2021

Deception Detection and Eye-Tracking

HOW BIASES AFFECT OUR VERACITY JUDGEMENT ABILITY

Tony Jungfer

Tutor: Slijkhuis, P.J.H. (BMS)

Abstract

The general ability of humans in detecting deception is just a little above chance and often based on luck or guessing. A major factor for this low veracity judgement ability are false biases people hold about deception cues. The goal of the current research was to examine the impact of different levels of knowledge about significant deception cues and biases on veracity judgement ability. To answer this question a comparative between-subject experimental research was conducted. Participants needed to interview a suspect in a mock-crime scenario while wearing eye-tracking glasses. Participants got randomly assigned to a training (N = 20) and no-training (N = 21) condition. A bias questionnaire was filled out to assess and differentiate between the levels of knowledge of both groups. To assess and compare the veracity judgement ability the fixation count on areas of interest were measured during the interview, as well as the lie detection accuracy through a questionnaire after the interview. There were two key findings of the present research. First, a short training was already enough to dismantle peoples biases and thereby heighten their conscious understanding of existing biases and correct deception cues. However, it was not enough to improve the practical ability to detect lies, called veracity judgement ability. This led to the conclusion that veracity judgement is a skill relying on long years of experience rather than theoretical training. Future research might be required to assess different forms and time spans of training needed to improve people's veracity judgement ability.

Keywords: Deception Detection, Veracity Judgement Ability, Biases, Eye-Tracking, Crime Scenario, Interview

Introduction

Two lies every day. This is the average number of lies a person tells each day, making deception a major aspect of human social behaviour and interaction (Ioannou, & Hammond, 2015). Even though deception is a common phenomenon, human deception skills are lacking with a deception detection accuracy of 54% (Bond, 2008; Hartwig, et al., 2004). This implies that the general ability of humans in detecting deception is just a little above chance and often based on luck or guessing (Bond, 2008; Hartwig, et al., 2004). Most daily lies people encounter are harmless, but some lies can have severe consequences for individuals and society as a whole (Wu et al., 2018). For example, a branch of criminals deceive individuals by selling fake products at their doorstep, luring them with special offers, and costing the victims a lot of money (Polizei-Beratung, 2020). Another example would be lying in court, which could affect the sentence by letting a perpetrator go free or it could lead to the framing of an innocent person (Wu et al., 2018). A major factor that influences the deception detection accuracy are false biases people hold about deception cues (Hartwig, et al., 2004). Deception cues are anomalies in the verbal and non-verbal behaviour of liars which in comparison only rarely occur in truth tellers. Therefore these cues are useful indications to detect deceit (DePaulo, et al., 2003). Due to a lack of knowledge or misinformation people direct their attention to false cues, reducing their veracity judgement ability (Hartwig, et al., 2004). Consequently, to reduce this misinformation a correct definition of deception and its cues is necessary.

Deception Cues

One definition of deception is that it is "a deliberate attempt to mislead others" (DePaulo, et al., 2003). This definition implies a conscious and active choice to deceive another person, excluding people who miscommunicate due to mistakes. Deceit often happens in face-to-face conversations between two or more individuals, for example, in a suspect interview or between friends. Due to this direct communication, there are enough possibilities for verbal and non-verbal cues (DePaulo, et al., 2003).

Regarding verbal cues, deceivers hesitate more, take longer pauses to answer and tend to have a higher pitch in their voice (Hartwig, et al., 2004; Zuckerman et al., 1981). Furthermore, the preparation of a deceiver has an influence on the cues the deceiver shows (DePaulo, et al., 2003). On the one hand, a deceiver might control their stories too much, emitting a feeling of rehearsal, less spontaneous interaction, and answers. On the other hand, a deceiver can also be underprepared, giving inconsistent and contradicting information

(DePaulo, et al., 2003). To conclude, no matter the preparation of the deceiver, there are always verbal cues a deceiver can show.

Non-verbal cues are a rich source of information especially in regards to people's intentions and deceptive behaviour (Runeson & Frykholm, 1983; Sebanz & Shiffrar, 2009). Common believed deception cues are non-verbal facial expressions, like increased gaze aversion and smiling, as people believe they are mostly out of conscious control (Hartwig, et al., 2004). Ekman and Friesen (1969), however, found that facial expressions are more under conscious control than lay-people think. Furthermore, the face can send many messages in a short time, making it even more difficult to identify deception cues. Contrary, the upper-face area, including cheeks, eyes and brows seem to be significant deception cues. The active control of this area is considered to be more difficult and thereby leading to more cues to detect deceit (DePaulo, et al., 2003; Porter, et al., 2012). An example for this is a genuine smile or Duchenne smile, which next to the raised mouth corners and the lips also includes the "orbicularis oculi surrounding the eyes, which pull the cheek up while slightly lowering the brow" (DePaulo, et al., 2003; Porter, et al., 2012). In a fake smile only the mouth corners and the lips get raised, as people only have these under their conscious control (Ekman et al., 1990). The body on the other hand is, in general, less under conscious control then the face, therefore being a more consistent channel to detect deceit (Ekman & Friesen, 1969). The most prominent significant non-verbal deception cues of the body are the increase in arm, hand, leg, and foot movements, as well as fidgeting (DePaulo, et al., 2003).

Verbal and non-verbal cues can be influenced by different factors, like emotions (DePaulo, et al., 2003; Ekman & Friesen, 1969). Deceivers will try to hide their real emotions when deceiving others and feign an emotion that supports their lie. Possible emotions when telling a lie are fear of being caught, the shame of telling a lie, or the delight of being able to deceive someone (DePaulo, et al., 2003). This active suppression of emotions leads to an unconscious micro expression of the real emotions, leading to emotional cues to detect the deceit, also called emotional leakage (Ekman & Friesen, 1969). When the intensity of the repressed emotion is high, emotional leakage occurs more frequently and lasts longer (Porter, et al., 2012). These emotional leakage cues are behaviours related to the emotion that the liar really feels. For example, when the liar feels shame, he might hold less eye contact and communicate more indirectly. When the liar however experiences fear, he might fidget more or has a more shacking and nervous voice (DePaulo, et al., 2003). To summarize, non-verbal emotional cues are unconscious, micro-behaviours of the real emotion the liar is trying to hide

(DePaulo, et al., 2003). Now that deception and its cues have been clearly defined, the differences of people in their deception detection abilities can be specified.

Deception Detection

As already mentioned the general deception detection ability of humans is just above chance, but there are experts in secret services and law enforcement with higher success rates with around 70% - 80% (Bond, 2008; Ekman & Friesen, 1969; Sebanz & Shiffrar, 2009; Vrij, 2004). Therefore, an interesting aspect needs to be understood, namely the difference in what experts do differently than lay-people. One main difference is that experts do not rely on intuition, but actively process information and use their experience, relying on successful schemas they used in past encounters with deceivers to detect deception (Bond, 2008). It seems that experts attend more to non-verbal information and are able to identify these cues faster (Bond, 2008; Sebanz & Shiffrar, 2009). Common schemas are fixating the face, especially lips, eyes, and cheeks and fixating movements in arm and leg areas (Bond, 2008; DePaulo, et al., 2003).

There are studies suggesting that experts' deception detection accuracy is only at chance level, too (Ekman & O'Sullivan, 1991; Vrij, 1993; Vrij, 2004). The limitation about these studies is that they have been conducted in a laboratory, therefore the stakes were not very high and there was no risk in being caught. However, Bond (2008) found a 70% - 80% deception detection accuracy in experts in a laboratory setting. Nonetheless, in a naturalistic suspect interview the stakes are higher, which could lead to more potential cues (Hartwig, et al., 2004). Lies in these high stakes scenarios, like covering crime, murder or terrorist intentions, are often accompanied by strong emotion such as hate, fear or excitement. These strong emotions are difficult to supress and are very prone to emotional leakage (Porter & ten Brinke, 2010). An example for this would be terrorists that need to hide their hate and resentment towards their victims while having a friendly chat with the airport staff (Porter & ten Brinke, 2010). These scenarios and emotions are almost impossible to replicate in a laboratory setting. Furthermore, in real suspect interviews, experts can use different deception detection strategies, like asking follow-up questions, demand elaborations and clarifications (Hartwig, et al., 2004).

