'intervista, con l'obiettivo primario di mettere nuovamente a proprio agio il teste per fare in modo che questi si possa rilassare e tranquillizzare.

La letteratura ha mostrato che l'Intervista Investigativa è una tecnica di intervista molto efficace (cfr. De Leo, Scali, & Caso, 2005). Una rassegna riferita ai primi dieci anni di applicazione della procedura (Geiselman, 1994, citato in De Leo, Scali, & Caso, 2005) ha trovato che come sia possibile raccogliere il 58% di informazioni corrette in più rispetto ad altre forme di intervista, e questo è parso essere vero per ogni tipologia di ricordo. Altri autori hanno replicato l'efficacia dell'Intervista Cognitiva (Cavedon & Calzolari, 1999; Hernandez-Fernaud & Alonso-Quecuty, 1997; Menon, & Stevenage, 1996, Mastroberardino 2011), rendendo quindi possibile concludere che questo protocollo è davvero utile e che può essere di aiuto nelle indagini investigative.

P.E.A.C.E.

P.E.A.C.E., acronimo di *Planning and Preparation, Engage and Explain, Account, Closure, ed Evaluation* (in italiano: Pianificazione e Preparazione, Spiegazione ed Ingaggio, Raccolta, Chiusura e Valutazione), è un protocollo sviluppato da Pearse e Gudjonsson (1996) al fine di ottenere una procedura di addestramento standard ed efficace da insegnare agli investigatori, in particolare agli appartenenti alle forze di polizia. Ciò assume particolare rilievo quando si tiene in considerazione come le forze di polizia sembrano essere impreparate rispetto agli aspetti psicologici e sociali in gioco durante l'interrogatorio, sia quando ad essere intervistate sono i principali sospettati di un crimine che quando vengono intervistati dei testimoni. Impreparazione che è presente in diversi paesi, come in Italia (cfr. Caso & Vrij, 2009)

e Regno Unito (Ainsworth, 2002; Baldwin, 1992; Gudjonsson, 2003; Shawyer & Milne, 2015; Snook, Eastwood, & Barron, 2014). Nella fase di Pianificazione e Preparazione l'intervistatore dovrebbe preparare l'intervista ponendo attenzione agli obiettivi principali ed agli aspetti chiave che si vogliono esplorare. Successivamente, nella fase Spiegazione ed Ingaggio, l'intervistato viene informato circa i motivi del colloquio e l'intervista parte solitamente in una forma conversazionale o discorsiva. La fase di raccolta ha invece come scopo principale quello di ottenere il maggior numero possibile di informazioni. Solitamente questo viene fatto attraverso l'Intervista Cognitiva (Dodson, Powers, & Lytell, 2015; Fisher & Geiselman, 1992), quando il teste è collaborativo, oppure con la tecnica *Conversation Management* (Shepherd, 1988) quando non lo è. Nella fase di Chiusura, l'intervistatore riassume quanto è stato detto fino a quel momento e chiede al teste se vuole in qualche modo modificare quanto detto. Infine, nella fase di Valutazione, vengono stimate l'impatto e l'importanza delle informazioni ottenute dal teste. La ricerca ha dimostrato che il protocollo PEACE risulta essere efficace, sia subito dopo che il personale è stato addestrato al suo uso, che dopo 6 mesi dalla fine del training stesso (McGurk, Carr & McGurk, 1993), dando credito a questa tecnica che risulta anche essere poco invasiva per la persona intervistata. Ciò nonostante, è stato suggerito che potrebbe essere possibile ottenere ulteriori miglioramenti se l'addestramento alla tecnica PEACE venisse ulteriormente perfezionato (Clarke, Milne, & Bull, 2011).

Behavioral Analysis Interview e tecnica dei nove passi di Reid

Una tecnica che ha uno scopo totalmente diverso da quelle precedenti è quella dei nove passi di Reid. Diversamente dall'Intervista Cognitiva e dal protocollo PEACE, infatti, l'obiettivo principale di questo approccio è quello di abbassare le difese dell'interrogato al fine di ottenere una confessione. Caso e Vrij (2009) fanno notare che i nove passi di Reid (Buckley, 2013; Inbau, 2013; Inbau, Reid, Buckley, & Jayne, 2013) vengono utilizzati solo quando esistono prove a carico dell'imputato o quando egli non è risultato credibile nella *Behaviorual Analysis Interview* (BAI), prima fase del processo investigativo secondo questo approccio. Vrij (2008), nota che la BAI è una tecnica di analisi della credibilità spesso utilizzata dalle forze di polizia statunitensi, al punto da essere tra i metodi che più spesso vengono insegnati alle forze dell'ordine. L'obiettivo principale della BAI è quindi quello di valutare, in una fase

preliminare, se sono presenti degli indizi che fanno pensare che l'interrogato stia mentendo o comunque nascondendo qualcosa. In caso positivo, come poc'anzi esposto, si procede all'interrogazione dell'indiziato a mezzo della tecnica dei nove passi. Inbau et al. (2013) sostengono che l'aspetto centrale della BAI è quello di provocare delle reazioni comportamentali nell'interrogato, reazioni che saranno poi esaminate dall'intervistatore al fine di valutarne la credibilità. La tecnica alterna domande aperte a domande pre-determinate e standardizzate e mira alla ricerca di specifici indizi di veridicità o menzogna (Caso & Vrij, 2009; Inbau et al., 2001, 2013; Vrij, 2008). Successivamente alla valutazione della credibilità dell'imputato, si procede ad un interrogatorio aggiuntivo più dettagliato, il cui scopo principale è quello di ottenere una confessione. La tecnica dei nove passi di Reid, consta dei seguenti passaggi:

Confronto diretto con la persona sospettata, dove la si informa che è considerata l'autore del reato.

Sviluppo dell'argomento, fase in cui la persona interrogata viene fronteggiata con tecniche di minimizzazione o massimizzazione a seconda del suo atteggiamento emotivo.

Gestione delle negazioni, momento in cui si cerca di persuadere il sospettato a dire la verità.

Fronteggiamento delle obiezioni.

Catturare e mantenere l'attenzione del sospettato.

Gestione della passività, fase in cui l'intervistatore dovrebbe cercare di mostrare di essere comprensivo ed emotivo, sempre col fine ultimo di incoraggiare la persona a dire la verità. Inoltre, l'intervistatore cerca di fare leva sul sentimento del rimorso presente nel colpevole.

Porre domande che comprendono alternative. In questo passaggio l'indagato ha la possibilità di giustificare il crimine commesso, ma ad ogni modo le domande vengono poste in modo tale che, quale che sia la risposta scelta dalla persona intervistata, si avrà come risultato una ammissione.

Ottenere un resoconto verbale di tutti i dettagli del crimine, il cui scopo ultimo è proprio quello di avere una ammissione di colpa completa e dettagliata.

Ottenere una confessione scritta, a partire dal passaggio precedente.

