[DECEPTION DETECTION AND EYE-TRACKING]

[How are they affected by cognitive load and training?]

[6.07.21]
[SENA BODUR]
[206816]
[Tutor: Peter Slijkhuis]

Abstract
In general, people are bad at detecting lies and do not know indicative non-verbal and verbal deception cues. Laypeople mostly guess and have an accuracy rate around chance level.

By incorporating cognitive load theory and the eye mind link, this research wants to explore how training in non-verbal deception cues can influence cognitive load and veracity judgment accuracy with the aid of eye-tracking. This is done in a comparative between-subject experimental research. There is a sample of 41 participants separated into training and no-training groups. Both groups participate in an interview, where an actor plays the role of a suspect in a mock-crime scenario. During this interview, the gaze behaviour of the participant gets recorded to estimate cognitive load and which cues they attend. Afterwards, they fill out three questionnaires about their biases, their cognitive load, and their judgement. It was investigated how they perform regarding veracity judgement accuracy and how cognitively demanding they perceived the interview. It turned out that the training did not influence the self-perceived cognitive load, fixation durations nor the veracity judgement accuracy. All hypotheses were rejected. This is probably because non-verbal cues are not reliable cues for deception. Further it might that the training was too short of giving significant results.

Keywords: *cognitive load, eye-tracking, veracity judgement ability, interview, non-verbal deception cues, training*

Table of contents

"No mortal can keep a secret, if his lips are silent, he chatters with his fingertips; betrayal oozes out of him at every pore"
(Freud, 1953)

Introduction

Deception is integrated into everyday life. Whether it is an applicant lying on their resume, teenagers are lying to their parents, or criminals lying in court. The topic has received a great deal of attention because of its predominance and fascination (DePaulo, Kashy, Kirkendol, Wyer, & Epstein, 1996). According to Mahon (2008), deception is the act of telling an untruthful statement with the intent of making other people believe the statement is truthful. In that definition, deception has three essential conditions. These would be the statement's veracity, the awareness of the sender that the statement is false, and the intent of covering this false statement as true (Fallis, 2010). In connection to this, he also states that the intention to deceive plays an important role. For example, some lies are used without an intention to deceive. These are called bald-faced lies that are usually not perceived by others as a lie. At the same time, lies with the intention of the sender to be a lie are considered deception. These are the types of lies that are supposed to be discovered by the interviewer (Fallis, 2010). Nonetheless, detecting deception is a complex task. Some cues, strategies, and tools will be discussed.

Deception detection strategies and Training

In general, people are mediocre at detecting lies. According to Hartwig, Granhag, Strömwall and Vrij (2004) neither laypeople nor experts are much better than chance (50%) at detecting deception. They claim this is caused by a mismatch of expectations and stereotypes on how liars behave. However, Porter, Juodis, ten Brinke, Klein, and Wilson (2010) showed that training can improve the judgement of the deception detectors. Bond (2008) had a similar finding that experts on deception detection have an accuracy from 80%-90%. Their accuracy score is higher because experts are experienced and use schemas which helped them in the past to distinct lies from the truth. This means that when people know about deceptive cues in more detail, they are more able to recognize deceit. Therefore, training might improve the veracity judgement accuracy of laypeople.

The training in the experiment by Porter et al. (2010) debunked myths around deception cues and explained which cues are indicative for lying. For example, non-verbal

signs for deception can be pressed lips, which can be indicative of nervousness or secrecy. Moreover, deceivers tend to raise their chin to appear more truthful (DePaulo, Lindsay, Malone, Muhlenbruck, Charlton, & Cooper, 2003). According to Porter, Ten Brinke and Wallace (2012), liars appear to smile more often. Nonetheless, their smiles are not genuine, which is recognizable through the orbicularis oculi which outlines the eye area. This muscle draws up the cheek while simultaneously letting down the brow area. Likewise, the upper face area is a more significant cue than the lower face area. This is due to the fact, that the upper face area cannot be consciously controlled that well unlike the lower face area. Not all non-verbal cues can be found in the face. For example, according to DePaulo et al. (2003) liars move their legs and arms more. In a similar fashion, there are some tools and methods which are assumed to detect deception. Nonetheless, most of these are criticized for being inaccurate (Nortje & Tredoux, 2019). In the following paragraph, some non-verbal and verbal methods and tools will be discussed.

There are significant non-verbal methods to detect deceit. The first example would be the Polygraph. It is deemed critical that polygraphs can prove reliably whether someone deceives or not (Nortje & Tredoux, 2019). Moreover, many people get anxious when interrogated and might fear consequences like being convicted wrongfully and thus would show an increased activation. On the one hand, it cannot be ruled out that truthtellers could be accused (false positive). On the other hand, there are several tactics of liars to manipulate the results (false negative). An example would be counting backwards or moving toes so that the difference in arousal between critical and control question is undetectable (Nortje & Tredoux, 2019).

Apart from the polygraph, evoked response potential (ERP) and fMRI can be used to detect deception. In connection to that, the right ventrolateral prefrontal, left ventrolateral prefrontal and medial premotor are brain areas which might show if a person lies (Nortje & Tredoux, 2019). This area is still researched and leaves room for improvement. Nonetheless, the same problem as with the polygraph remains that the focus on physiological readings might be misleading. In addition to that, the ERP and fMRI are highly impractical and time-consuming to be used frequently. Besides that, it is not realistic to convince an interviewee to undergo this procedure. Since physiological measures are flawed, it might be more effective to take a closer look at verbal deception detection measures.