It is not only important to know what experts do to perform better in deception detection, but also what lay-people do wrong. One major aspect are the biases or wrong believes people have about deception cues (Hartwig, et al., 2004). Lay-people seem to focus their attention more on verbal cues, like speech disturbances, longer pauses and holes in the

story, when screening for lies (Vrij, 2000). There is a mismatch between the actual deception cues and the biases, leading to misapplied attention and reduced accuracy in deception detection (Hartwig, et al., 2004). Ekman and Friesen (1969) predicted that a liar is more successful when the target of the lie is not even aware or considering the fact he is being lied to. This can be connected to the honesty effect, the tendency of wanting to believe that someone is truthful (Hartwig, et al., 2004). Having an honesty effect about a person increases the chance that a lie gets by undetected and lay-persons seem to exhibit the honesty effect more than, for example, police officers (Vrij, 1993). The opposite of the honesty effect is also possible, which is a high suspicion or belief that someone is lying (Hartwig, et al., 2004). Overall there seems to be a discrepancy between the cues experts and lay-people attend to, which possibly explains the difference in deception detection accuracy. A method to further investigate the influence of people's biases about deception on attention might be the technology of eye-tracking, as visual attention is closely linked to eye-movement (Chun, & Wolfe, 2001).

The Eye-Mind Link and Eye-Tracking

"What you see is determined by what you attend to" (Chun, & Wolfe, 2001) and attention is determined by once goals and the personal relevance of the information. When a person wants to detect deceit, their attention and what they see will be limited by what they personally deem as important (Chun, & Wolfe, 2001). This makes them vulnerable to personal biases which shifts the focus and gaze away from the real cues. Furthermore, people have the unconscious drive to move the eyes to the stimulus that is currently processed, known as the eye-mind link (Carter & Luke, 2020). Coming back to deception detection, the use of eye-tracking can be very useful in researching veracity judgement ability. We can gain an understanding of the subject's decision making processes and visual attention by tracking the cues they focus on. For example it can be tracked on what cues the subjects focused when they rejected or accepted that someone is lying (Carter & Luke, 2020). Bond (2008) already used eye-tracking for this purpose and since 2008 eye-tracking evolved, becoming more reliable and valid, especially in naturalistic settings (Carter & Luke, 2020). For these reasons, eye-tracking is considered a useful tool in measuring people's veracity judgement ability.

Eye-tracking is a method, predominately used in social sciences, which allows recording of eye-movements across time (Carter & Luke, 2020). Although it is an older technology, the usage in deception detection is new. It provides information about the attentional shift and gaze behaviour of the participants (Carter & Luke, 2020). Gaze behaviour

is constructed of two components, the fixation on stimuli and saccades or rapid eye movements between fixations (Frischen et al., 2007). As raw fixations and saccades happen in milliseconds, a large amount of data gets generated. Due to this clusters of fixations get observed rather than single fixation points (Carter & Luke, 2020). Lastly, eye-tracking has the advantage of being non-invasive and unobtrusive, employing external cameras or the use of glasses (Borys & Plechawska-Wójcik, 2017).

Current Research

The goal of the current research is to establish whether it is possible to improve veracity judgement ability through a training about deception detection biases. Wrong misconceptions or biases are considered a main factor for the general low veracity judgement ability of people, consequently resolving those should improve people's veracity judgement ability (Hartwig et al., 2004). A theoretical training about correct and significant deception cues should be enough to increase people conscious awareness about deception, as biases are often a result of lack of knowledge or misinformation. The deception cues used for the current research are mainly based on Bond (2008), DePaulo et al. (2003) and Vrij (2000) findings, namely the non-verbal cues of the upper face, legs and arms. Veracity judgement ability can, on the one hand, be measured through the number of accurately detected lies, also called deception detection accuracy. Therefore, higher level of knowledge through training should increase the number of accurately detect lies (Hartwig et al., 2004). On the other hand, due to the eye-mind link we can also gain an understanding about veracity judgement through tracking the visual attention and decision making process related to eye-movements (Carter & Luke, 2020). Eye-tracking with the tool areas of interest and the metric fixation count is therefore of particular importance to measure veracity judgement ability. Areas of interest is a tool that only collects metrics in specific chosen regions, like significant deception cues (Farnsworth, 2020). The metric used is the number of fixations that are counted in each area of interest (Farnsworth, 2020).

Research Question

Based on the information gathered in the literature the following research question was compiled "What is the impact of the level of knowledge about significant deception cues and deception detection biases on veracity judgement ability of suspect interviewers?"

Hypotheses:

H1: The trainings group has a more conscious understanding of existing biases than the control group.

H2: The trainings group has a higher deception detection accuracy than the control group.

H3: The trainings group has a higher number of fixations on arm and leg areas, as well as upper face areas than the control group.

Method

Participants

The participant population consisted of 48 participants. Due to an insufficient gaze sample of around 50%, seven participants were excluded, which led to a final sample of 41 participants. The mean age was 22 years, and the age range was between 20 and 34 years. The participants included 20 men, 20 women and 1 non-binary. Of these participants 6 were Dutch, 33 were German and 2 were from other countries. Ethical approval was obtained before recruiting participants. The study was advertised through a participant's credit system (SONA) and social media posts. Participants were compensated with a treat and one SONA credit if applicable for their time in the half an hour long study.

Design

This research was a comparative between-subject experimental design, with one independent and two dependent variables. The independent variable was the level of knowledge and training of the participants. There was an 'untrained' control group, next to the trained experimental group. The dependent variable is veracity judgement ability which is measured through deception detection accuracy and fixation counts on areas of interest or correct deception detection cues.

Materials

For this research, a couple of materials were required. First of all, a HP Pavilion x360 Convertible 14-ba1xx laptop was used for the participant to fill out the consent form (see Appendix A), their demographics, a truth-or-lie-questionnaire based on the interview guide (see Appendix E), a questionnaire in regard to biases of the participants (see Appendix B) and the NASA-TLX (Appendix C). Qualtrics is the platform through which these forms were filled out. The participants also got a crime scenario with all the important information they needed to know and their role or task (see Appendix D). Furthermore, they got a question guide with example questions for the semi-structured interviews (see Appendix E). The researchers also got a pre-prepared scenario for their role as a witness and perpetrator (see Appendix F). The researchers also had a protocol sheet and a training sheet with them to manage the research (see Appendix G and Appendix H).

The eye-tracking glasses that were used are the Eye Tracking Tobii Pro Glasses 2. They are one of the most widely used and efficient eye tracking glasses (Farnsworth, 2020). The glasses are unobtrusive and also suitable for everyday use which fits our methods best.

Procedure

In the control group the participants were asked to sign the consent form, before the experiment started. Next, the researcher handed the participants the documents with the description of the scenario and role. The participants now had 10 minutes to study this role. After the eye tracker had been set up and everyone had memorised their role, the experiment began. During the experiment, the participants played the role of the interrogator. Their goal was to find out if the suspect is guilty or not. For this purpose, they had been given a framework of questions by the researcher. In addition to that, they were allowed to ask follow-up questions to the suspect. After the interrogation, the participants were debriefed and asked to fill out the truth-or-lie-questionnaire, bias-questionnaire and the NASA-TLX.

In the experimental group before the experiment started, the participants were asked to give written consent for taking part in the study. In the experimental group, the participants were asked to fill out a short bias questionnaire and received a short training (see Appendix H) based on that. During this part, the researcher discussed the right beliefs and wrong assumptions about deception detection with the participant. Next, the researcher handed the participant the documents with the description of the scenario and role. The participants had 10 minutes to memorise this role. After the eye tracker had been set up, the experiment began. During the experiment, the participants played the role of the interrogator. Their goal was to find out if the suspect is guilty or not. For this purpose, they had been given a framework of questions by the researcher. In addition to that, they were allowed to ask follow-up questions to the suspect. After the interrogation, the participants were debriefed and asked to fill out the truth-or-lie-questionnaire, bias-questionnaire and the NASA-TLX.