Nonostante la sequenza BAI-9 passi di Reid sia molto conosciuta e sovente utilizzata da alcune forze di polizia (Vrij, 2008), la ricerca ha mostrato i rischi che si corrono nell'utilizzare sia l'una che l'altra. Hovarth, Jayne, e Buckley (1994) sostengono che la BAI permette di valutare la veridicità degli indagati con un'accuratezza molto alta, circa 86%. Ciò nonostante, Vrij (2008) mette in evidenza gli aspetti critici della ricerca di Hovarth e colleghi. Anche altre ricerche hanno sottolineato l'inefficacia della BAI (Blair & Kooi, 2004; Kassin & Fong, 1999; Mann, Vrij, & Bull, 2004). Inoltre, la tecnica dei nove passi è stata duramente criticata (Gudjonsson, 2003; Kassin, 1997; Memon, Vrij, & Bull, 2003), perché può portare la persona intervistata a fornire delle false confessioni (Caso & Vrij, 2009). In effetti, questo protocollo prevede l'utilizzo di strategie di confronto diretto, minimizzazione e massimizzazione. Tali strategie però fanno aumentare il rischio di ottenere una falsa confessione (Caso & Vrij, 2009; Moore & Fitzsimmons, 2011). Nel confronto diretto, l'intervistatore muove forti accuse verso il sospettato, anche sulla base di prove di colpevolezza che non devono per forza essere reali (cfr. Caso & Vrij, 2009). La tecnica di minimizzazione prevede che l'intervistatore si mostri comprensivo nei confronti del sospettato e giustifichi moralmente quanto successo, mentre quella di massimizzazione prevede che l'intervistatore faccia leva sulla paura che la persona interrogata prova, lasciandogli credere che ormai è inutile continuare a mentire (Horgan, Russano, Meissner, & Evans, 2012). Altri fattori di estrema importanza che incidono sul rischio di una falsa confessione sono l'isolamento, la fatica e la stanchezza dovuta a lunghi interrogatori (Moore & Fitzsimmons, 2011). Infine, interrogatori estremamente suggestivi e tecniche coercitive possono persino portare al sospettato a convincersi di essere colpevole e formarsi un'immagine mentale ed un ricordo vivido dell'evento a sostegno della confessione stessa (Henkel & Coffman, 2004). In conclusione, nonostante la forte pubblicità di cui questi due protocolli godono, sarebbe opportuno smettere di usarli vista la loro inefficacia, almeno fino a quando non dovessero essere disponibili ricerche che ne sostengono la validità.

Assessment Criteria Indicative of Deception

Un protocollo relativamente recente, che si basa su un approccio totalmente diverso da quello della BAI e della tecnica dei nove passi di Reid, è quello sviluppato da Colwell e colleghi (per una trattazione completa, cfr. Colwell, Hiscock-Anisman, & Fede, 2013). Infatti, esso predilige un approccio etico e non pressante piuttosto che uno accusatorio come quello della tecnica Reid. Nell'Assessment Criteria Indicative of Deception (ACID), vengono integrate le tecniche dell'intervista cognitiva (leggermente modificata e chiamata *Reality Interview,* Colwell, Hiscock-Anisman, & Memon, 2002) con l'analisi delle dichiarazioni del sospettato. La tecnica ha ricevuto diverse conferme sperimentali, sia con persone che parlavano in Inglese che con persone che parlavano Arabo, e madrelingua spagnoli che parlavano in Inglese (Colwell et al., 2013). Inoltre, il protocollo è stato accolto positivamente anche da altri autori (cfr. Vrij, 2015a). L'ACID consta di una prima fase di *baselining*, dove viene creato *rapport* con la persona intervistata al fine di metterla a proprio agio. Successivamente, viene seguito un protocollo semi-strutturato dove si alternano domande aperte a domande forzate e tecniche di massimizzazione mnemonica come quelle presenti nell'Intervista Cognitiva. Nella fase conclusiva, la persona viene invitata a raccontare nuovamente la storia secondo il proprio punto di vista. Il presupposto di base è che le persone oneste otterranno benefici dalle tecniche messe in atto nello svolgersi dell'intervista, tecniche che permetteranno loro di ricordare un maggior numero di dettagli. Al contrario, una persona disonesta, non otterrà alcun beneficio, dal momento che tali tecniche non fungeranno da suggerimenti utili al recupero di ulteriori informazioni. Colwell et al. (2013) basano quindi la tecnica su ciò che chiamano *Differential Recall Enhancement*. Ovvero, le persone oneste aggiungeranno maggiori dettagli nelle fasi dell'intervista successive alle tecniche mnemoniche, rispetto alle persone disoneste. Inoltre, gli autori considerano anche altri aspetti, come il numero totale di parole, che risulta essere maggiore per gli onesti, e il *Type Token Ratio*, il quale è una misura di complessità linguistica delle dichiarazioni dei sospettati e si misura come rapporto tra numero di parole uniche (non ripetute) e numero totale di parole, e che invece risulta essere maggiore per i disonesti (Colwell, Hisckock, & Memon, 2002; Colwell et al., 2013). Le ricerche sperimentali sono state molto positive, in quanto hanno permesso di raggiungere una buona accuratezza

anche con un training relativamente breve. Colwell et al. (2012), per esempio, hanno trovato che le persone che hanno partecipato ad un percorso di formazione sulla tecnica ACID della durata di sole 2 ore e mezza hanno raggiunto un'accuratezza nel valutare correttamente la credibilità delle persone intervistate del 70%. Ciò porta alla luce il potenziale di questa tecnica di intervista nei procedimenti giudiziari.

Strategic Use of Evidence

Lo *Stragetic Use of Evidence* (SUE, uso strategico delle prove) è stato introdotto da Granhag (2010) ed è un modello che si basa su un livello strategico ed uno tattico, che mira quindi a tenere in considerazione quali sono le strategie (e le contro-strategie) messe in atto dalle persone intervistate (cfr. Granhag & Hartwig, 2015). L'aspetto strategico è di natura prettamente psicologico-sociale, e ha come focus il comportamento e i processi cognitivi dell'intervistato e la sua relazione interpersonale con l'intervistatore. Questa parte del modello è quindi indipendente dal caso sottoposto ad indagine giudiziaria (allo specifico crimine) e consta dei seguenti 4 aspetti:

1. La percezione che il sospettato ha riguardo a quante (e quali) prove l'intervistatore possiede
2. Le contro-strategie messe in atto dal sospettato per fronteggiare le strategie utilizzate dall'intervistatore
3. Le risposte (verbali) fornite dall'indagato
4. I processi psicologici di cambio di prospettiva e di teoria della mente che si attivano nell'intervistato.

Gli aspetti tattici invece sono più concreti e sono le tecniche messe in atto dall'intervistatore al fine di ottenere la verità da parte dell'intervistato. Le tecniche che compongono la parte tattica della SUE[34] sono:

1. La raccolta di tutte le informazioni possibili riguardanti il caso sotto indagine
2. Il tipo di domande poste all'intervistato

[34] Per una descrizione più dettagliata della tecnica SUE si consiglia la lettura di Granhag (2010), Granhag, Strömwall, Willén, e Hartwig (2013), Granhag e Hartwig (2015) e Tekin, Granhag, Strömwall, e Vrij (2016).

3. I modi ed i tempi per rivelare all'intervistato le informazioni (prove) di cui l'intervistatore è in possesso.

Il cuore del modello è quello di non lasciar trasparire di quali informazioni l'intervistatore è in possesso. In questo modo, un sospettato di fatto colpevole si troverà a fare delle dichiarazioni incoerenti con le prove. Una volta che questo sarà accaduto l'intervistatore potrà far notare tali incongruenze all'indagato e chiedergli ulteriori spiegazioni, facendogli notare che nascondere delle informazioni (o fornire delle informazioni false) ha minato la sua credibilità. Alla persona intervistata viene quindi data la possibilità di spiegarsi meglio e di giustificare queste incongruenze. È qui che entra in gioco il fattore chiave della tecnica: una volta che il sospettato è messo alle strette tenderà a cambiare strategia ed a fornire indicazioni più precise perché teme che facendo il contrario risulterà nuovamente incongruente con quanto l'intervistatore già conosce. In realtà, in questa fase di confronto l'intervistatore possiede poche informazioni (a volte nessuna), ma il sospettato è spinto a credere il contrario. In questa maniera la persona che conduce l'interrogatorio riuscirà ad ottenere nuovi indizi (Tekin, Granhag, Strömwall, & Vrij, 2015). Purtroppo, la tecnica necessita che le forze dell'ordine siano di fatto in possesso di alcune prove (non divulgate tramite i media), rendendola quindi inapplicabile quando questo non è il caso. Quando invece delle prove sono disponibili, la SUE ha mostrato un grande potenziale[35], con punte di accuratezza dell'85% nel discriminare chi mente da chi dice la verità (Hartwig, Granhag, Strömwall, & Kronkvist, 2006). Le forze di polizia potrebbero quindi beneficiare di questa tecnica allo stesso modo di come potrebbero beneficiare dell'utilizzo dell'ACID.