Not only the body can expose a deceiver, but also the way they speak can be indicative of a lie. For example, deceivers talk more slowly and interrupt their speech more often than truthtellers (Hartwig, Granhag, Strömwall, & Vrij, 2004). Based on verbal cues,

several verbal measures have been developed. Verbal measures are mostly interview techniques with different intentions. Some suspect interviewers are focused on obtaining a confession, which influences the interrogation style (Hartwig et al., 2004). For example, the nine-step approach by Inbau, Reid and Buckley (1986) has been criticized for leading to false confessions. The reason for this is because it emphasizes possible advantages of confessing and discourages denial. Next to this would be the information-gathering interrogation, also called, ethical or inquisitorial interrogation (Hartwig et al., 2004). This puts more emphasis on acquiring information in full measure and being open-minded. During the interview the witness report should not be judged but should rather be consumed and evaluated afterwards. According to Hartwig et al. (2004) the downside of this method is that it is usually impractical in an interrogation setting.

In conclusion, these techniques (non-verbal and verbal) used in deception detection research and settings prove to be inaccurate, impractical, and unreliable. Therefore, it can be stated that the area of non-verbal and verbal deception detection methods leaves room for improvement and research.

Eye tracking and eye-mind theory

In recent years eye tracking gave a prospective outlook to be a valid physiological measure of lie detection. In the past eye tracking has already been utilized as a lie detection device. Good examples for this are Eye Detect or the AVATAR, Automated Virtual Agent for Truth Assessments in Real-Time (Bessonova & Oboznov, 2018). The reason for this gain in popularity are the advantages eye-tracking as a physiological lie detector has to offer. According to Bessonova and Oboznov (2018), eye-trackers have high measuring accuracy, good calibration and the results are easy to administer. In addition to this, it offers high-frequency data, even the most subtle changes in gaze behaviour can be recorded and analysed. For deception detection, this is an indispensable feature. Considering, that most non-verbal cues are subtle and otherwise difficult to recognize. As mentioned by Bessanova and Obznov (2018) the reasons for this are that eye movements are involuntary and cannot be consciously controlled. Eyes, although not completely proven, are connected to the mind and give a good reflection of various mental processes. Research by Just and Carpenter (1980) and Rayner (1978) have established the eye-mind theory. To elaborate, human beings can only see focussed within the foveal area, due to eye anatomy. The gaze moves according to the stimulus we are engaged with at the moment. As a result, the gaze fixates on this stimulus, also called a fixation, and moves on to the next, after processing it.

Fixations are steady gaze points, while saccades are the gaze movement between two fixations. Regarding this, longer fixation durations imply concentration and processing time. For example, when focused on reading, the eye has longer fixation durations on the word which is currently being processed. In comparison, when searching for an object, the gaze tends to move around more (Just and Carpenter, 1980; Rayner, 1978). Consequently, a higher cognitive load could be represented by fixation durations (Freire, Eskritt, & Lee, 2004). In addition to that, it has been found by Rayner (1998) that large saccade amplitudes are indicative of high concentration as well. When their participants were asked to perform cognitively demanding tasks, their saccade amplitude lessened. Since this was demanding higher concentration and it would result in more fixations. This could be an indicator of cognitive load as well. In conclusion, it would be of use to track eye movement to take a further look at the connection between deception and gaze behaviour.

Cognitive load and the eye mind theory

As stated earlier, there is a connection between gaze behaviour, people's intentions, and cognitive load. In a similar fashion, the research area of deception detection explored the connections between (the intention of) lying and cognitive load. Burgoon et al. (1989) and Köhnken (1989) state that a lie is more calculated and thus mentally demanding. They claim that a person who deceives tries to observe others and appear convincing to them. Consequently, they judge the reactions carefully and adapt to them (Burgoon et al., 1989; Köhnken, 1989). Different cognitively loaded methods have been introduced to test the credibility of suspects. For example, the suspect tells the events of a story backwards (Vrij et al., 2008). As indicated prior by Ekman and Friesen (1974), because of this increased cognitive load for the deceiver, the body "leaks" micro-expressions as a result. Examples could be inconsistencies in their answers and increased response times.

For the interviewer this is crucial information. In an interview setting, the interviewer should be aware of the impact the situation has on the interviewee to be able to interpret the behaviour rightfully.

Notably, the investigative interview can be cognitively demanding for the interviewers as well. According to Hanway, Akehurst, Vernham and Hope (2021), the Cognitive load theory (CLT) describes three different types of load which are applicable in most settings. These types are intrinsic load, extraneous load, and germane load. First, intrinsic load refers to the basics of information processing and how complex it is. Second, the extraneous load is influenced by external factors, for example being under time pressure. Third and last, the

germane load is describing the load for learning and application of new processes and skills. Particularly, the germane load is fundamental to build and extend cognitive schemas.
To relate this back to the interviewer, there are several intrinsic factors that increase the cognitive load. For example, the interviewer needs to listen in a concentrated way, and remember in detail the shared information. Furthermore, they are also taking notes and verbalize hypotheses (Hanway, Akehurst, Vernham, & Hope, 2021). Especially, given the case that they are newly learning the information, it might be a highly cognitively demanding task. For experts, these tasks are probably more automatic, therefore easier to administer and less cognitively loading. Simultaneously, they need to think about their next question and develop an appropriate strategy. This is a complex, mentally demanding task. And a mentally demanding task might influence the accuracy of veracity judgments negatively (Hanway, Akehurst, Vernham, & Hope, 2021). It would therefore be of interest to find ways to decrease the cognitive load on the inexperienced interviewers to improve their veracity judgements. Bond (2008) found that experts have a high veracity judgement accuracy, up to 90%, through practice. This raises the question if training could improve the cognitive load and simultaneously influence the veracity judgement accuracy too.