Analysis Plan

To answer the first hypothesis as well as the second hypothesis the data set of the Qualtrics questionnaires needed to be used. Initially there were two data sets, one for the training group and one for the no-training group. In SPSS these two data sets got combined and brought in line. After that a variable for the group is added to differentiate the participants of the training and no-training group. The last part of ordering and cleaning the data set was to delete the participants that did not fulfil the before mention inclusion criteria. As the bias questionnaire (see Appendix B) only had two options to choose from for each component the frequencies were calculated, as means would not say much about the data. The frequencies were contrasted in bar graphs to get an overview of the data. To see if the training about bias

made a difference in regards to conscious understanding and level of knowledge about biases an independent sample t-test is conducted to assess the p-values of the questions. If there is a p-values below 0.05 there would be a significant difference between the groups, if not then the groups scored very similar on the component. The same analyses were run for the truth-or-lie-questionnaire which was based on the interview question guide (see Appendix E)

For the third hypothesis another data set needed to be used. First the raw video data of the eye-tracker was uploaded in the Tobii Pro Lab app and followed by the two snapshots of the researcher in the position they were interviewed in. Then each video got a label, assigning it to the group the participant was part of, namely training or no-training. Lastly the participant that did not fulfil the inclusion criteria were deleted. After the preparation the videos got coded by hand, as the automatic coding assistant did not reliably work due to the participant not looking at an unmoving picture but a moving and changing suspect. After the coding procedure the areas of interest on each snapshot needed to be added. These areas of interest included upper face, lower face, right arm, left arm, legs, brows, eyes, lips, hair, environment, torso and nose. After marking the areas of interest the data was analysed by the program and different results were exported in a excel file. For this study the data of fixation counts on areas of interest was used. The data set was cleared of unnecessary and distracting information and the two tables for each snapshot combined into one. After sorting the data set, it was imported into SPSS for analysis. Before the analysis the variable for the group of the participant was added. Also the variables right arm and left arm were combined to the variable arms. To get an overview over the data the means and standard deviation of each area of interest got calculated and contrasted in a cross table. After that an independent sample t-test was also conducted to see if there is a significant difference between the training and no-training group. If there is a p-values below 0.05 there would be a significant difference between the groups, if not then the groups scored very similar on the component.

Results

Hypothesis 1

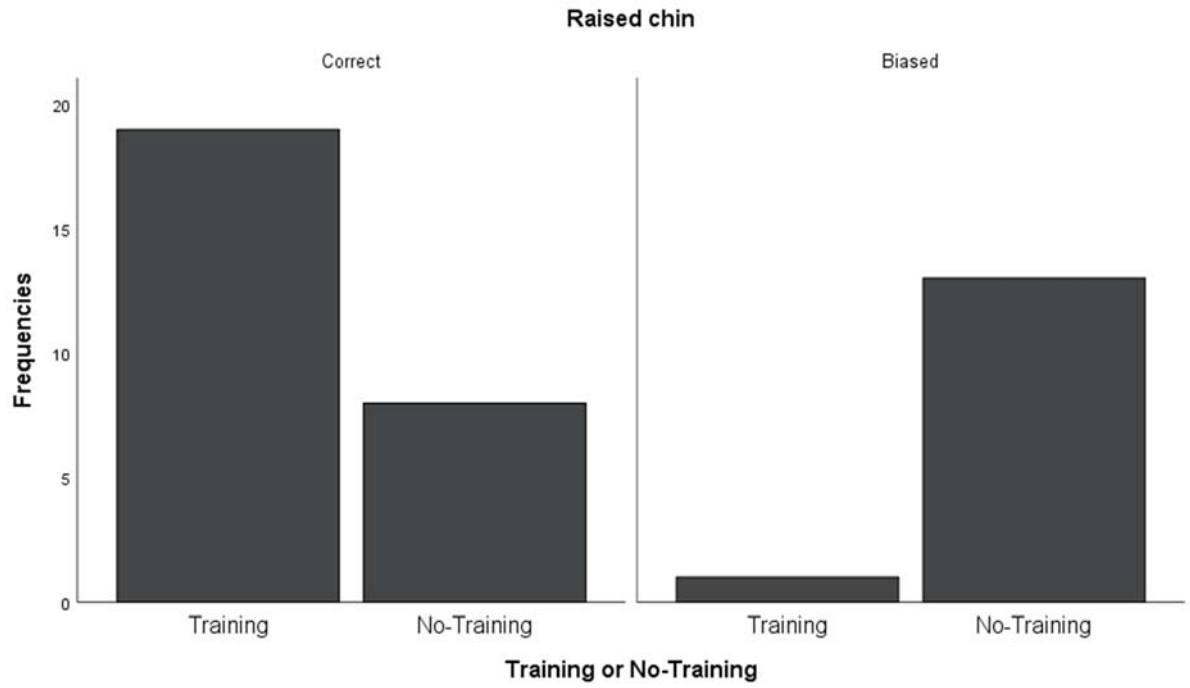
To assess whether there is a difference between the training and no-training group regarding their conscious understanding of existing biases, the amount of correct and biased answers for each component of the questionnaire are compared. The following components are all consistent with the hypothesis, as there is a significant difference between the training and no-training group. The first component of the bias questionnaire is about detecting non-verbal cues as a deception cue, $t(39) = -3.08$, $p < .05$ (see Appendix I). The third component of the bias questionnaire is about detecting raised chin as a deception cue, $t(39) = -4.68$, $p < .05$ (see Graph 1). The fourth component of the bias questionnaire is about detecting pupil dilation as a deception cue, $t(39) = -3.78$, $p < .05$ (see Appendix I). The fifth component of the bias questionnaire is about detecting upper face area as a deception cue, $t(39) = -3.08$, $p < .05$ (see Appendix I). The sixth component of the bias questionnaire is about detecting brows and cheeks as deception cues, $t(39) = -4.68$, $p < .05$ (see Appendix I). Last the ninth component of the bias questionnaire is about detecting increased arm movement as a deception cue, $t(39) = -2.12$, $p < .05$.

The following components, however, contradict with the hypothesis, as there is not a significant difference between the training and no-training group. The second component of the bias questionnaire is about detecting pressed lips as a deception cue, $t(39) = -1.78$, $p = .083$. The seventh component of the bias questionnaire is about detecting fidgeting as a deception cue, $t(39) = -1.78$, $p = .083$. The eighth component of the bias questionnaire is about detecting increased foot movement as a deception cue, $t(39) = -1.37$, $p = .178$.

Based on the results the hypothesis that “the trainings group has a more conscious understanding of existing biases than the control group” can be accepted, as most of the results show a significant difference regarding biases between the two groups.

Graph 1

Frequencies of correct and biased conception on the deception cue of raised chin



Hypothesis 2

To assess whether there is a difference between the training and no-training group regarding their deception detection accuracy, the amount of believed truths and lies for each question of the interview are compared. The following components all contradict with the hypothesis, as there is not a significant difference between the training and no-training group. The first question of the interview is a truth question regarding the name of the suspect, t (39) = -.96, p = .335. The second question of the interview is a truth question regarding the age of the suspect, t (39) = -1.37, p = .178. The third question of the interview is a truth question regarding the origin of the suspect, t (39) = .04, p = .973. The fourth question of the interview is a lie question regarding what the suspect witnessed, t (39) = .12, p = .904. The fifth question of the interview is a lie question regarding the date, time and duration that was witnessed, t (39) = -.29, p = .775. The sixth question is a truth question regarding the place of the incident, t (39) = -1.78, p = .083. The seventh question is a lie question regarding signs and description of suspects, t (39) = -1.11, p = .274. The eight question is a lie question regarding the reason the suspect was near the incident, t (39) = -.78, p = .439. The ninth question is a lie question regarding the reason the incident occurred, t (39) = .10, p = .992. The tenth question is a lie question regarding if the suspect was alone or not, t (39) = -.57, p = 0.569. The eleventh question is a lie question regarding information about other witnesses, t

(39) = .75, p = .457. The last question was about whether the participant considered the suspect guilty or not, t (39) = -.93, p = .357 (see table 1).