Tecniche di intervista non strutturate

In letteratura sono presenti diverse tecniche di intervista (semi)strutturate utilizzabili dalle forze di polizia nello svolgimento delle investigazioni, sia per ottenere più informazioni possibili dalla fonte, che per valutarne la credibilità. Similmente, sono presenti delle tecniche da utilizzare quando l'indagato è collaborativo ma anche quando non lo è. Sono presenti però anche studi empirici che hanno valutato

[35] Per una review Hartwig, Granhag, & Luke (2014).

l'efficacia di alcune strategie specifiche, le quali però non sono ancora sfociate in protocolli standardizzati. Ciò nonostante, è importante esaminarne pregi e difetti.

Baseline

Si legge spesso, sia nella letteratura scientifica che in quella divulgativa, che al fine di fronteggiare le differenze interpersonali negli indizi di menzogna, sarebbe opportuno creare una baseline comportamentale dell'indagato. Ovvero, in una prima fase dell'intervista chiamata *baseline*, l'intervistatore dovrebbe porre all'intervistato alcune domande dalle quali si aspetta, certamente, una risposta onesta. Tali risposte verranno poi utilizzate come riferimento con cui confrontare le risposte che l'indagato darà alle domande relative all'evento sotto indagine (o domande *target*). L'idea di base di questo approccio è che ogni persona mostra indizi di inganno diversi: una può muovere di più le mani quando mente, un'altra quando è onesta. Quindi, al fine di valutare la credibilità di una persona non vanno ricercati specifici indizi di menzogna predeterminati. Al contrario, si valuterà la presenza di differenze comportamentali tra le due fasi dell'intervista, nel qual caso si potrà concludere che la persona intervistata sta mentendo. Uno degli aspetti più problematici della tecnica di *baselining* è che è spesso stato suggerito di creare una baseline con domande irrilevanti e risposte totalmente indipendenti dall'evento sotto indagine (cfr. Frank, Yarbrough, & Ekman, 2006). D'altra parte, come notano Ewens, Vrij, Jang, e Jo (2014) in questo modo si corrono forti rischi di falsi positivi. Infatti, è presumibile che non solo le persone che mentono, ma anche quelle che sono oneste, mostreranno differenze tra le due fasi dell'intervista. In effetti, il carico cognitivo ed emotivo, e soprattutto la posta in gioco, sono comprensibilmente differenti tra quando si parla di qualcosa di innocuo rispetto a quando si parla di qualcosa che è sotto indagine giudiziaria. Al fine di fare chiarezza, Ewens et al. (2014) hanno condotto un esperimento dove hanno chiesto a metà dei loro partecipanti di essere onesti e all'altra metà di mentire, dopo aver loro posto domande innocue utilizzate come *baseline* (alle quali tutti i partecipanti hanno risposto dicendo la verità). Come previsto, sia le persone oneste che quelle disoneste hanno mostrato delle differenze comportamentali tra le due fasi dell'intervista. Tali risultati hanno quindi mostrato che in effetti una *baseline* costruita in questo modo non è efficace e anzi bisognerebbe evitare di usarla al fine di evitare grossolani errori di giudizio. Ciò nonostante, questi risultati non bocciano totalmente la tecnica di

baselining, quanto piuttosto il modo in cui questa viene creata. In effetti, sia Ewens et al. (2014) che Caso e Vrij (2009) e Vrij (2008) suggeriscono che un diverso tipo di *baseline* potrebbe essere efficace. Ovvero, se la fase *target* e quella di *baseline* condividono contesto, posta in gioco, carico cognitivo ed emotivo (per quanto difficile sia creare un tale tipo di riferimento) allora la strategia potrebbe funzionare. Vrij (2008) ha chiamato questa *baseline* "paragonabile" (*comparable*), proprio per sottolineare la similitudine che dovrebbe esserci tra le due fasi. Ad oggi, sono disponibili due soli studi sperimentali che hanno testato l'efficacia di una *comparable truth baseline* (Palena, Caso, Vrij, & Orthey, 2018; Palena, Caso, Carlotto, De Mizio, & Marciali, 2017). Palena et al. (2018) hanno valutato le differenze sia verbali che non verbali presenti tra la fase di *baselining* e la fase *target*, ed hanno replicato i risultati di Ewens et al. (2014) secondo i quali una tecnica di *baselining* costruita con domande irrilevanti (chiamata "*small talk*") è inefficace. Inoltre, essi hanno trovato che quando si analizza il comportamento non verbale neanche una *comparable truth baseline* risulta essere efficace in quanto, nuovamente, sia le persone oneste che quelle disoneste mostravano delle differenze tra le due fasi dell'intervista. Vrij (2016) nota che ciò è probabilmente dovuto alla difficoltà di creare una *baseline* di tipo non verbale. Ciò nonostante un altro risultato interessante è disponibile. Quando ad essere analizzate erano le differenze di tipo verbale, allora una *comparable truth* baseline è risultata essere efficace, dal momento che solo coloro che stavano mentendo mostravano delle differenze tra la fase di baseline e quella target. Tali risultati, come anche Vrij (2016) sostiene, ci portano a concludere due cose: primo, la tecnica di *baseline* è molto delicata e non dovrebbe essere utilizzata dalle forze di polizia fino a quando non verrà sviluppata una strategia standardizzata pulita ed efficace; secondo, la tecnica di *baselining* ha del potenziale se l'investigatore si concentra sugli aspetti verbali delle dichiarazioni dell'indagato, ma è controproducente se ad essere analizzato è il comportamento non verbale. Purtroppo, ad oggi non sono presenti studi sull'accuratezza degli osservatori nel valutare la credibilità del mittente quando viene utilizzata la tecnica di *baselining.*

Porre domande inaspettate

La letteratura ha mostrato che solitamente quando siamo onesti riferiamo un maggior numero di dettagli rispetto a quando mentiamo (DePaulo et al., 2003). In effetti, il

criterio "quantità di dettagli" della CBCA[36] è uno dei criteri più discriminativi (cfr. Caso & Vrij, 2009; Vrij, 2008). D'altra parte, bisogna anche considerare che una persona che sa che la sua credibilità verrà valutata, tenderà a preparare una storia credibile e coerente e ad immaginare quali domande le verranno poste (Granhag, Andersson, Strömwall, & Hartwig, 2004; Granhag, Ströwall, & Jonsson, 2003; Hartwig, Granhag, & Strömwall, 2007). Come conseguenza, la differenza della quantità di dettagli tra bugia e verità tenderà a diminuire (DePaulo et al., 2003). Vrij (2015b) nota che questa preparazione è efficace solo se l'indagato è in grado di anticipare correttamente quali domande gli verranno poste. In questi casi, una persona che mente sarà sottoposta ad un minore carico cognitivo quando risponderà alle domande che si è preparato, ma ad un maggiore carico cognitivo quando risponderà a quelle per le quali non si è preparato, rispetto ad una persona onesta, la quale invece non dovrebbe mostrare grosse differenze rispetto ai due tipi di domande. Infatti, una persona onesta risponde alle domande inaspettate così come risponde a quelle previste: ricordando. Al contrario, le persone che mentono rispondono alle domande previste ricordando la storia che hanno preparato, ma devono creare sul momento risposte credibili alle domande inaspettate. Questo ha portato alcuni ricercatori a valutare l'efficacia delle domande inaspettate quando l'obiettivo primario è proprio quello di valutare l'onestà del teste. Tale strategia è risultata efficace sia quando le persone venivano intervistate rispetto ad eventi passati (Lancaster, Vrij, Hope, & Waller, 2013) che quando il focus dell'intervista era riferito ad intenzioni future (Warmelink, Vrij, Mann, Jundi, & Granhag, 2012). Ciò che funziona in questa tecnica è quindi il fatto di porre domande non previste, le quali possono essere sviluppate in maniera diversa a seconda della specifica situazione. A titolo di esempio, Leins e colleghi (2011) hanno chiesto ai loro partecipanti di visitare una stanza e li hanno poi interrogati. Successivamente ai partecipanti è stato chiesto di descrive la stanza nella quale erano stati, prima a voce e poi disegnandola. I risultati hanno mostrato che le persone che mentivano si sono contraddette di più delle persone oneste. Ovvero, c'era