Current Research

A well-thought-out training might be of advantage to decrease the cognitive load and improve veracity judgements. This training would incorporate different non-verbal deception cues. As mentioned, these might expose the deceiver. Especially, the upper face area, eyes and brows are of specific interest in this domain because they are under less conscious control than the lower face area. Therefore, this study will investigate if the training group would pay more attention to this than to other non-verbal cues. In addition to this, fixation durations can be indicative of cognitive load. Longer duration times can be a signifier for higher cognitive load. To aid this, the participants will be asked to fill out the NASA-TLX which measures their self-perceived cognitive load. The target group is limited to laypeople because the training incorporates basic knowledge, experts are most likely aware of. In addition to this, we wanted to research how the training could affect laypeople in particular. This study wants to investigate how training affects the cognitive load and the veracity judgement accuracy of laypeople which leads to the following research question and hypotheses:

Research question:
What is the effect of training in significant (non-)verbal deception cues on cognitive load and veracity judgement accuracy of suspect interviewers?

Hypothesis:
H1: Interviewers in the training group will have more accurate veracity judgements than interviewers in the control group.
H2: Interviewers in the training group will have longer fixation durations on the upper face area (e.g. eyes and brows) than interviewers in the control group.
H3: Interviewers in the training group will have a lower self-perceived cognitive load than interviewers in the control group.

Method

Participants

The participant population consisted of 48 participants. All of the participants are laypeople and do not have any professional experience with investigative interviews. 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 person. 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 three 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 variables were the veracity judgement accuracy of the participants, the perceived cognitive load of the participants and the non-verbal deception cues (more specifically, the upper face area, eyes, and brows).

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 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 suitable for everyday use which fits our methods best.

Procedure

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 bias-questionnaire and the NASA-TLX.

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 could ask follow-up questions to the suspect. After the interrogation, the participants were debriefed and asked to fill out the bias-questionnaire and the NASA-TLX.

Data Analysis

For the three hypotheses, there were three different ways to analyse the data.

For the first hypothesis, the veracity judgement accuracy of the trained group and the untrained group was compared to each other. For that, the accuracy scores of the participants were calculated. For each correct answer, the participant got one point. A correct answer was defined as the participant recognizing the lie or the truthful statement. The highest reachable score was 14 and the lowest 0. The participant had a high accuracy score if their score was between 12 and 14. If their score was between 12 and 8, their score was deemed average. If the score was below 8, than it was deemed low accuracy. To determine the differences at each sub question between the trained group and the untrained group, an independent samples t-test was conducted. In addition to this a bar chart was created to visualize the mean differences.

To test for the second hypothesis, the eye tracking videos were mapped manually using the tobii-pro lab software. Automatic mapping was not suitable for this research because the results ended up being highly inaccurate. The fixation durations of both groups on the upper face areas, the eye area and the brow area will be compared using an independent samples t-test. In addition to this, a qualitative comparison was done by comparing four gaze plots to each other. These four gaze plots visualize the fixation durations on the male and female suspect and between the training and no-training group.

Lastly, for the third hypothesis, the NASA-TLX was utilized. To assess the subjective mental workload of the participant, they were asked to fill out the NASA TLX. For this, the weighted average was calculated. The fourth question is negatively phrased to avoid response bias. This presents the self-perceived cognitive load of the participant. These scores of the training group and of the control group were compared to each other by using an independent sample t-test.

Results

After the research has been completed, the gaze samples were coded, and the questionnaire data was analysed. In the following, the three hypotheses were tested, and the outcomes will be presented.

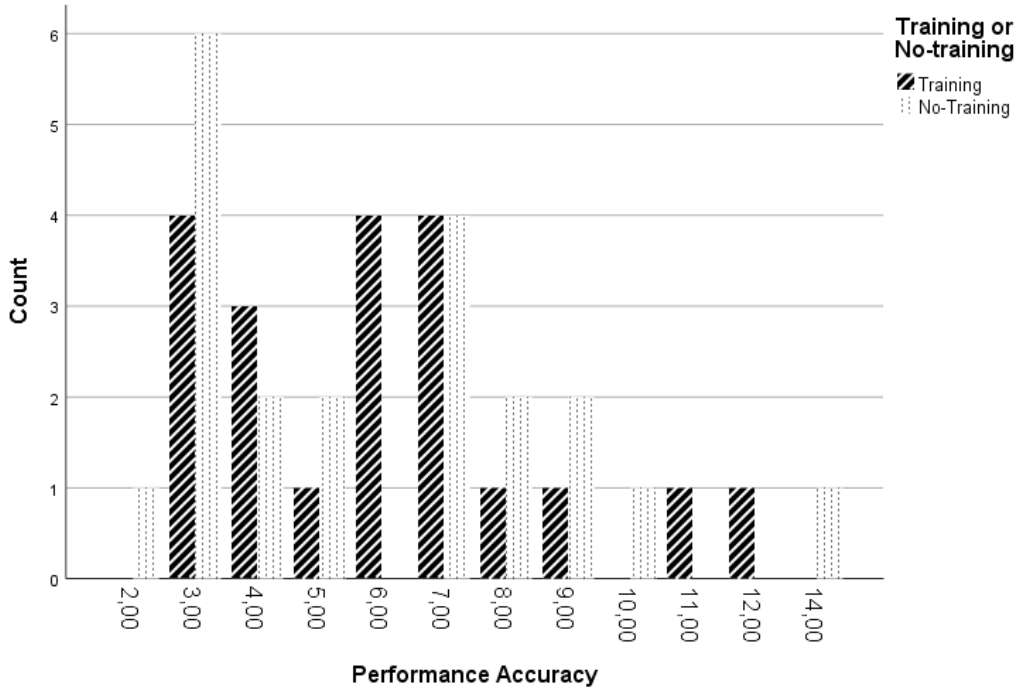
Hypothesis 1:
"Interviewers in the training group will have more accurate veracity judgements than interviewers in the control group"

The results for each score are visualised in Fig. 1. In the no-training condition six out of 20 participants reached a score of 3, which is considered a low score. Only four out of 21 participants in the training condition reached this low score. The lowest score is 2, which was reached by one participant in the no-training condition. Four people in the training condition and in the no-training reached a score of 7, which is still considered a low score. Only one participant reached the full 14 points. This participant belongs to the no-training condition. The second highest score 12, has been achieved by a participant in the training condition.