The twelfth question is a lie question regarding additional information the suspect might have. Even though, there is a significant difference between the training and non-training group, t (39) = 2.16, p < .05, contradicting the hypothesis the accuracy of the training group only lies at chance level (see Appendix J).

Based on the results the hypothesis that "the trainings group has a higher deception detection accuracy than the control group" must be rejected, as there are no significant differences regarding the amount of detected lies between the two groups.

Table 1

Frequencies on the question whether the suspect is considered guilty or not

	definitely yes	Probably yes	Might or might not	Probably not	Definitely not	Total
Training	0	6	6	7	1	20
No-Training	1	3	4	12	1	21
Total	1	9	10	19	2	41

Hypothesis 3

To assess whether there is a difference between the training and no-training group regarding fixations on areas of interest for deception detection, the average fixation count per area of interest is compared (see table 2). First, there is no significant difference between the scores for the training (M = 1075.15, SD = 400.20) and no-training (M = 892.74, SD = 355.63) group regarding the overall fixations; t (37) = 1.50, p = 0.142.

During the training it was explained that the upper face is a more consistent leakage channel then the lower face, as it is less under conscious control. Still, there is no significant difference between the scores for the training (M = 433.90, SD = 224.63) and no-training (M = 351.89, SD: 187.83) group regarding fixation on the upper face; t (37) = 1.23, p = 0.225. Part of the upper face are the brows, which are an indicator for a truthful smile. There is no significant difference between the scores for the training (M = 57.20, SD = 48.11) and no-training (M = 48.35, SD = 39.13) group regarding fixation on the brows; t (37) = 0.62, p = 0.542. Part of the upper face are the eyes, where especially the pupils can leak if someone is lying. There is no significant difference between the scores for the training (M = 114.40, SD =

73.58) and no-training (M = 87.68, SD = 55.34) group regarding fixations on the eyes; t (37) = 1.28, p = 0.210. There is also no significant difference between the scores for the training (M = 122.00, SD = 123.89) and no-training (M = 121.79, SD = 110.14) group regarding fixations on the lower face; t (37) = 0.01, p = 0.996. Increased pressing of the lips is one of the few cues of the lower face areas. There is no significant difference between the scores for the training (M = 50.20, SD = 57.40) and no-training (M = 42.58, SD = 40.65) group regarding fixations on the lips; t (37) = 0.48, p = 0.637.

The arms and legs are one of the most consistent leakage channels of deception, as the body is less under conscious control then the face. There is no significant difference between the scores for the training (M = 14.60, SD = 17.54) and no-training (M = 7.47, SD = 8.36) group regarding fixations on the arms; t (37) = 1.61, p = 0.117. There is also no significant difference between the scores for the training (M = 1.75, SD = 3.43) and no-training (M = 1.63, SD = 2.06) group regarding fixations on the legs; t (37) = 0.13, p = 0.897.

Based on the results the hypothesis that "the trainings group has a higher number of fixations on arms and legs, as well as the upper face than the control group" has to be rejected, as there are no significant differences regarding the fixations count on areas of interest between the two groups.

Table 2

Average fixation count on areas of interest

	Upper face	Lower Face	Arms	Legs	Brows	Eyes	Lips	Distractions	Total
Training	433.90	122.00	14.60	1.75	57.20	114.40	50.20	281.10	1075.15
No-Training	351.89	121.79	7.47	1.63	48.35	87.68	42.58	231.16	892.74
Total	785,79	243.79	22.07	3.38	115.55	202.08	92.78	512.26	1967,89

Discussion

The purpose of this study was to examine whether there was an effect of a short training about deception detection and one's own biases on veracity judgement ability by comparing the results of a training and no-training group. There were two key findings of the present research. First, a short five-minute training was already enough to dismantle peoples biases and thereby heighten their conscious understanding of existing biases and correct deception cues. However, it was not enough to improve the practical ability to detect lies, also called veracity judgement ability.

For the first hypothesis it was expected that people with training had a more conscious understanding of existing biases than people that did not get trained. Consistent with this hypothesis our results showed that a short training seemed to be already enough to dismantle people's biases and heighten their conscious understanding of existing biases and correct deception cues. Before the training the participants had biases about deception cues such as non-verbal cues, raised chin, pupil dilatation, the upper face, brows, cheeks and arm movements. One explanation for this might be, that these are very specific and unknown deception cues. These results were also in line with the study of Vrij (2000), who found that lay-people seem to focus their attention more on verbal cues rather than non-verbal cues. Even though participants of the no-training group had difficulties identifying these as correct cues, the training group was able to identify them after a short training. The participants did not think about these cues themselves, but after explaining the cues and giving examples they stated that these cues seem obvious and evident. The cues pressed lips, fidgeting and increased foot, on the other hand, did not seem to be cues that the participants had biases about. Both the training and no-training group equally detected them as significant cues. Based on the comments of the participants themselves, a reason for this might have been that these cues were very evidently cues associate with deception and nervousness.

For the second hypothesis it was expected that people with training had a higher deception detection accuracy than people that did not get trained. Contradicting with this hypothesis our results showed that a short training is not enough to improve deception detection accuracy. Regarding every question there seemed to be no significant difference between the training and no-training group. Furthermore, there was no evident overall pattern, leading to the conclusion that even after this training, the participant's decision was based on luck or guessing (Bond, 2008; Hartwig, et al., 2004). A possible explanation for these results could be that even though the participants were trained, they still lacked the substantial experience expert have in detecting lies (Bond, 2008). Based on Bonds (2008) study using

and applying experience was the most successful strategy experts used. Another explanation could be that the participants were subjected to the honesty effect, the tendency of wanting to believe that someone is truthful (Vrij, 1993). This should not have been the case, as the participant got primed to be suspicious, by letting them know that the witness was also a suspect. To conclude, even though the conscious understanding about deception detection was successfully heightened by the small training, the deception detection accuracy was not improved. This leads to the conclusion that veracity judgement ability is not dependent on theoretical knowledge acquired through a small training, but on the participant's experience.

For the third hypothesis it was expected that people with training had a higher fixation count on the areas of interest upper face area and extremities than people that did not get trained. Contradicting with this hypothesis our results showed that there was no significant difference between the training and no-training group regarding fixations on arms and legs, as well as the upper face. Even though there were no differences between the groups, the significant deception cue of the upper face, including eyes, brows and hair, was the most fixated area. As it is common to hold eye contact during a conversation, this high fixation count might not be an indication for an active processing of deception cues (Argyle & Dean, 1965). This is further supported by the high fixation count on the hair area, which indicates a general distraction in this area. Overall there was a high fixation count on distractions supporting the indication that there was lack of active processing of significant deception cues. Furthermore, even though it was explained to the training group that legs and arms are a significant deception cue, both groups focused least on the extremities. Contrary both groups seemed to be very distracted with the body, environment and as already mentioned hair, rather than focusing on significant cues. All this leads to the conclusion that instead of actively processing significant cues, both groups seem to look more randomly around or guess, contributing to the low overall veracity judgment ability (Bond 2008; DePaulo, et al., 2003; Vrij, 2000). These findings are also an extension on the second hypothesis, further strengthening the conclusion that experience rather than theoretical understanding of biases is required for an improved veracity judgement ability (Bond, 2008).

Impact

The results of this study mainly build on the existing evidence of Vrij (2000), who found that a main bias of lay people was a focus on verbal cues rather than non-verbal cues. The results of this study support this claim, but also gained additional information about biases lay-people had about non-verbal cues and deception cues they did not have biases

about. On the one hand, lay-people in this study tended to think a lowered chin, smaller pupils, the lower face area and a decrease in arm movements are deception cues. On the other hand, a majority correctly assumed that pressed lips, fidgeting and increased foot movements are deception cues. The results of this study could in practice be used as an extension to the bias list of Vrij (2000).