[36] CBCA è l'acronimo di *Criteria Based Content Analysis* che, insieme ad un colloquio strutturato ed al confronto con gli elementi di prova, compone lo *Statement Validity Assessment* (SVA). Lo SVA mira a valutare la credibilità di un resoconto sulla base del contenuto verbale. L'analisi del contenuto si basa proprio sulla CBCA, che è quindi la componente centrale della SVA. La CBCA è composta da 19 criteri i quali, se presenti, aumentano l'attendibilità del resoconto. Per una descrizione dettagliata cfr. Caso & Vrij, 2009, De Leo, Scali, e Caso, 2005, e Vrij, 2008.

una minore congruenza tra quanto riportato a voce e quanto riportato con il disegno. Nell'esempio precedente l'aspetto non previsto dagli intervistati risiedeva nella *modalità* con la quale veniva posta la domanda. D'altra parte, può essere anche il contenuto della domanda stessa ad essere inaspettato. Ad esempio, Vrij e colleghi (2009) hanno chiesto ai loro partecipanti di andare a pranzo insieme (onesti) o fare finta di averlo fatto (disonesti). I partecipanti hanno avuto persino tempo di preparare la storia insieme. Durante l'intervista però, l'intervistatore ha posto alcune domande che i partecipanti non si aspettavano, ad esempio "*Chi ha finito il pasto per primo?*", oppure "*Rispetto al vostro tavolo e al posto dove eri seduto, dov'erano i commensali più vicini?*". Utilizzando questo approccio è stato possibile raggiungere un'accuratezza di circa l'80%. Quindi, viste le potenzialità di questa tecnica, sarebbe opportuno cercare di sviluppare un protocollo efficace che ne faccia un uso corretto.

Incoraggiare a dire di più

Caso e Vrij (2009) e De Leo, Scali, e Caso (2005) sottolineano l'importanza del porre domande efficaci agli intervistati al fine di ottenere il maggior numero di informazioni (accurate) possibile, mettendo al contempo il teste (o sospettato) a proprio agio ed evitando di suggestionarlo. In effetti, riuscire ad ottenere delle risposte elaborate da parte dell'intervistato rende più semplice valutare l'onestà della persona interrogata. In primo luogo, perché c'è più materiale che è possibile analizzare: è diverso valutare la credibilità di una persona che risponde "sì" o "no" rispetto ad una che racconta in dettaglio l'evento accaduto. In secondo luogo, incoraggiando a dire di più si mettono in difficoltà le persone che mentono, in quanto il loro carico cognitivo aumenta in conseguenza alla richiesta di riportare ulteriori (e quindi false, probabilmente create sul momento) informazioni. Inoltre, una credenza che hanno la maggior parte delle persone è che per essere credibili bisogna raccontare una storia in modo chiaro, coerente e privo di contraddizioni, quando la ricerca ha dimostrato che una produzione non strutturata è invece indice di verità (Caso e Vrij, 2009). Terzo, come si è visto nella sezione relativa all'ACID di Colwell et al. (2013), la richiesta di dire di più permette alle persone oneste di ricordare ulteriori dettagli (veri) del caso sotto investigazione. Uno dei metodi utilizzati è quello di fare ascoltare o leggere alla persona intervistata un *model statement*. Ovvero, le viene chiesto di leggere o ascoltare l'intervista di una terza persona che ha fornito un adeguato numero di

dettagli. In questa maniera si stimola l'intervistato a dire di più e ad aggiungere ulteriori informazioni. La ricerca ha mostrato che in effetti questa richiesta ha del potenziale, e che i dettagli aggiunti dalle persone oneste erano più plausibili di quelli aggiunti dalle persone disoneste. Di conseguenza, attraverso specifiche analisi statistiche, è stato possibile raggiungere un'accuratezza nel distinguere gli onesti dai disonesti dell'80% (Leal, Vrij, Warmelink, Vernham, & Fisher, 2015, citato in Vrij, 2016).

Dettagli verificabili

Sebbene buona parte delle ricerche mostri che quando diciamo la verità forniamo più dettagli rispetto a quando diciamo una bugia, ci sono alcuni casi in cui questa differenza diminuisce, ad esempio quando abbiamo la possibilità di prepararci all'intervista (DePaulo et al., 2003). Questo accade perché un indagato sa che inserendo maggiori dettagli risulterà più credibile (Vrij, 2016). Inoltre, la letteratura ha anche mostrato che addestrare la persona intervistata rispetto al come risultare credibile, rende vana l'efficacia di strumenti di analisi verbale come la CBCA (Caso, Vrij, Mann, & De Leo, 2006; Vrij, Akehurst, Soukara, & Bull, 2002, 2004). D'altra parte, un indagato può aggiungere dettagli quando sa che questi non potranno essere verificati dalla polizia e quando questi, come abbiamo visto nella tecnica SUE, non contraddicono le prove di cui sono in possesso gli inquirenti. Da questo ragionamento è nata l'ipotesi di Nahari, Vrij, e Fisher (2014a, 2014b) di esplorare se l'ammontare di dettagli verificabili differisce tra persone che dicono la verità e persone che dicono una bugia. La tecnica è risultata efficace sia nel primo studio pubblicato che in studi successivi. Inoltre, l'efficacia della tecnica sviluppata da Nahari, Vrij, e Fisher (2014a) è risultata maggiore quando è stata analizzata la frequenza dei soli dettagli verificabili rispetto a quando è stato analizzato il rapporto tra dettagli verificabili e dettagli non verificabili. Nel primo caso, gli autori hanno trovato un'accuratezza del 79%, mentre nel secondo del 71%. Forse ancora più interessante, l'accuratezza calava al 63% quando non si differenziava tra dettagli verificabili e non verificabili, ed ogni differenza tra onesti e disonesti spariva quando si analizzavano solo quelle informazioni che non era possibile verificare. Questi risultati indicano che potrebbe essere utile analizzare le dichiarazioni del sospettato sia sulla base dei dettagli non verificabili che, soprattutto, di quelli che invece è possibile controllare.