An independent-samples t-test was conducted to compare the veracity judgement accuracy of the training group to the no-training group. As seen in Appendix L and Appendix M there was no significant difference in the scores for training (M= 6.05, SD=2.58) and no-training (M=5.9, SD=3.08); t(39)=-0.16, p= 0.87. Therefore, the first hypothesis can be rejected.

Figure 1

The veracity judgement accuracy compared between the training and no-training group



Note The label Count on the Y-axis refers to the number of participants who reached the score on the x-axis

Hypothesis 2:

"Interviewers in the training group will have longer fixations durations on the upper face area (e.g. eyes and brows) than interviewers in the control group."

Regarding the second hypothesis, an independent samples t-test was conducted to compare the fixations durations on non-verbal deception cues (namely, upper face area, eyes and brow area) in the training group and no-training group. In addition to these four gaze plots were created to visualize the fixation durations on the male and female suspect.

Taking an explorative look at Fig. 2, it becomes evident that the experimental and control group do not differ much. The experimental group has more distractions on the side than the control group. This means that they have fixation durations on the table or other objects in the background. Both groups have many attention clusters on the face.

In Fig. 3, a similar trend can be recognized. Experimental and control group, both have many attentional clusters on the face. Nonetheless, the experimental group has many attentional clusters on the side, apart from the suspect, on the table and the white board in the

background. From this explorative look, it can be said that there is no difference between training and control group, considering both subjects.

The difference between fixation duration means in the training group and the no-training group regarding the Brow Area is indifferent. For the Brow Area, there is no significant difference in the scores for training (M=20.21, SD= 17.41) and no-training (M=20.37, SD=20.25); t(37)=-0.026, p=0.979.

Next to that, the difference between training and no training group on the fixation duration mean on the eyes is barely noticeable too. For the eye area, there is no significant difference in the scores for training (M=45.15, SD= 33.71) and no-training (M=36.08, SD=24.13 either, t(37)=0.962, p=0.342, as seen in Table 3 and 4.

Last, the mean is slightly higher for the training than for the no-training group when comparing the upper face area. Nonetheless in Table 3 and 4, there is no significant difference found in the scores for training (M= 164.60, SD= 95.74) and no-training (M=142.45, SD= 71.47) for the upper face area, t(37)=0.815, p=0.42. Therefore, the second hypothesis can be rejected.

Table 3

Group Statistics comparing the fixation duration of the non-verbal deception cues between the training and no-training group

Group Statistics

	Group	N	Mean	Std. Deviation	Std. Error Mean
BrowArea	Training	20	20,21	17,41	3,89
	No Training	19	20,37	20,25	4,64
Eyes	Training	20	45,15	33,71	7,54
	No Training	19	36,08	24,13	5,54
UpperFaceArea	Training	20	164,60	95,74	21,41
	No Training	19	142,45	71,47	16,40

Figure 2

Gaze Plot of 11 participants in the experimental group (left) and gaze plot of 8 participants in the control group (right) on the female suspect

Figure 3

Gaze Plot of 12 participants in the experimental group (left) and gaze plot of 8 participants in the control group (right) on the male suspect

Hypothesis 3:

"Interviewers in the training group will have a lower self-perceived cognitive load than interviewers in the control group."

An independent-samples t-test was conducted to compare the self-perceived cognitive workload in the training group and no-training (=control) group. The self-perceived cognitive load has been measured by using the NASA-TLX scores for each question and for the total cognitive workload (=average of all subscales). In addition to this a bar chart was created to visualize the mean differences (Fig. 4).

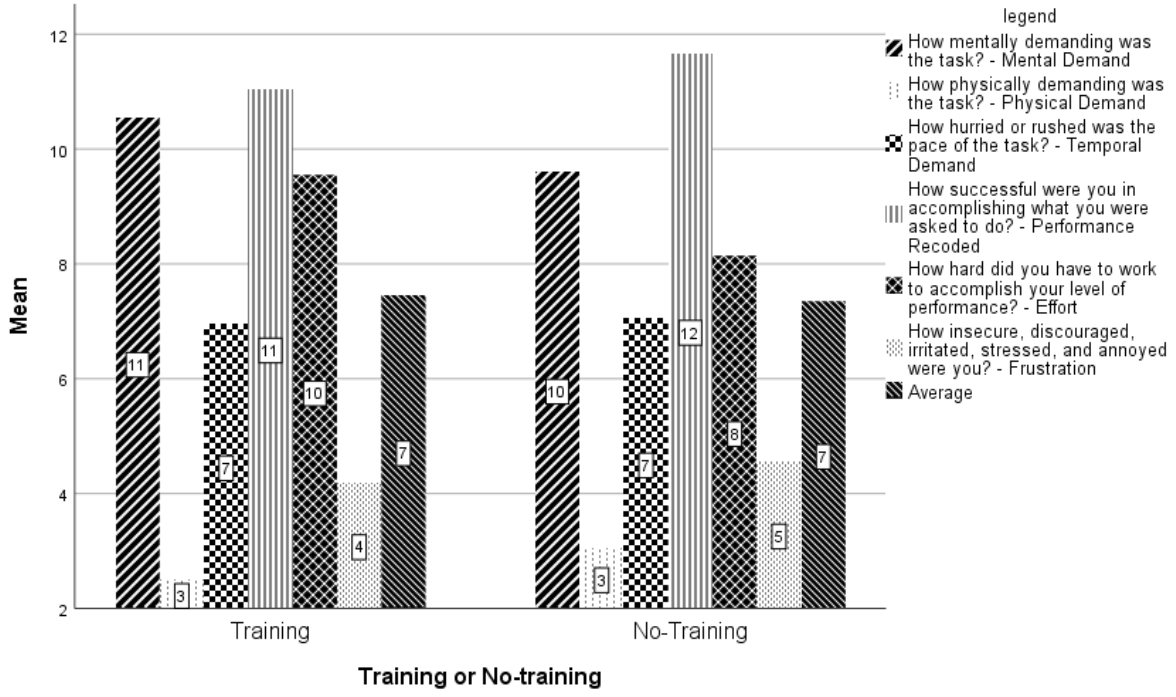
As can be seen in Figure 4 and Table 5 (see Appendix L), the means in both groups are similar. The mean for the mental demand sub question is slightly higher in the training group (M= 10.55) than it is in the no-training group (M= 9.62). The scale for effort also displays a slightly higher mean for the training (M=9.55) than the no-training group (M=8.14). The bar representing the Average of NASA-TLX scores shows that the mental workload score is similar for both groups.