However, the results of this study contradict with the findings of Hartwig et al. (2004). They concluded that the biases people have about deception detection are a major factor influencing their veracity judgment ability. The current study, however, found that there was no difference in veracity judgment ability, as neither the accuracy in detecting lies nor the active fixations on significant cues improved, when people got informed about their biases. The results of the current study would support Bond´s (2008) findings, that deception detection is a skill relying on long years of experience. This means that a short training or the mere knowledge about biases is not enough to develop this skill. This has two major implications for practice. First, police academies or other institutions that train detectives or investigators should focus their schooling about veracity judgement more on practical exercise and experience rather than conducting theoretical testing. Second, based on these results the focus in research should be shifted away from the impact of theoretical understanding to the impact of experience and practical training on veracity judgement ability.

Limitations

There were at least three major potential limitations we found concerning the results of this study. The first limitation concerns the trainings duration of the study, as the training about the biases of the participants was on average only five minutes long. It could be concluded that this short training was just not enough so achieve a significant difference between the two groups. Even though this can be seen as a limitation, it also can be seen as the starting point of a gradual process to discover the amount of training needed to achieve improvement.

The second limitation concerns the overall timeframe of the study. The study only took around 30 minutes, meaning that between the two times the biases questionnaire was filled out only 15 minutes on average pass. It could be argued that the outcomes of the second bias questionnaire were not a result of long-term memory understanding, but mere a short-term remembrance of the correct answers. However, before the second bias questionnaire was filled out the participants had to do two demanding tasks, namely an interview and another

questionnaire. This would have made it more difficult to just remember the answers. But it cannot be ignored that there was the potential for this limitation as there was not follow-up questionnaire a few days later testing if the biases were truly eliminated.

Third it cannot be ignored that this study required direct contact with participants during the Corona pandemic. Even though enough participants were reached, around 20 more participants might have led to different results. The problem was that due to Corona only few participants signed up and at the end it became less, leading to us reaching only the sufficient level of participants.

Much work remains to be done before a full understanding of the effects of training about deception bias on veracity judgment ability is established. For one a study could examine if the short training has long-term effect on the participants biases. This could be achieved by conducting this study but adding a third bias questionnaire a few days or even weeks later. Through this the limitation of short term memory could be solved. Other studies could examine different lengths of training, to assess the amount training needed for increased performance. Next to different lengths of training, different focuses of training could be assessed, too. For example pure theoretical training, pure practical training or a combination of both could be applied.

Conclusion

Although the broader scope of the current results must be established by the future research, the present study has shown a tendency for the possibility that training could improve veracity judgement ability of people. While the veracity judgement ability in itself did not improve, the conscious understanding about biases got heightened through the training. If further investigated, this might be the start to reduce people's susceptibility to fraud and deception.

References

Argyle, M., & Dean, J. (1965). Eye-contact, distance and affiliation. *Sociometry*, 289-304.

Bond, G. D. (2008). Deception detection expertise. *Law and Human Behavior*, *32*(4), 339-351.

Borys, M. & Plechawska-Wójcik, M. (2017). Eye-tracking Metrics in perception and visual attention research. *European Journal of Medical Technologies* 3(16), 11-23.

Buller, D. B., & Burgoon, J. K. (1996). Interpersonal deception theory. Communication Theory, 3, 203–242.

Carter, B. T., & Luke, S. G. (2020). Best practices in eye tracking research. *International Journal of Psychophysiology*, *155*, 49-62.

Chun, M. M., & Wolfe, J. M. (2001). Visual attention. *Blackwell handbook of perception*, *272310*.

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

Ekman, P., Davidson, R., & Friesen, W. (1990). The Duchenne smile: Emotional expression and brain physiology: II. *Journal of Personality and Social Psychology*, 58(2), 342–353. doi:10.1037/ 0022-3514.58.2.342.

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

Ekman, P., & O'Sullivan, M. (1991). Who can catch a liar? *American Psychologist*, 46, 913-920

Farnsworth, B. (2020). 10 Most Used Eye Tracking Metrics and Terms [Blog post]. Retrieved from https://imotions.com/blog/10-terms-metrics-eye-tracking/

Frischen, A., Bayliss, A. P., & Tipper, S. P. (2007). Gaze cueing of attention: visual attention, social cognition, and individual differences. *Psychological bulletin*, 133(4), 694.

Hartwig, M., Granhag, P. A., Strömwall, L. A., & Vrij, A. (2004). Police officers' lie detection accuracy: Interrogating freely versus observing video. *Police Quarterly*, *7*(4), 429-456.

Ioannou, M., & Hammond, L. (2015). The detection of deception within investigative contexts: Key challenges and core issues.

Lomer, D. (2017). 44 Investigation Interview Questions for the Complainant, Subject and Witnesses - Take your investigation interviews beyond the who, what, where, when, why and how of what happened. *I-Sight Software*. Retrieved from https://i-

sight.com/resources/44-investigation-interview-questions-for-complainant-subject-and-witnesses/

Polizei-Beratung (2021). Betrüger an Ihrer Haustür. Retrieved from https://www.polizei-beratung.de/themen-und-tipps/betrug/haustuerbetrug/

Porter, S., & ten Brinke, L. (2010). The truth about lies: What works in detecting high-stakes deception?. *Legal and criminological Psychology*, *15*(1), 57-75.

Porter, S., Ten Brinke, L., & Wallace, B. (2012). Secrets and lies: Involuntary leakage in deceptive facial expressions as a function of emotional intensity. *Journal of Nonverbal Behavior*, *36*(1), 23-37.

Runeson, S., & Frykholm, G. (1983). Kinematic specification of dynamics as an informational basis for person-and-action perception: Expectation, gender recognition, and deceptive intention. *Journal of Experimental Psychology: General*, 112, 585-615.

Sebanz, N., & Shiffrar, M. (2009). Detecting deception in a bluffing body: The role of expertise. *Psychonomic bulletin & review*, 16(1), 170-175.

Vrij, A. (1993). Credibility judgments of detectives: The impact of nonverbal behavior, social skills and physical characteristics on impression formation. *Journal of Social Psychology*, 133, 601-611

Vrij, A. (2000). Detecting lies and deceit: The psychology of lying and its implications for professional practice. Chichester, UK: Wiley.

Vrij, A. (2004). Why professionals fail to catch liars and how they can improve. Legal & Criminological Psychology, 9, 159-181.

Wu, Z., Singh, B., Davis, L., & Subrahmanian, V. (2018, April). Deception detection in videos. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 32, No. 1).

Zuckerman, M., DePaulo, B. M., & Rosenthal, R. (1981). Verbal and nonverbal communication of deception. In L. Berkowitz (Ed.), Advances in experimental social psychology (Vol. 14, pp. 1–59). New York: Academic Press.

Appendix A

You are invited to participate in a research study as part of a bachelor's thesis on Eye tracking and Deception Detection at the University of Twente.

In this study, you will be asked to play the role of an investigator in a crime setting. For this, you will wear an eye tracker which will record your gaze behaviour. Your task will be to identify if your interviewee is guilty in the provided crime-scenario or not. The purpose of this research is to find a correlation between eye-gaze behaviour and lie detection. After the interview, the participant will be debriefed about the study.

Namely, you should be aware, that this study will be about a crime scenario, which might be a sensitive topic to you.

Else, there is no known risk in participating in this study. When agreeing to participate, you agree to the interview, to being video recorded, using the eye tracker and to complete the surveys. Additionally, you agree for the researcher to keep your contact information and might be contacted for follow-up or future research.

The participant can withdraw their consent at any time, but should contact one of the two researcher to do so. Names, dates, locations, and other confidential data will be anonymised by the researcher. Nonetheless, the anonymised data will be shared within the research team and their supervisor. The data will be stored anonymously and used by the research team. It will be published in the respective bachelor thesis of the students, but it will not be used for any commercial purposes. The data might be reused in an academic context. There are no third parties involved.