Interviste di coppia

Sebbene buona parte della letteratura sulle tecniche di intervista e valutazione della credibilità si concentri sul singolo individuo, sia esso sospettato, vittima o testimone, è evidente come nella vita reale i crimini siano commessi anche da coppie o gruppi di individui. Caso e Vrij (2009) notano che questo è un importante gap presente nella letteratura scientifica in quanto è importante sviluppare dei protocolli di intervista specifici per queste situazioni. Uno dei primi studi a riguardo è stato svolto da Vrij et al. (2012). Gli autori notano che nella pratica giudiziaria si tende a separare le persone ed ad intervistarle individualmente, quando potrebbe essere più produttivo interrogarle contemporaneamente. Nel loro esperimento, gli autori hanno intervistato 43 coppie di partecipanti, metà delle quali mentivano e metà dicevano la verità. Sulla base della letteratura precedente che suggerisce come le persone oneste adottino come strategia quella di "dire semplicemente le cose per come sono accadute" ed i bugiardi quella di "raccontare una storia semplice", gli autori hanno ipotizzato che le coppie di individui che dicevano la verità si sarebbero interrotte l'un l'altra più spesso delle coppie di bugiardi. Allo stesso modo, gli autori hanno ipotizzato che gli onesti avrebbero corretto il loro compagno più spesso dei disonesti, così come che avrebbero aggiunto più informazioni a quanto detto dal loro partner. Tutte queste previsioni sono state supportate e sono state interpretate prevalentemente sulla base della teoria della memoria transattiva (Wegner, 1987). Questo primo studio ha poi stimolato lo sviluppo di ulteriori paradigmi sperimentali volti, da una parte, alla ricerca di specifici indizi di menzogna nell'interrogatorio di coppie o gruppi; dall'altra, all'obiettivo di sviluppare una tecnica di intervista che possa essere efficacemente utilizzata in queste specifiche situazioni. Dalla revisione sulle interviste di gruppo di Vernham e Vrij (2015) si possono trarre due conclusioni principali. Primo, quando vengono intervistate delle coppie (o dei gruppi) di sospettati, diventano disponibili degli indizi di menzogna di tipo prevalentemente sociale e comunicativo (cfr. Driskell, Salas, & Driskell, 2012). Ovvero, quando ad essere intervistate erano delle persone che mentivano, mancavano tutti quegli indizi e segnali di interazioni socio-comunicativa che invece appaiono quando sono due persone oneste ad interagire. Tra questi, gli autori riportano: porsi domande a vicenda, finire la frase del proprio partner, e guardarsi. Ciò nonostante, anche aspetti di tipo cognitivo sono in gioco, come la

codifica ed il recupero di una memoria condivisa. Secondo, nel caso in cui ci sia più di un sospettato, le interviste collettive non dovrebbero essere sostituire quelle individuali. Piuttosto, entrambe le tecniche dovrebbero essere utilizzate, in quanto ognuna di esse è di aiuto per il raggiungimento dell'obiettivo finale. Tra le strategie che sono state esplorate pare avere rilievo l'utilizzo di un approccio di presa di turno forzato. In Vernham, Vrij, Mann, Leal, e Hillman (2014), questa procedura è stata utilizzata interrompendo, ogni venti secondi, la persona che stava parlando, per chiedere al relativo partner di continuare la risposta. I risultati hanno mostrato che le coppie di bugiardi trovavano più difficile fronteggiare questa richiesta, in quanto non riuscivano a continuare il racconto iniziato dal proprio partner in modo fluido, tendevano ad aspettare di più prima di rispondere e, per fronteggiare questo aumento nel tempo di latenza, ripetevano l'ultima parte della frase detta dal loro compagno. I risultati di Vernham et al. (2014), quelli di altri studi (Jundi et. al, 2013; Vernham et al., 2014) e la prima review sulle interviste di gruppo (Vernham & Vrij, 2015), hanno quindi mostrato che tale tecnica di intervista ha del potenziale, in quanto permette la comparsa di indizi di menzogna altrimenti assenti nelle interviste dei singoli individui, quali ad esempio segnali comunicativi e di interazione sociale. Questi risultati sottolineano ancora una volta l'importanza degli indizi verbali e sociali (Buller & Burgoon, 1996) nella valutazione della credibilità, nonostante spesso venga dato invece maggior rilievo agli aspetti non verbali (Global Deception Research Team, 2006). D'altro canto, la letteratura ha chiaramente mostrato la migliore efficacia dei primi rispetto ai secondi. Infatti, anche Palena (2015) ha trovato come nelle interviste di coppia gli indizi di tipo non verbale abbiano avuto poca efficacia. Rimane comunque importante e di potenziale utilità continuare la ricerca sulle interviste di gruppo al fine di sviluppare un protocollo strutturato.

Conclusioni

A partire dal 2012, la letteratura scientifica sulla raccolta di informazioni e sulla valutazione della credibilità ha visto un netto cambio di rotta. Si è passati dalla ricerca di indizi di menzogna allo sviluppo di tecniche di intervista investigativa. Sebbene lo scopo sia in entrambi i casi quello di mettere in evidenza l'inganno, è comunque presente una differenza di base. Nel primo caso, lo sforzo era diretto solo a valutare quali indizi potessero indicare la presenza di una bugia, spesso a prescindere dal

contesto nel quale veniva condotta l'intervista e dalla struttura di quest'ultima. Nel secondo, si cerca di trovare dei metodi per far venire a galla tali indizi. Ciò è successo in quanto è parso chiaro, dopo diverse decadi di ricerca, il fatto che i marcatori di veridicità ed inganno sono deboli ed inaffidabili. Se da una parte tecniche investigative di successo, e con una solida base teorica, sono rimaste attive a tutt'oggi dopo più di 20 anni di pratica (come nel caso dell'Intervista Cognitiva), altre tecniche più mirate alla valutazione della credibilità hanno cominciato a comparire sulla scena internazionale. Tutte quante sono basate su tre presupposti teorici principali: i) il carico cognitivo che si viene a creare quanto si mente; ii) gli aspetti di interazione sociale tra la persona che interroga e l'interrogato; iii) le strategie messe in atto dalle persone intervistate per risultare credibili. Gli approcci e i protocolli che si basano su questi assunti teorici hanno mostrato un grosso potenziale, spendibile anche nel ramo applicato delle indagini di polizia giudiziaria. Di particolare rilievo, ad oggi, ci sono il protocollo ACID di Colwell et al. (2013) e l'uso strategico delle prove di Granhag (2010). Altri approcci, i quali mancano di una solida base teorica, si sono invece dimostrati inefficaci. È questo il caso particolare della BAI e della tecnica dei nove passi di Reid (Inbau et al., 2011). Sarebbe quindi opportuno che le forze di polizia smettessero di usare tali protocolli investigativi al fine di evitare grossolani errori.

Inoltre, diverse ricerche particolarmente recenti hanno mostrato la potenziale efficacia di ulteriori strategie investigative, sebbene queste non siano ancora confluite in un vero e proprio protocollo standardizzato. Tra queste abbiamo visto ad esempio l'analisi dei dettagli verificabili, l'utilizzo di una *baseline* "paragonabile", il porre domande che sono inattese e l'incoraggiare la persona intervistata a dire di più di quanto non abbia già fatto. C'è poi il caso particolare delle strategie utilizzate per l'intervista di coppie o gruppi di individui, fino ad ora trascurate, le quali invece risultano di essenziale importanza nei procedimenti giudiziari. Alcune di esse hanno già dimostrato un impatto positivo sull'accuratezza dell'osservatore, che dalla letteratura sappiamo essere generalmente molto basso, circa il 54% (Bond & DePaulo, 2006, 2008). D'altra parte, ognuno di questi metodi è più spendibile in un contesto rispetto che in un altro. L'Intervista Cognitiva, la PEACE, la ACID ed il *model statement,* possono essere utilizzate solo quando abbiamo di fronte una persona collaborativa, o che vuole mostrarsi tale. Se ci troviamo in questa posizione abbiamo