As seen in Table 5 (see Appendix L) and Table 6 (see Appendix M), it becomes evident that, except for physical demand, none of the mean differences are significant. For physical demand, there is a significant difference in the scores for training (M=2,5, SD=1,73) and no-training (M=3,05, SD=2,8) conditions, t(39)=0,75, p<0,05.

Taking into consideration the average total scores, there is no significant difference for training (M=7.46, SD=2.66) and no-training (M=7.34, SD=2.38) conditions, t (39) =0.149, p=0,883. Therefore, the third hypothesis can be rejected.

Figure 4

Bar Chart representing the scores per subscale and mean average of the NASA-TLX for the training and no-training group



Discussion

This study examined the effect of training about non-verbal deception cues on the veracity judgement accuracy and cognitive load of laypeople.The results show that all hypotheses are rejected thus an effect of training could not be observed In the following, we will discuss the three hypotheses of the current research, compared to the results and previous findings.

First, the results show that there was no significant correlation between being in the training session and having more accurate veracity judgements. This suggests that the veracity judgements of this sample did not improve through training. Additionally, Fig. 1 showed that the participants in the training condition did not necessarily have better scores. Sometimes they made several mistakes, such as the no-training group. It was hypothesized that their judgements would have been more accurate. This was because in the past there could be some improvements made just through training of deceptive cues. This was suggested by Porter et al. (2010). They also specified in their research that deception detection schemas based on facial expressions improved better than it did for narratives for example. This supported the hypothesis. Nonetheless, it is not surprising that in this study, the training was ineffective. The training by Porter et al. (2010) endured two days, provided more in-depth information, and practice. Despite this longer session, the results only showed some improvement (from 50% to 60% accuracy). This study only had a short 5-minute session which is not comparable to a two-day unit. As stated in the bachelor thesis by Jungfer (2021) the participants gained knowledge about non-verbal deception cues but were not able to improve their veracity judgements. Porter et al. (2010) and Bond (2008) argued that the increase in veracity judgement accuracy can be due to practice and not only through training. Therefore, expertise is vital to accurately detect deceit. Important to mention in the research by Porter et al. (2010) is that they suppose that their training was effective because of myth dissolution. A similar effect was noticed by Jungfer (2021) that the biases towards deceptive cues could be easily resolved and registered.

Second, the results show that there was no significant difference between the training and no-training group regarding fixation duration on significant non-verbal cues, namely the upper face area, eyes and brows. It was presumed that the trained participants incorporate their newly gained knowledge and focus on these non-verbal cues. This was hypothesized because they were trained on these cues. If the participants could apply the knowledge, they previously learned, they should have paid more attention to these cues. Nonetheless, this was

not evident in the eye-tracking data. This could be because the training was brief and that they could not internalize and practice the new knowledge. In addition to this, non-verbal cues are highly criticized to be significant cues for deception (Hartwig et al, 2014; Nortje & Tredoux, 2019). People can show various cues for a variety of reasons (e.g., nervousness) which makes it difficult to set them as accurate signifiers of deception. Therefore, the assumption could not hold. In addition to this, a longer fixation duration is indicative of higher cognitive load (Freire et al., 2004), but it did not differ much in both groups which suggests that it was not more or less cognitively loading for either group.

Third and last, there was no significant correlation found between self-perceived cognitive load and training. The results suggest that the training did not influence the self-perceived cognitive load since the scores are similar. Considering, the participants in the training condition learned new information, there should have been a difference to the control condition. According to cognitive load theory (Hanway et al, 2021), the cognitive load could have been higher because the participants needed to form new cognitive schemas (germane load) and paid more attention to non-verbal cues. In addition to this, they needed to process the information given by the suspect, process it and form a decision based on this (intrinsic load). Nonetheless, it was hypothesized that the cognitive load will be lower because the training was supposed to lower the pressure of the participant. The idea was that the training makes the participant aware of what they should pay attention to, so that they could form a cognitive schema and apply it soon after. It was supposed to make the participant familiar with the situation and therefore less loading for them to interact (Hanway et al., 2021).

Nonetheless, this was not the case. It could be argued that this might have changed when the training would have been more extensive. Cognitive load can be influenced by several outside (extrinsic load) or personal factors as well (Hanway, 2021). For example, one participant said that she scored high on the frustration scale because she was frustrated since she does not like to do interviews at all. Doing an interview per se can be already cognitively loading for some people while it is easier to do for others, especially with a structured question guide.

Impact of Research

This study indicated that non-verbal cues are ineffective as signifiers of deception. This finding aligns with previous literature. Nortje and Tredoux (2019) suggested that physiological measures are prone to error and not reliable. The participants who were trained in de-

ceptive cues scored similarly to the ones in their veracity judgement accuracy than those who were not trained. Nonetheless, it needs to be considered that there is a finding (Porter et al., 2010) which suggests that training and/or practice of these cues leads to a significant improvement of deception detection. This is not in line with our finding, but this could be attributed to the briefness of the training.