Researchers:

Sena Bodur, s.y.bodur@student.utwente.nl

Tony Jungfer, t.n.jungfer@student.utwente.nl

Supervisor:

Peter Slijkhuis, p.j.h.slijkhuis@utwente.nl

If you have questions about your rights as a research participant, or wish to obtain information, ask questions, or discuss any concerns about this study with someone other than the researcher(s), please contact the Secretary of the Ethics Committee of the Faculty of Behavioural, Management and Social Sciences at the University of Twente by ethicscommittee-bms@utwente.nl

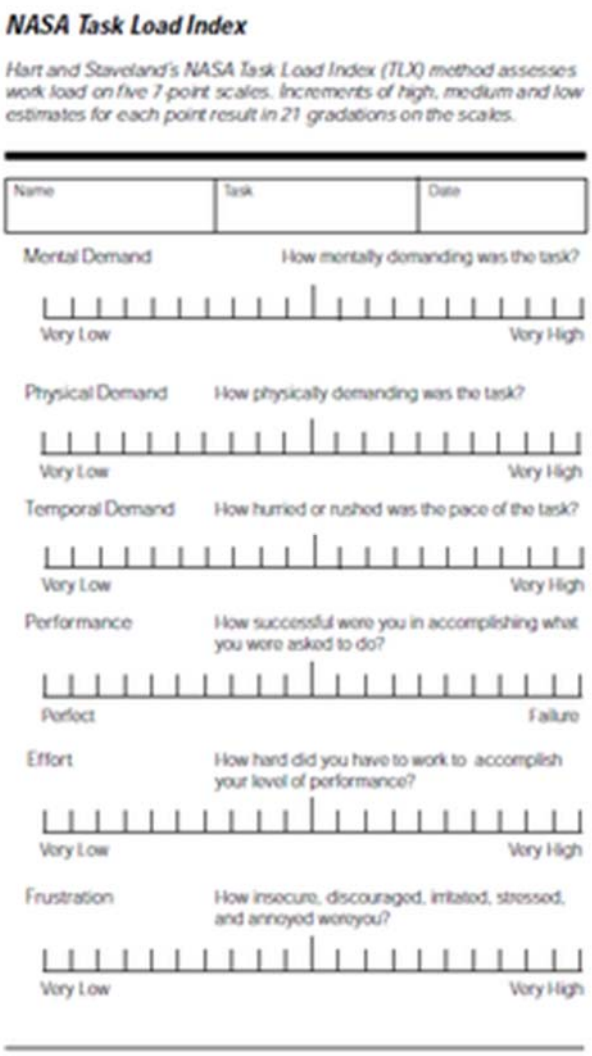
Appendix B

Preconception Questionnaire:

When wanting to detect deceit or a lie on which of the following cues would you focus?

Non-verbal cues	○ ○	Verbal cues
Pressed lips	○ ○	Loose lips
Raised chin	○ ○	Lowered chin
Pupil dilation	○ ○	Pupil shrinking
Upper face areas	○ ○	Lower Face areas
Brows and cheeks	○ ○	Mouth corners
Fidgeting	○ ○	Sitting still
Increase in foot movements	○ ○	Decrease in foot movements
Increase in arm movements	○ ○	Decrease in arm movements

Appendix C



NASA Task Load Index

Hart and Staveland's NASA Task Load Index (TLX) method assesses work load on five 7-point scales. Increments of high, medium and low estimates for each point result in 21 gradations on the scales.

Name	Task	Date

Mental Demand — How mentally demanding was the task?

Very Low — Very High

Physical Demand — How physically demanding was the task?

Very Low — Very High

Temporal Demand — How hurried or rushed was the pace of the task?

Very Low — Very High

Performance — How successful were you in accomplishing what you were asked to do?

Perfect — Failure

Effort — How hard did you have to work to accomplish your level of performance?

Very Low — Very High

Frustration — How insecure, discouraged, irritated, stressed, and annoyed wereyou?

Very Low — Very High

Appendix D

Scenario for the interviewer

On Friday the 26 of March 2021 at around 23.00 the clothes shop "America Today" in Enschede centre was vandalized. Unknown suspects broke into the shop by demolishing the windows and destroyed the interior and set it on fire. When the police arrived the perpetrator/s was/were gone, but a few witnesses were already at the crime scene, waiting for the police. Based on previous witness interviews there should have been between 2-5 perpetrators, most gave this rough indication. Furthermore, some witnesses think they were probably adolescents, judging by the sound of their voices and laughter. Also, they were probably armed with bats or hammers. One witness saw someone run away with a dark green hoodie, jeans, and white Nikes.

You are a police interviewer and your job is to interview another witness that was at the crime scene and determine the value of their information. Based on previous interviewees this person could also potentially be a suspect, as some said that this young person joined the crowd of witnesses later when the police were already there. Furthermore, another witness reported that this individual seemed very nervous. So your job is also to determine if the suspect/witness is innocent or not.

Here is a list of all the information already gathered:

- 2-5 perpetrators
- Adolescent (16-20 years old)
- Armed with hammers and bats
- One wore a dark green hoodie, jeans, and white Nikes.
- They demolished the store and set it on fire
- The police was first called around 23:00
- One witness might be a suspect, as they joined the crowd later, seemed very nervous

Appendix E

Question Guide

Questions to Ask the Witnesses (Lomer, 2017):

1. Welcome, the witness/suspect. Introduce yourself and what is going to happen.
2. What is your name?
3. What is your age?
4. Where are you from?
5. What did you witness?
6. What was the date, time, and duration of the incident you witnessed?
7. Where did it happen?
8. Did you see any signs of suspects? Give a description of potential suspects.
9. Why were you near the incident?
10. Do you know why the incident occurred?
11. Were you alone? (If no: Can someone confirm that they were with you?)
12. Do you know anyone else who saw the incident?
13. Is there anything else you want to tell me that I haven't asked you?

Appendix F

Suspect scenario

You are a 19-year-old adolescent, you like parties, alcohol and do not care what anyone says. Furthermore, you really like the thrill of doing illegal and dangerous activities, anything else bores you. On Friday the 26 of March 2021 sometime late in the night you (you think it was between 10 pm and 12 pm, as you already drunk a little bit) and 2 of your friends decided to have some fun in the city centre of Enschede. You grabbed some bats and golf clubs from your home and decided to destroy the clothing store "America Today" that kicked you out once. You demolished most of the shop and set fire to it before you heard the police coming and a group of people outside of the shop. You decided to leave through a back window, but your friends were faster than you. You decided to secretly join the crowd of witnesses instead to cover up your involvement.

Your friends are the same age as you. One has worn a dark green hoodie, jeans, and white Nikes. The other has worn a black jacket, sweatpants, and black shoes.

Unfortunately, you have to participate in an interview as a witness in the police station, your job is to hide and discredit your and your friend's involvement in this to save yourself from punishment.

Appendix G

Protocol

Step 1: The Preparation (max. 20 minutes)

Researchers set up the eye tracking glasses and the program

Researchers prepare the scene/ lab

Participant walks in

Participant reads, agrees, and signs consent form on laptop

Researcher hands participant the eye tracker and they put it on

Calibration of the eyetracker

Explanation of the procedure to the participant/ handing over of the documents

Documents:

- Paper which explains the role of the Interrogator and the scenario
- a Question guide (semi-structured interview)

Participants have 10 minutes to prepare themselves for this role.