la possibilità di raggiungere un buon livello di accuratezza: tra il 70% se la decisione viene presa da un essere umano e l'80% quando la classificazione viene effettuata da analisi statistiche quali l'analisi discriminante (Leal et al., 2015). Purtroppo, nel caso di reticenza da parte dell'indagato questi approcci diventano inapplicabili. La SUE è meno sensibile a questo problema. Infatti, anche se il sospettato parla poco, è sufficiente che contraddica le prove di cui sono in possesso le forze dell'ordine. A questo punto, la tecnica SUE spingerà la persona indagata a fornire ulteriori (nuovi) dettagli i quali, a loro volta, permetteranno di raggiungere un livello di accuratezza che può arrivare anche all'85% (Hartwig et al., 2006). La tecnica di *baselining*, l'utilizzo di domande che il sospettato non ha previsto e l'analisi dei dettagli verificabili condividono la possibilità di fare confronti tra le varie dichiarazioni dello stesso soggetto. Purtroppo, ad oggi non sono presenti studi sull'impatto che la tecnica di *baselining* ha sull'accuratezza degli osservatori. Considerando però che in letteratura sono presenti dei lavori che hanno riscontrato delle differenze oggettive tra onesti e disonesti (Palena et al., 2017, 2018) è plausibile che chi interroga possa trarre benefici dal suo utilizzo. Ad ogni modo, la presenza di due singoli studi non è sufficiente a trarre conclusioni solide e affidabili, per cui è necessario che essi vengano replicati. L'utilizzo di domande non previste e dei dettagli verificabili hanno invece già dimostrato la loro utilità, dal momento che attraverso il loro utilizzo è possibile raggiungere classificazioni corrette circa nell'80% dei casi. Considerando la variabilità delle indagini, dei reati e delle persone intervistate, sarebbe auspicabile lo sviluppo di ulteriori protocolli di intervista investigativa che tengano conto, in modo armonioso, di tutti i risultati che sinora sono stati trovati. Se da una parte è vero che è impossibile creare una tecnica valida per ogni tipo di indagine, è vero anche che sarebbe utile sviluppare delle tecniche di intervista standardizzate da utilizzare a seconda dell'indagine in corso. Inoltre, nel farlo, i ricercatori dovrebbero cercare di sfruttare tutti i processi psicologici, sociali e comunicativi in gioco, evitando di concentrarsi solo su uno di questi aspetti. Lo sviluppo di tali protocolli permetterebbe sia un supporto investigativo nello svolgimento delle indagini, al fine di aumentare la probabilità di individuare tempestivamente un colpevole, che di tenere in considerazione i princìpi etici, i diritti e le necessità della persona intervistata.

La situazione internazionale sembra andare in questa direzione, dove non è raro trovare psicologi ed accademici (soprattutto in nord Europa) impiegati come *expert witnessess*. Questo è comprensibile quando si considera che le tecniche di intervista e di interrogatorio riassunte in questo elaborato paiono essere uno strumento efficace. La situazione italiana è al momento diversa, salvo per i casi di abusi sessuali su minori e, recentemente, quelli che riguardano donne maltrattate. Ad esempio, nel Regno Unito le istituzioni accademiche e le forze di polizia collaborano al fine di sviluppare specifiche linee guida e protocolli di intervista investigativa, mentre questo tipo di collaborazione è raramente presente in Italia. Sarebbe quindi auspicabile, anche nel nostro paese, una collaborazione tra accademia e forze di polizia.

Referenze

Ainsworth, P. B. (2002). *Psychology and Policing*. Cullompton: Willan.

Baldwin, J. (1992). *Video taping police interviews with suspects: An evaluation*. London: Home Office Police Department.

Blair, J. P., & Kooi, B. (2004). The gap between training and research in the detection of deception. *International Journal of Police Science & Management*, *6* (2), 77-83.

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

Bond, C. F., Jr., & DePaulo, B. M. (2008). Individual differences in judging deception: Accuracy and bias. *Psychological Bulletin, 134*(4), 477-492.

Buckley, J. P. (2013). The Reid Technique of Interviewing and Interrogation. In T. Williamson (Ed.), *Investigative Interviewing* (pp. 190-206). New York: Routlege.

Bui, D. C. (2015). *Using the Encoding Specificity Principle to Assess the Nature of the Secondary Memory Component of Working Memory*. Washington University in St. Louis.

Buller, D. B., & Burgoon, J. K. (1996). Interpersonal deception theory. *Communication theory, 6* (3), 203-242.

Caso, L., & Vrij, A. (2009). *L'interrogatorio giudiziario e l'intervista investigativa.* Bologna: Il Mulino.

Caso, L., Vrij, A., Mann, S., & De Leo, G. (2006). Deceptive responses: The impact of verbal and non-verbal countermeasures. *Legal and Criminological Psychology*, *11* (1), 99-111.

Cavedon, A., & Calzolari, M. G. (1999). Interrogare un testimone: tecniche di interrogatorio a confronto. *Rivista di Psicologia Giuridica*, *I/II*, 23-34.

Clarke, C., Milne, R., & Bull, R. (2011). Interviewing suspects of crime: The impact of PEACE training, supervision and the presence of a legal advisor. *Journal of Investigative Psychology and Offender Profiling*, *8* (2), 149-162.

Colwell, L. H., Colwell, K., Hiscock-Anisman, C. K., Hartwig, M., Cole, L., Werdin, K., & Youschak, K. (2012). Teaching professionals to detect deception: The efficacy of a brief training workshop. *Journal of Forensic Psychology Practice*, *12* (1), 68-80.

Colwell, K., Hiscock-Anisman, C., & Fede, J. (2013). Assessment criteria indicative of deception: An example of the new paradigm of differential recall enhancement. In B. S. Cooper, D. Griesel, & M. Ternes (Eds.), *Applied issues in investigative interviewing, eyewitness memory, and credibility assessment* (pp. 259-291). New York: Springer.

Colwell, K., Hiscock-Anisman, C. K., & Memon, A. (2002). Interviewing techniques and the assessment of statement credibility. *Applied Cognitive Psychology*, *16* (3), 287-300.

De Leo, G., Scali, M., & Caso, L. (2005). *La testimonianza: Problemi, metodi e strumenti nella valutazione dei testimoni.* Bologna: Il Mulino.

DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological Bulletin*, *129* (1), 74-118.

Dodson, C. S., Powers, E., & Lytell, M. (2015). Aging, confidence, and misinformation: Recalling information with the cognitive interview. *Psychology and aging*, *30*(1), 46.

Driskell, J. E., Salas, E., & Driskell, T. (2012). Social indicators of deception. *Human Factors, 54* (4), 577-588.

Ewens, S., Vrij, A., Jang, M., & Jo, E. (2014). Drop the small talk when establishing baseline

behaviour in interviews. *Journal of Investigative Psychology and Offender Profiling, 11* (3), 244-252. DOI: 0.1002/jip.1414

Fisher, R. P., & Geiselman, R. E. (1992). *Memory-enhancing techniques in investigative interviewing: The cognitive interview*. Springfield, IL: Thomas.

Fisher, R. P., Geiselman, R. E., Raymond, D. S., Jurkevich, L. M., & Warhaftig, M. L. (1987). Enhancing enhanced eyewitness memory: Refining the cognitive interview. *Journal of Police Science & Administration, 15*, 291-296.

Frank, M. G., Yarbrough, J. D., & Ekman, P. (2006). Investigative interviewing and the detection of deception. In T. Williamson (Ed.), *Investigative interviewing: Rights, research and regulation* (pp. 229–255). Cullompton, Devon: Willan.

Global Deception Research Team. (2016). A world of lies. *Journal of cross-cultural psychology*, *37*, 60-74.

Granhag, P. A. (November 18-19, 2010). The Strategic Use of Evidence (SUE) technique: A scientific perspective. High Value Detainee Interrogation Group (HIG; FBI). HIG Research Symposium: Interrogation in the European Union, Washington, DC.

Granhag, P. A., Andersson, L. O., Strömwall, L. A., & Hartwig, M. (2004). Imprisoned knowledge: Criminals' beliefs about deception. *Legal and Criminological Psychology*, *9* (1), 103-119.