In this study training had no significant impact on the veracity judgement accuracy and the cognitive load. Nonetheless, maybe repeating this study design on a larger scale with modification (e.g., more training) could have led to more significant results. This suggests that research in this area has a prospective outlook. It would be of advantage to further research it. Improvement of the training and raising awareness about cues might have a positive influence on the future work of suspect interviewers.

Limitations

Some of the results might be based on the condition of some limitations in the study design. The first limitation is the length of the training. A longer training session could have led to significant results. Practice and feedback could have improved the judgement accuracy of the participants. Nonetheless, it would have gone beyond the scope to make a longer training session. The goal was to analyse if a short training could shift the attention of the participants to non-verbal cues. The participants managed to require some short-term knowledge about deceptive cues but could not apply this knowledge appropriately. In line with this, the biases of the participants could be debunked, and the participants showed improvement (Jungfer, 2021). There was no effect on the veracity judgements, but to get a rough impression of a possible effect of training, this was sufficient.

Next to that, it can be criticized that the scenario is not naturalistic. To explain, the whole scenario is just acting, and the suspect is an actor. One could argue that the actor cannot replicate the non-verbal deception cues correctly and lacks the nervosity/emotional state of a real suspect. This influences the judgement accuracy of the interviewer. In addition to that the participant is aware that the scenario is acted and therefore, the distinction between lying and truth-telling is more complicated. However, the actor knows the real (fictional) story and needs to come up with a good, consistent story on the spot. The actor may not be perceived as nervous or otherwise, they could be exposed as a liar. In a similar fashion, a real suspect can construct a story, rehearse it several times and reduce their nervousness through this. This makes it difficult to detect deception, which was replicated in this study.

In accordance with this, another limitation were the notes. The participants could take notes during the interview and had all the material (question guide, training sheet in the training group, role sheet) laid out to them. This could be highly distracting for a participant. The people looked up to the participants and down again on the notes. This can distort the eye-tracking results because of the ongoing movement. In addition to this, the participant cannot always fixate on the suspect, and this might lead to less fixation durations and missing out on important cues. Both are important to consider because this could have changed the results on a high range. If the participants would have always fixated on the participant, they might have noticed more cues. This could have improved their veracity judgements. In addition to this there might have been longer fixation durations which could have indicated a higher cognitive load in one of the groups. Nonetheless, the study was designed that way so that the participant did not spend a lot of thought on remembering all the information but could instead focus on the conversation. For example, if the research team would have not provided the question guide, then the participants would have used up more time and thought on constructing questions. In addition to this, it would have been difficult to standardize the procedure, if each participant would have needed to think of their own questions. For the participant, it is less cognitively loading to look up the information about the incident on the sheet instead of remembering it all by heart. It is not realistic to expect the participant to remember all the details of the incident, all the questions and the whole training within half an hour:

Additionally, the study was limited by the fact that the sample only included laypeople and could not draw a comparison to experts. Nonetheless, the scope of the study was limited to laypeople and finding experts who are willing to participate would have been a complicated task. In addition to this, we provided basic knowledge to the laypeople, and the experts probably are aware of the information. Therefore, we limited the scope to laypeople to estimate if a short training could bring some significant effects.

Future Research

Further, this study could be redone with some modifications in the future. First, it could be possible to replicate the study but with a more extensive training. It would be interesting to see how extensive training could appeal to the long-term memory. Instead of having one short session, there could be multiple short sessions. For example, there could be a session every second day of the week. In these sessions the non-verbal cues will be explained and then practiced. After two weeks the interview can be conducted. Then it can be tested if the

veracity judgements accuracy improved in comparison to a control group.

Second, in future research it could be possible to replicate the experiment but with an advanced interview strategy. In this experiment a standardized question guide was provided to the participants. In the future, it could bring interesting results to teach the participants one interview strategy e.g. information-gathering interrogation (Hartwig et al, 2014) and test if it improves their scores and how it affects the cognitive load. Alternatively, it could be tested how different groups of participants perform when using different interview methods. According to Nortje and Tredoux (2019) interview strategies hold a great value for the future of deception detection. Therefore, it might be more practical to focus on developing newer interview styles and other interview questions.

Third, it would be interesting to see how laypeople performed in comparison to experts. Therefore, doing the experiment again with four groups, instead of two. This could be done by having one training and one control group for the laypeople, one training and one control group with experts. The training should be extensive enough that the experts learn some new information, but simple enough that laypeople can easily follow. It should mainly be focused on non-verbal cues but can incorporate verbal strategies too. Each training group would receive the training. Then it would be compared how the experts and laypeople performed in both training and control conditions:

Fourth and last, the experiment could be changed by switching the roleplay with videotapes. For example, the participant could look at a video of a real crime-investigation or someone who is lying and not only acting. This could be of advantage because it increases the authenticity of the non-verbal cues since it is difficult to replicate this in an acting, lab or non-naturalistic setting.

Conclusion

The answer to the research question is that training had no significant effect on cognitive load or the veracity judgement accuracy. The rejection of the first hypothesis showed that a short training is insufficient while more practice has a prospective outlook for significant results. The second hypothesis showed that non-verbal cues are not necessarily significant deceptive cues. Lastly, the third hypothesis showed the self-perceived cognitive load did not improve or was impaired by the training session. Even though all hypotheses could be rejected, it can be stated that this study gave an interesting outlook on the relation between deception detection and eye tracking. As mentioned above, repeating this study with more practice session might lead to more significant results.

Reference

Bessonova, Y. V., & Oboznov, A. A. (2018). *Eye movements and lie detection.* Paper presented at the International Conference on Intelligent Human Systems Integration.