In the experimental group:

They fill out the bias questionnaire on a laptop prior to the interrogation

participant gets a short training, explaining what they should pay special attention to (5 minutes)

Training:

1. Fill out bias questionnaire
2. discussion of the questionnaire, informing about right beliefs and correcting wrong assumptions
3. explaining what Experts do differently
   → experts pay more attention to non-verbal cues
   → they fixate face areas, like lips, eyes, nose and cheeks
   → non specifically pupil dilation, changes, chin raise and pressing lips
   → less genuine smiles (a genuine smile can be recognized by raised lips **and** the "orbicularis oculi surrounding the eyes, which pull the cheek up while slightly lowering the brow" (Porter, et al., 2012).)
   → also explain that they focus on movements in increased movement or fidgeting in arm and leg areas.
4. Ask them to repeat the information/ ask three question in the questionnaire

In the control group:

no training and no questionnaire

Step 2: The Experiment (15 minutes)

The participant takes the role of the Interrogator in this setting

The other researcher takes the role of the suspect and has a specific role to play (this role is planned and thoroughly studied beforehand, Appendix E)

They act out an interrogation which is timely limited to 15 minutes

Step 3: The Debriefing (max 10 minutes)

After the Experiment, the participant (in control group and experimental group) will be asked to fill out a questionnaire (about biases) and fill out the NASA-TLX

The researcher can now debrief the participant and tell them about further (prior withhold) details

Step 4: After the experiment

Cleaning of materials (eye trackers, laminated papers, laptop)

Appendix H

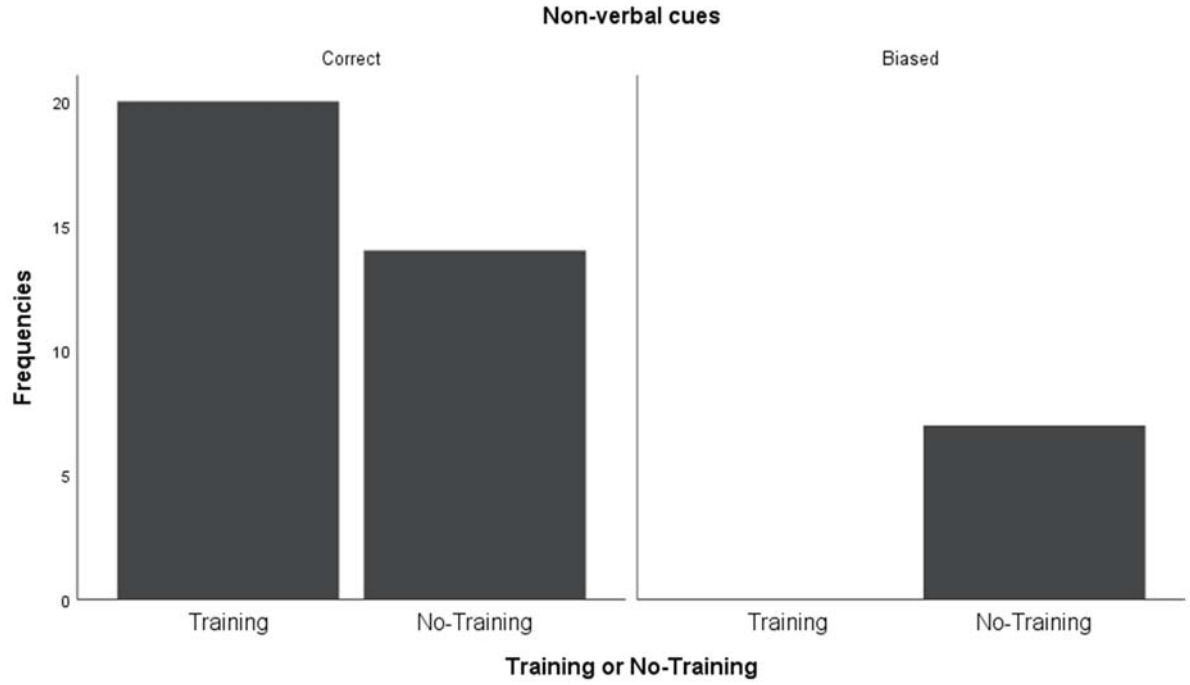
Prior to the training the participant is asked to fill out the bias questionnaire. After they are finished, the researcher takes a look at their results. Based on that, they explain to the participant which beliefs were right, and which were misconceptions. As follows, the researcher explains how detection experts identify the lies of their suspects. The information is based on previous research, specifically based on the work of DePaulo et al (2003), Porter et al (2012) and Ekman and Friesen (1969). After the training was finished the participants were asked to repeat the information they got in their own words to show understanding. This training was conducted with the following checklist, which the researcher used for orientation.

1. Experts pay more attention to non-verbal cues than verbal cues. Explain that stories can be very well constructed with time to prepare, but behaviour or non-verbal cues are less under conscious control. Therefore it is advised to not look for cues in the story but closely look at the body language.
2. One of the areas experts focus on is the face. One cue that might leak deceit is an increase in pressing lips, as biting lips for example is often a sign of nervousness or that someone is hiding something. Additionally an increase in chin raises might leak deception, as people who lie tend to look away at the ceiling rather than on the floor. Another reliable lie detection cue are the pupils because they cannot be consciously controlled. Pupils dilate when someone is telling a lie. Nonetheless, this is difficult to detect. Hence, it is advised to focus more on other cues.
3. When people lie there is an increase in fake smiles. A genuine smile can be recognized by raised lips **and** the "orbicularis oculi surrounding the eyes, which pull the cheek up while slightly lowering the brow" (Porter, et al., 2012).
4. Next, it needs to be explained that the main focus on the face should be on the upper face area. It is explained that this area is less under conscious control then the lower face area and is therefore more reliable. Furthermore, it includes more significant cues than the lower face area, namely the eyes, brows and cheek.
5. Lastly, it is also important to focus on the arm and leg areas. Increased movement or fidgeting in these areas are often a reliable sign for deceit, as these areas are under least conscious control.

Appendix I

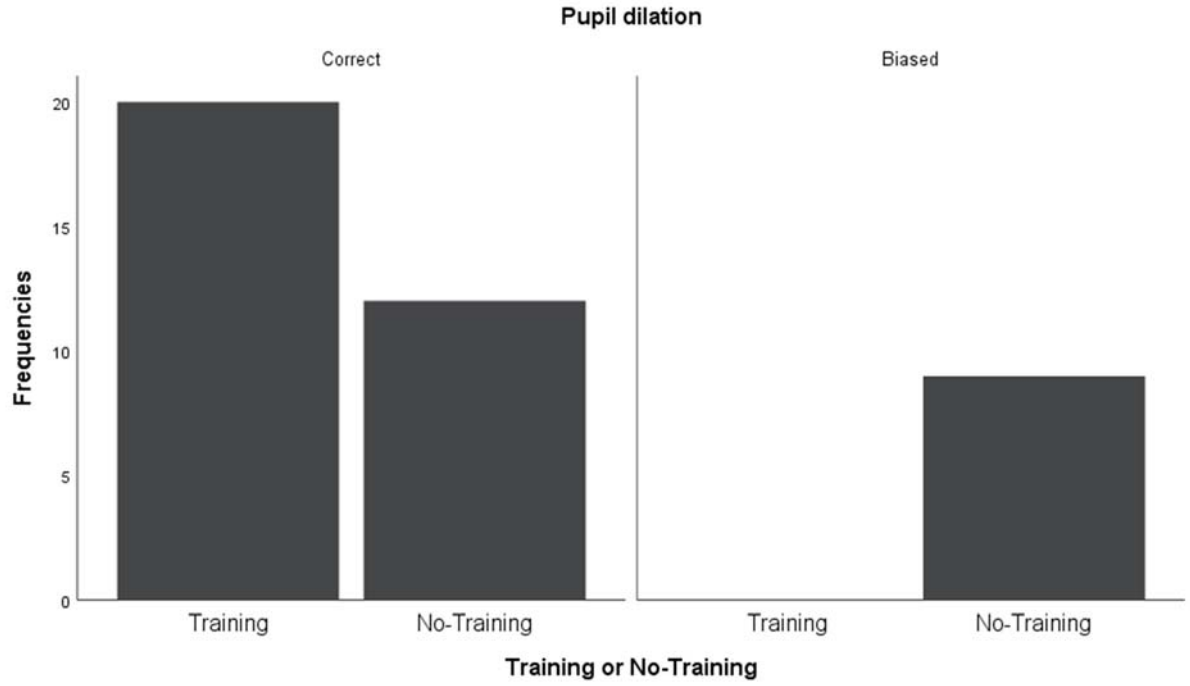
Graph 2

Frequencies of correct and biased conception on the deception cue of non-verbal behaviour