Granhag, P. A., & Hartwig, M. (2015). The Strategic Use of Evidence Technique: A conceptual overview. A cognitive approach to lie detection. In P. A. Granhag, A. Vrij, & B. Verschuere

(Eds.), *Detecting deception: Current challenges and cognitive approaches* (pp. 231-252).

Chichester: John Wiley and Sons.

Granhag, P. A., Strömwall, L. A., & Jonsson, A. C. (2003). Partners in crime: How liars in collusion betray themselves. *Journal of Applied Social Psychology, 33* (4), 848-868.

Granhag, P. A., Strömwall, L. A., Willén, R. M., & Hartwig, M. (2013). Eliciting cues to deception by tactical disclosure of evidence: The first test of the Evidence Framing Matrix. *Legal and Criminological Psychology*, *18* (2), 341-355.

Gudjonsson, G. H. (2003). *The psychology of interrogations and confessions: A handbook.* Chichester: Wiley.

Hartwig, M., Granhag, P. A., & Luke, T. (2014). Strategic use of evidence during investigative interviews: The state of the science. In D. C Raskin, C. R. Honts, & J. C. Kircher (Eds.), *Credibility assessment: Scientific research and applications* (pp. 1-36). Waltham, MA: Academic Press.

Hartwig, M., Granhag, P. A., & Strömwall, L. A. (2007). Guilty and innocent suspects' strategies during police interrogations. *Psychology, Crime & Law*, *13* (2), 213-227.

Hartwig, M., Granhag, P. A., Strömwall, L. A., & Kronkvist, O. (2006). Strategic use of evidence during police interviews: When training to detect deception works. *Law and human behavior*, *30*(5), 603-619.

Henkel, L. A., & Coffman, K. J. (2004). Memory distortions in coerced false confessions: A source monitoring framework analysis. *Applied Cognitive Psychology*, *18*(5), 567-588.

Hernández-Fernaud, E., & Alonso-Quecuty, M. (1997). The cognitive interview and lie detection: A new magnifying glass for Sherlock Holmes?. *Applied Cognitive Psychology*, *11* (1), 55-68.

Horgan, A. J., Russano, M. B., Meissner, C. A., & Evans, J. R. (2012). Minimization and maximization techniques: Assessing the perceived consequences of confessing and confession diagnosticity. *Psychology, Crime & Law*, *18*(1), 65-78.

Horvath, F., Jayne, B., & Buckley, J. (1994). Differentiation of truthful and deceptive criminal suspects in behavior analysis interviews. *Journal of Forensic Science*, *39* (3), 793-807.

Inbau, F. E. (2013). *Essentials of the Reid technique*. Burlington, MA: Jones & Bartlett Publishers.

Inbau, F. E., Reid, J. E., Buckley, J. P., & Jayne, B. C. (2013). *Criminal interrogation and confessions, 5th edition*. Burlington, MA: Jones & Bartlett.

Jundi, S., Vrij, A., Mann, S., Hope, L., Hillman, J., Warmelink, L., & Gahr, E. (2013). Who should I look at? Eye contact during collective interviewing as a cue to deceit. *Psychology, Crime & Law*, *19*, 661-671. DOI: 10.1080/1068316X.2013.793332.

Kassin, S. M. (1997). The psychology of confession evidence. *American Psychologist, 52* (3), 221-233.

Kassin, S. M., & Fong, C. T. (1999). "I'm innocent!": Effects of training on judgments of truth and deception in the interrogation room. *Law and Human Behavior*, *23* (5), 499-516.

Lancaster, G. L., Vrij, A., Hope, L., & Waller, B. (2013). Sorting the liars from the truth tellers: The benefits of asking unanticipated questions on lie detection. *Applied Cognitive Psychology*, *27* (1), 107-114.

Leal, S., Vrij, A., Warmelink, L., Vernham, Z., & Fisher, R. (2015). You cannot hide your telephone lies: Providing a model statement as an aid to detect deception in insurance telephone calls. *Legal and Criminological Psychology*, *20*, 129–146. DOI:10.1111/lcrp.12017.

Leins, D., Fisher, R. P., Vrij, A., Leal, S., & Mann, S. (2011). Using sketch drawing to induce inconsistency in liars. *Legal and Criminological Psychology*, *16*(2), 253-265.

Mann, S., Vrij, A., & Bull, R. (2004). Detecting true lies: police officers' ability to detect suspects' lies. *Journal of Applied Psychology*, *89* (1), 137-149.

Mastroberardino S. (2011). *L'intervista cognitiva,* Carocci, Roma.

McGurk, B. J., Carr, M. J., & McGurk, D. (1993). *Investigative interviewing courses for police officers: An evaluation*. Home Office Police Research Group.

Memon, A., & Stevenage, S. V. (1996). Interviewing witnesses: What works and what doesn't. *Psychology*, *96,* 217-240.

Memon, A. A., Vrij, A., & Bull, R. (2003). *Psychology and law: Truthfulness, accuracy and credibility*. Chichester: John Wiley & Sons.

Milne, R., & Bull, R. (1999). *Investigative Interviewing: Psychology and practice.* Chichester: Wiley.

Moore, T. E., & Fitzsimmons, C. L. (2011). Justice imperiled: False confessions and the Reid technique. *Crim. LQ*, *57*, 509-542.

Nahari, G., Vrij, A., & Fisher, R. P. (2014a). Exploiting liars' verbal strategies by examining the verifiability of details. *Legal and Criminological Psychology*, *19* (2), 227-239.

Nahari, G., Vrij, A., & Fisher, R. P. (2014b). The verifiability approach: Countermeasures facilitate its ability to discriminate between truths and lies. *Applied Cognitive Psychology*, *28* (1), 122-128.

Palena, N. (2015). *Non-Verbal Communication and Mimicking Behaviours among Truth Tellers and Lying Couples*. (Unpublished Master's degree dissertation). Huddersfield University, Huddersfield, UK.

Palena, N., Caso, L., Vrij, A., & Orthey, R. (2018). Deception detection through small talk and comparable truth baselines. *Journal of Investigative Psychology and Offender Profiling, 1-9.* doi: 10.1002/jip.1495.

Palena, N., Caso., L., Carlotto, G., De Mizio, L., & Marciali, M. (2017). Efficacia della tecnica di *baselining* nella valutazione della credibilità. *Giornale Italiano di Psicologia, 4*, 905-916. doi: 10.1421/88773..

Pearse, J., & Gudjonsson, G. H. (1996). Police interviewing techniques at two south London police stations. *Psychology, Crime and Law*, *3* (1), 63-74.

Shawyer, A., & Milne, R. (2015). PEACE in fraud interviews: Assumptions and detection of guilt and the impact on interviewer behaviour. *Crime Prevention and Community Safety*, *17*(1), 30-46.

Shepherd, E. (1988). Developing interview skills. In P. Southage (Ed.), *New direction in police training* (pp. 170-188). London: HMSO.

Snook, B., Eastwood, J., & Barron, W. T. (2014). The next stage in the evolution of interrogations: The PEACE model. *Canadian Criminal Law Review*, *18*(2), 219.

Sooniste, T., Granhag, P. A., Strömwall, L. A., & Vrij, A. (2015). Statements about true and false intentions: Using the Cognitive Interview to magnify the differences. *Scandinavian Journal of Psychology*, *56*(4), 371-378.

Tekin, S., Granhag, P. A., Strömwall, L. A., & Vrij, A. (2016). How to make perpetrators in denial disclose more information about their crimes. *Psychology, Crime & Law*, *22* (6), 561-580.

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, *80* (5), 359-380.

Vernham, Z., & Vrij, A. (2015). A review of the collective interviewing approach to detecting deception in pairs. *Crime Psychology Review*, *1* (1), 43-58.