Burgoon, J. K., Kelley, D. L., Newton, D. A., & KEELEY-DYRESON, M. P. (1989). The nature of arousal and nonverbal indices. *Human Communication Research, 16*(2), 217- 255.

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

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

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

DePaulo, B. M., Kashy, D. A., Kirkendol, S. E., Wyer, M. M., & Epstein, J. A. (1996). Lying in everyday life. *Journal of personality and Social Psychology*, *70*(5), 979.

Ekman, P., & Friesen, W. V. (1974). Detecting deception from the body or face. *Journal of personality and Social Psychology*, *29*(3), 288.

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

Fallis, D. (2010). Lying and deception.

Farnsworth, B. (2020, May 12). 10 Most Used Eye Tracking Metrics and Terms [Blog post]. Retrieved from https://imotions.com/blog/10-terms-metrics-eye-tracking/

Freire, A., Eskritt, M., & Lee, K. (2004). Are eyes windows to a deceiver's soul? Children's use of another's eye gaze cues in a deceptive situation. *Developmental psychology, 40*(6), 1093.

Freud, S. (1953). Fragment of an analysis of a case of hysteria (1905 [1901]). In *The Standard Edition of the Complete Psychological Works of Sigmund Freud, Volume VII (1901-1905): A Case of Hysteria, Three Essays on Sexuality and Other Works* (pp. 1-122).

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

Hanway, P., Akehurst, L., Vernham, Z., & Hope, L. (2021). The effects of cognitive load during an investigative interviewing task on mock interviewers' recall of information. *Legal and criminological psychology*, *26*(1), 25-41.

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.

Inbau, F. E., Reid, J. E., & Buckley, J. P. (1986). *Criminal interrogation and confessions.* Baltimore: Williams & Walkins.

Jungfer, T. (2021). *Deception Detection and Eye-tracking: How biases effect our veracity judgement ability* (thesis).

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological review, 87*(4), 329.

Köhnken, G. (1989). Behavioral correlates of statement credibility: Theories, paradigms, and results. In *Criminal behavior and the justice system* (pp. 271-289): Springer.

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/

Mahon, J. E. (2008). The definition of lying and deception.

Nortje, A., & Tredoux, C. (2019). How good are we at detecting deception? A review of current techniques and theories. *South African Journal of Psychology*, *49*(4), 491-504.

Porter, S., Juodis, M., ten Brinke, L. M., Klein, R., & Wilson, K. (2010). Evaluation of the effectiveness of a brief deception detection training program. *Journal of Forensic Psychiatry & Psychology, 21*(1), 66-76.

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.

Rayner, K. 1978. Eye movements in reading and information processing. Psychological Bulletin, 85, 618-660.(Just & Carpenter, 1980)

Rayner, K. 1998. Eye movements in reading and information processing: 20 years of research. Psychological Bulletin, 124, 372–422. [23] Richards, J., & G

Van Edwards, V. (2020, July 10). Is lie to me lying to you? The truth behind the hit show. Retrieved March 23, 2021, from https://www.scienceofpeople.com/is-lie-to-me-true/

Vrij, A. (2004). Why professionals fail to catch liars and how they can improve. *Legal and criminological psychology*, *9*(2), 159-181.

Vrij, A., Mann, S. A., Fisher, R. P., Leal, S., Milne, R., & Bull, R. (2008). Increasing cognitive load to facilitate lie detection: The benefit of recalling an event in reverse order. *Law and human behavior*, *32*(3), 253-265.

Walczyk, J. J., Griffith, D. A., Yates, R., Visconte, S. R., Simoneaux, B., & Harris, L. L. (2012). Lie detection by inducing cognitive load: Eye movements and other cues to the false answers of "witnesses" to crimes. *Criminal Justice and Behavior*, *39*(7), 887-909.