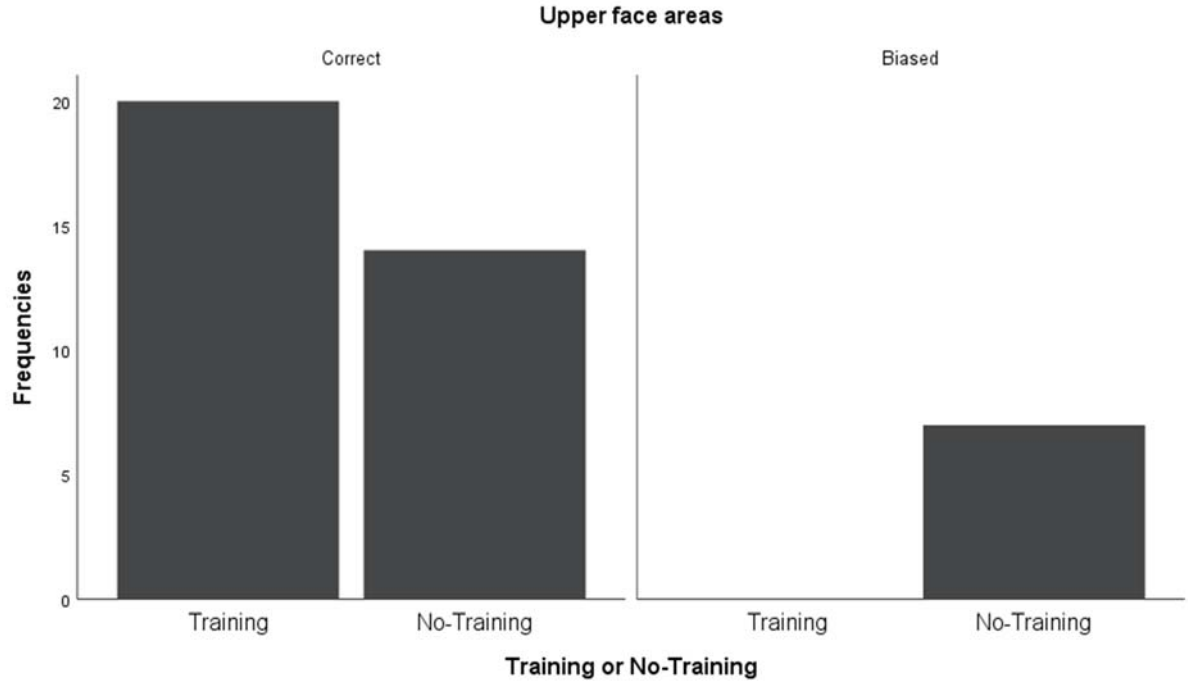
Graph 3

Frequencies of correct and biased conception on the deception cue of pupil dilation



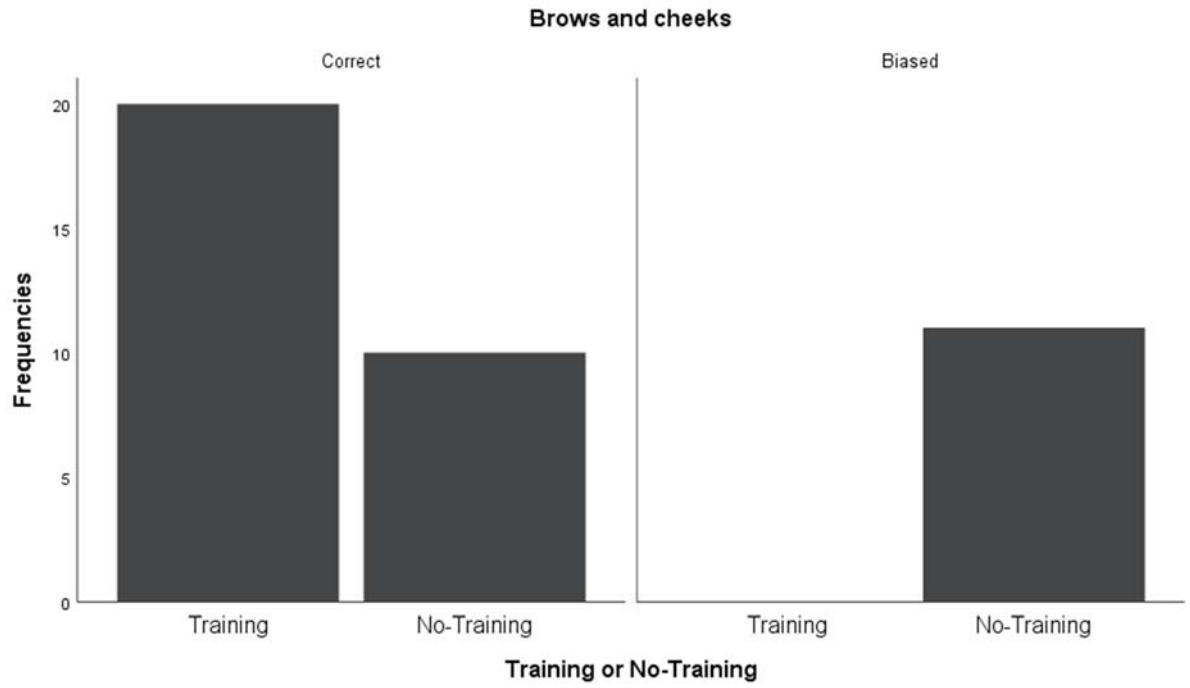
Graph 4

Frequencies of correct and biased conception on the deception cue of upper face areas



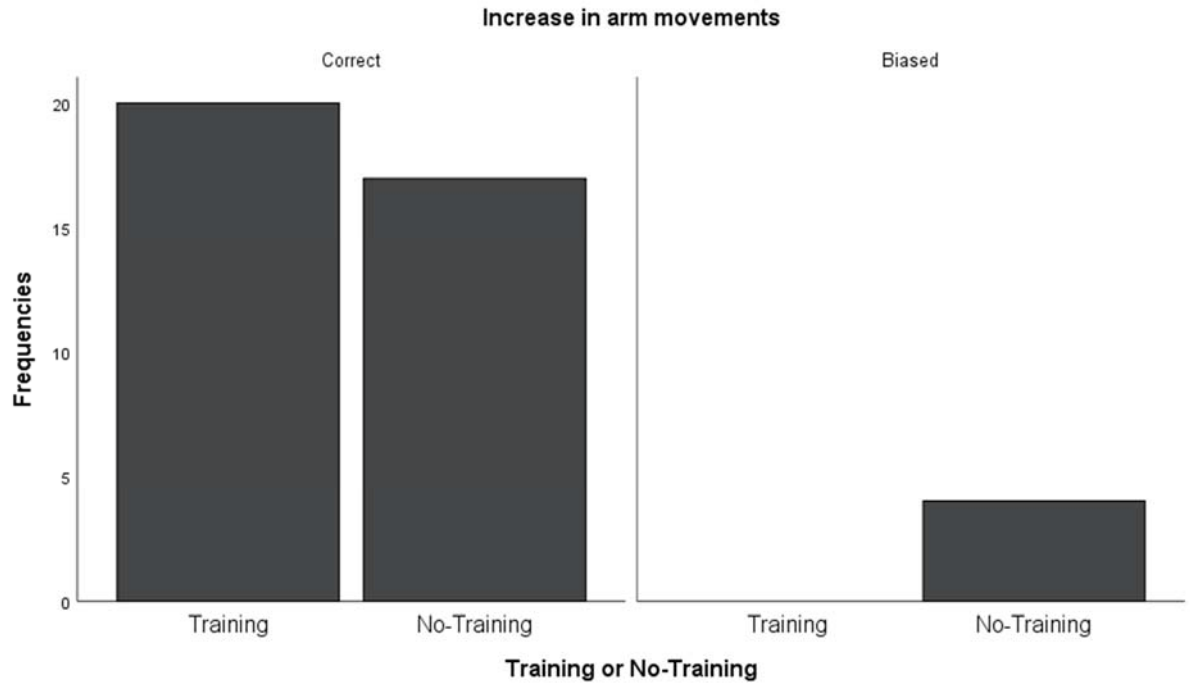
Graph 5

Frequencies of correct and biased conception on the deception cue of brows and cheeks



Graph 6

Frequencies of correct and biased conception on the deception cue of increased arm movements



Appendix J

Graph 7

Frequencies on the question whether the suspect has anything else to add