Vernham, Z., Vrij, A., Leal, S., Mann, S., & Hillman, J. (2014). Collective interviewing: A transactive memory approach towards identifying signs of truthfulness. *Journal of Applied Research in Memory and Cognition*, *3*, 12-20. DOI: 10.1016/j.jarmac.2014.01.001.

Vernham, Z., Vrij, A., Mann, S., Leal, S., & Hillman, J. (2014). Collective interviewing: Eliciting cues to deceit using a turn-taking approach. *Psychology, Public Policy, and Law*, *20* (3), 309-324.

Vrij, A. (2008). *Detecting lies and deceit: Pitfalls and opportunities*. Chichester: John Wiley and Sons.

Vrij, A. (2015a). Verbal lie detection. In P. A. Granhag, A. Vrij, & B. Verschuere (Eds.), *Detecting deception: Current challenges and cognitive approaches* (pp. 205-230). Chichester: John Wiley and Sons.

Vrij, A. (2015b). A cognitive approach to lie detection. In P. A. Granhag, A. Vrij, & B. Verschuere (Eds.), *Detecting deception: Current challenges and cognitive approaches* (pp. 205-230). Chichester: John Wiley and Sons.

Vrij, A. (2016). Baselining as a Lie Detection Method. *Applied Cognitive Psychology*, *30* (6), 1112-1119.

Vrij, A., Akehurst, L., Soukara, S., & Bull, R. (2002). Will the truth come out? The effect of deception, age, status, coaching, and social skills on CBCA scores. *Law and human behavior*, *26* (3), 261.

Vrij, A., Akehurst, L., Soukara, S., & Bull, R. (2004). Let me inform you how to tell a convincing story: CBCA and reality monitoring scores as a function of age, coaching, and deception. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement*, *36* (2), 113.

Vrij, A., Jundi, S., Hope, L., Hillman, J., Gahr, E., Leal, S., Warmelink, L., Mann, S., Vernham, Z., & Granhag, P. A. (2012). Collective interviewing of suspects. *Journal of Applied Research in Memory and Cognition*, *1*, 41-44.

Vrij, A., Leal, S., Granhag, P. A., Mann, S., Fisher, R. P., Hillman, J., & Sperry, K. (2009). Outsmarting the liars: The benefit of asking unanticipated questions. *Law and human behavior*, *33*(2), 159-166.

Warmelink, L., Vrij, A., Mann, S., Jundi, S., & Granhag, P. A. (2012). The effect of question expectedness and experience on lying about intentions. *Acta Psychologica*, *141* (2), 178-183.

Wegner, D.M. (1987). Transactive memory: A contemporary analysis of the group mind. In B. Mullen & G. R. Goethals (Eds.), *Theories of group behaviour* (pp. 185-208). New York: Springer-Verlag.

Activity report

Below are reported the activities done and all the outputs (conference presentations, scientific papers, etc.) produced during this three-years PhD project. The versions of the papers reported above (Appendices) may change from the published version. The lack of a Hyperlink within the text means either that the paper is not still ready or that the paper is not part of this PhD project.

In preparation

Caso, L., Palena, N., & Monticciolo, R. (In preparation). Utilità delle tecniche di intervista sviluppate in ambito accademico in contesti reali: Il parere delle Forze di Polizia di Bergamo. [IP1]

Palena, N., Caso, L., Granhag, P. A., Orthey, R., Monticciolo, R., & Vrij, A. (in preparation). Reciprocity: Does it helps in Scharff interviews? [IP2]

Caso, L., Palena, N., Vrij, A., & Melocchi, L. (in preparation). Implementing the use of a checklist with comparable truth and small talk baselines: The effect on laypersons' lie detection accuracy. [IP3]

Papers under review

Caso, L., Morganti, F., Palena, N., & Vrij, A. (under review). Dual-Task, working memory and deception. [UR1] (not part of this PhD Project)

Papers published in Italian peer-reviewed journals

Palena, N., Caso, L., Carlotto, G., De Mizio, L., & Marciali, M. (2017). Efficacia della tecnica di baselining nella valutazione della credibilità (Efficacy of the baseline technique when assessing credibility). *Giornale italiano di psicologia, 44*(4), 905-916. doi:10.1421/88773. [PA1]

Caso, L., & Palena, N. (2018). Tecniche di interrogatorio con soggetti adulti: una rassegna internazionale. *Rassegna Italiana di Criminologia, 2*, 130-140. [PA5]

Papers published in International peer-reviewed journals

Palena, N., Caso, L., Vrij, A., & Orthey, R. (2018). Detecting deception through small talk and comparable truth baselines. *Journal of Investigative Psychology and Offender Profiling, 15*(2), 124-132. doi:10.1002/jip.1495. [PA2]

Caso, L., Palena, N., Vrij, A., & Gnisci, A. (2019). Observers' performance at evaluating truthfulness when provided with Comparable Truth or Small Talk Baselines. Psychiatry, *Psychology and Law*. doi:10.1080/13218719.2018.1553471. [PA3]

Caso, L., Palena, N., Carlessi, E., & Vrij, A. (2019). Police accuracy in truth/lie detection when judging baseline interviews. *Psychiatry, Psychology, and Law*. doi: 10.1080/13218719.2019.1642258. [PA4]

Palena, N., Caso, L., & Vrij, A. (2019). Detecting Lies via a Theme-Selection Strategy. Frontiers in Psychology, 9(2775). doi:10.3389/fpsyg.2018.02775. [PA6]

Caso, L., Maricchiolo, F., Livi, S., Vrij, A., & Palena, N. (2018). Factors affecting Observers' Accuracy when Assessing Credibility: The Effect of the Interaction between Media, Senders' Competence and Veracity. *The Spanish Journal of Psychology, 21,* 1-10. doi:10.1017/sjp.2018.54. [PA7]

Costa, M., Simone, A., Vignali, V., Lantieri, C., & Palena, N. (2018). Fixation distance and fixation duration to vertical road signs. *Applied Ergonomics, 69,* 48-57. doi: 10.1016/j.apergo.2017.12.017. [PA8]

Orthey, R., Palena, N., Vrij, A., Meijer, E. H., Leal, S., Blank, H., & Caso, L. (2019). Effect of time pressure on strategy selection and strategy execution in forced choice testing. *Applied Cognitive Psychology, 33*(5), 974-979. doi: 10.1002/acp.3592 [PA9].

Italian conference presentations

Palena, N., & Caso, L. (2017). *Efficacia delle tecniche di intervista in ambito investigativo*. Paper presented at the III convegno Nazionale di Psicologia Giuridica, Milano, Italy. [CP1]

Caso, L., Palena, N., & Pontigia, L. (2018). *Analisi dell'accuratezza degli operatori della giustizia nel riconoscimento della falsa testimonianza attraverso il baselining*.

Paper presented at the XV Congresso dell'Associazione Italiana di Psicologia (sezione Psicologia Sociale), Bari, Italy. [CP2]

International conference presentations

Palena, N., Caso, L., Vrij, A., & Orthey, R. (2018). *Interviewing suspects with small talk and comparable truth baselines.* Paper presented at the EAPL, Turkey, Finland. [CP3]

Italian poster presentations

Pontigia, L., Palena, N., & Caso, L. (2018). I silenzi nella testimonianza minorile. Poster presented at the XV Congresso dell'Associazione Italiana di Psicologia (sezione Psicologia Sociale), Bari, Italy. [PP1]

International poster presentations

Pontigia, L., Palena, N., & Caso, L. (2018). Evaluating children's statements credibility: a comparison between Criteria-Based content analysis and reality monitoring. Poster presented at the EAPL, Turku, Finland. [PP2]

Index