Appendices

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

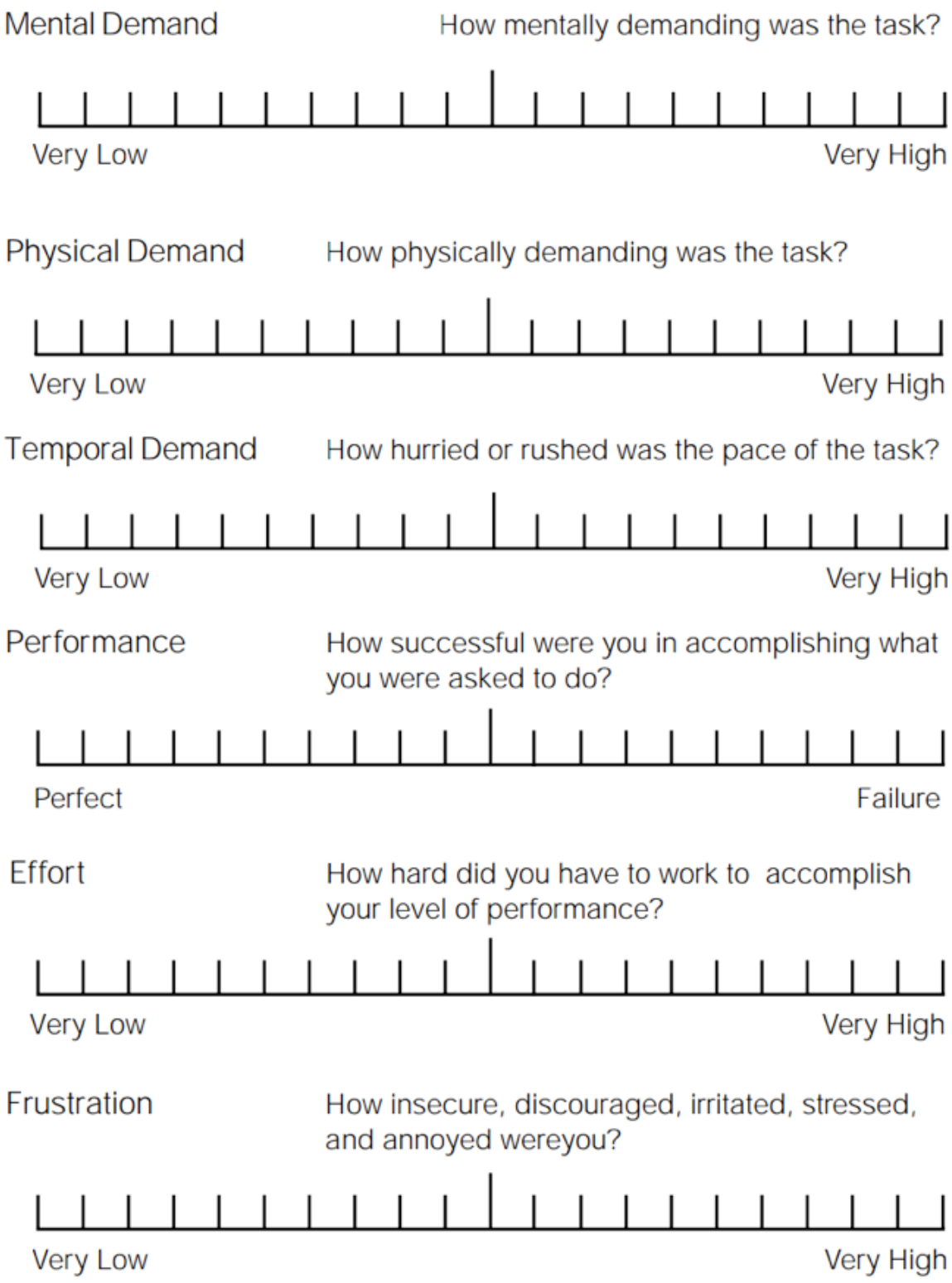
Appendix B

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


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 yes: 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

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

Training

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 can not 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

Table 1:

Group Statistics comparing the veracity judgement accuracy between the training and no training group

	Training or No-training	N	Mean	Std. Deviation	Std. Error Mean
Veracity Judgement Accuracy	Training	20	6,05	2,58	,58
	No-Training	21	5,90	3,08	,67

Appendix J

Table 2:

Results of the Independent Samples T-Test comparing the veracity judgement accuracy between the training and no-training group

		Levene's Test		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Veracity Judge-ment Accu-racy	Equal variances assumed	1,45	,24	,16	39	,87	,16	,89	-1,66	1,95
	Equal variances not assumed			,16	38,4	,87	,16	,89	-1,65	1,94

Appendix K

Table 4:

Independent Samples T-Test comparing the fixation duration of the non-verbal deception cues between the training and no-training group

		Levene's Test		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
BrowArea	Equal variances assumed	,65	,43	-,03	37	,98	-,16	6,04	-12,39	12,08
	Equal variances not assumed			-,03	35,55	,98	-,16	6,06	-12,46	12,14
Eyes	Equal variances assumed	,58	,45	,96	37	,34	9,08	9,43	-10,04	28,19
	Equal variances not assumed			,97	34,44	,34	9,08	9,35	-9,92	28,08
Upper FaceArea	Equal variances assumed	,27	,60	,81	37	,42	22,15	27,17	-32,9	77,2
	Equal variances not assumed			,82	35,08	,41	22,15	26,97	-32,59	76,89

Appendix L

Table 5

Group Statistics for the NASA-TLX subscales between the training and no-training group

	Training or No-training	N	Mean	Std. Deviation	Std. Error Mean
Mental Demand	Training	20	10,55	4,52	1,01
	No-Training	21	9,62	4,85	1,06
Physical Demand	Training	20	2,50	1,73	,39
	No-Training	21	3,05	2,80	,61
Temporal Demand	Training	20	6,95	5,53	1,24
	No-Training	21	7,05	4,31	,94
Performance Recoded	Training	20	11,05	3,56	,8
	No-Training	21	11,67	4,29	,94
Effort	Training	20	9,55	4,61	1,03
	No-Training	21	8,14	3,51	,77
Frustration	Training	20	4,20	3,19	,71
	No-Training	21	4,57	4,41	,96
Average	Training	20	7,47	2,67	,6
	No-Training	21	7,35	2,39	,52

Appendix M

Table 6:

Independent Samples T-Test comparing the NASA-TLX subscales between the training and no-training group

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference Lower	Upper
Mental Demand	EVA	,09	,76	,64	39	,53	,93	1,47	-2,04	3,9
	EVnA			,636	39	,53	,93	1,46	-2,03	3,9
Physical Demand	EVA	5,22	,028	-,748	39	,459	-,55	,73	-2,03	,93
	EVnA			-,76	33,58	,45	-,55	,72	-2,02	,92
Temporal Demand	EVA	2,60	,16	-,06	39	,95	-,1	1,54	-3,22	3,02
	EVnA			-,06	35,9	,95	-,1	1,55	-3,25	3,05
Performance Recoded	EVA	1,05	,31	,5	39	,62	,62	1,24	-3,12	1,88
	EVnA			,5	38,3	,62	,62	1,23	-3,11	1,87
Effort	EVA	,66	,42	1,1	39	,28	1,4	1,28	-1,17	3,99
	EVnA			1,1	35,51	,28	1,4	1,28	-1,2	4,01
Frustration	EVA	1,46	,23	-,3	39	,76	-,37	1,2	-2,81	2,07
	EVnA			-,31	36,42	,76	-,37	1,2	-2,8	2,06
Average	EVA	,8	,36	,15	39	,88	,117	,79	-1,48	1,72
	EVnA			,15	38,02	,88	,12	,79	-1,49	1,72

Note. EVA = Equal variances assumed, EVnA = Equal variances not assumed