Master thesis

Improving veracity judgement accuracy: an eye-tracking study

Jessica M. Schimmel

s2247755

Faculty of Behavioural, Management and Social sciences (BMS), University of Twente

Department: Psychology of Conflict, Risk and Safety

First supervisor: P.J.H. Slijkhuis, MSc.

Second supervisor: Dr. M. Stel

April 28, 2021

Abstract

People have proven to be very skilled liars and the detection of deception remains a challenge until this day. Namely, people perform around chance level when trying to detect deception (Bond & DePaulo, 2006). This experimental study investigated ways to improve accuracy in deception detection. This is done by studying the receiver of lies instead of the sender and by using an eye-tracker on the receiver. With a 2x2 between-subjects design the effects of role type (receiver or observer) and medium type (face-to-face or video) on veracity judgement accuracy and gaze behaviour were investigated. Also, it was tested whether indirect measures outperformed direct measures. Further, the effects of relational and behavioural familiarity on accuracy were investigated. Lastly, additional analyses explored the relation between gaze behaviour (of the receiver of lies on distinct areas of the body of the liar) and accuracy. There were no effects of the role and medium type on accuracy. Role type and medium type did, however, elicit differences in gaze behaviour. Yet, differences in gaze behaviour did not translate to differences in accuracy, indicating that accuracy is not dependent on what has been perceived. However, further analyses (regardless of the conditions) examined the relation between the time that receivers spent on looking at distinct areas on the body of the liar and accuracy. The time that the receiver spent looking at the feet and legs predicted positively accuracy rates, whereas the time spent on hands and arms significantly impaired accuracy. Time spent on the face seemed unrelated to accuracy. These findings seem to support the existence of nonverbal cues to deception which will have implications for the discipline of deception detection. Indirect measures, relational and behavioural familiarity did not have an effect on veracity judgement accuracy.

Keywords: veracity judgement accuracy, eye-tracking, (nonverbal) cues to deception, relational familiarity, behavioural familiarity, direct measures, indirect measures

Introduction

Deception is something pervasive in life and ranges from small non-significant lies to situations where stakes are high. Deception is defined as a deliberate act to foster a belief in others which the deceiver considers as false (Zuckerman et al., 1981). Important aspects of this definition are that deception is considered as intentional and directed towards another. Therefore, self-deception, sarcasm, and mistaken lies are not considered deceptive. This definition, however, includes concealments and distortions.

There is a consensus in the literature that humans are poor at detecting lies. Namely, people perform around chance level when trying to detect deception (Levine, 2015; Bond & DePaulo, 2006; Hartwig et al., 2004). This study will examine how deception detection can be improved. First, current valid cues to deception will be addressed. Later it will be examined whether looking at these cues improves accuracy. Also, detection performance on direct and indirect measures will be compared. Other factors that might influence deception detection are being familiar with the deceiver and/or being familiar with cues to deception. Lastly, potential influences of the medium type and conversational involvement (i.e. role type) will be discussed. We will start with cues to deception.

Cues to deception

A deception cue is a signal or indication that someone is lying. For instance, fidgeting could be a signal that accompanies deceitful behaviour (DePaulo et al., 2003). Most of the cues that were held to indicate deceptive messages seemed unrelated to deception. DePaulo et al. (2003) did an extensive meta-analysis of 158 cues to deception. Of those cues, only a few seemed significantly related, while the vast majority appeared unrelated to deception. Nonverbal cues that repeatedly have shown to be related to deception are hand movements (i.e. illustrators), and fidgeting. Also, verbal and vocal involvement, verbal immediacy, cooperativeness, thinking hard and indifference seem related (Sporer & Schwandt, 2007; DePaulo et al., 2003; Bond et al., 2015; Vrij et al., 2004; Mann & Vrij, 2006).

In addition, there is support that reliable cues can be derived from emotions (muscles) in the face (Stel & Van Dijk, 2018). This study showed that people were able to distinguish between real and fake negative emotions. It seems to indicate that fake negative emotions leak the genuine emotion. Ekman and Friesen (1969) introduced the leakage hypothesis. It is reasoned that – based on Darwin's research of micro emotional expressions – a distinction can be made between muscles that can voluntarily and/or involuntarily be activated. If muscles cannot be activated voluntarily (i.e. deliberate), it is neither possible to voluntarily inhibit

muscles that are activated spontaneously while experiencing an emotion. This ‘leakage’ of emotions exhibits the real experienced emotions despite attempts of concealments (i.e. faking macro expressions). Involuntary muscles are mainly found in the face. Porter et al. (2012) and Warren et al. (2009) found strong support that some facial muscles, in particular muscles in the upper face, are not controllable and cannot be inhibited and therefore leak emotions. Accuracy rates for observers of the face however were lower than observers of the body. This was explained that the majority was misled by the macro expressions, and did not notice the barely perceptible micro-expressions. Micro-expressions distinct themselves from macro expressions in duration: micro-expressions occur within a fraction of a second, whereas macro expressions last for two to five seconds.

Body movements are in general more controllable and therefore, theoretically, leak less of the genuine experienced emotions. However, observers of the body were found to be more accurate in comparison to observers of the face (Ekman, 2003). This is because people put less effort in censoring their body movements based on the belief that observers mainly focus on the face (Gullberg & Holmqvist, 2001; Ekman, 2003). Yet overall, the effects found for all the nonverbal cues appeared to be weak in their relation to deception (DePaulo et al., 2003; Vrij et al., 2019).

Nonverbal cues have not received lots of support, either due to lack of a comprehensive underlying theory or inaccuracy of measuring cues (DePaulo et al., 2003; Bond & DePaulo, 2006; Bond et al., 2015). Vrij et al., (2019) state that nonverbal cues have not proven to be very reliable so far, because these cues have not been studied correctly. They suggest that there should be differentiation within a nonverbal cue. For instance, when considering nonverbal cues for facial expressions, smiles seem an obvious one. Smiles in general do not seem to be related to deception, yet when making a distinction between types of smiles (real and false smiles) either one might be related (Ekman & Friesen, 1982). Similarly, most deception research does not make a distinction between, for instance, differences in leg movements and might therefore seem unrelated to deception. But when taking subcategories into account, a cue to deceit might emerge. More concretely, a restless leg might not be related to deception, but switching crossing legs might be related, for instance.

Valid cues to deception are closely linked to the discussion of the influence of direct and indirect measures on veracity judgement accuracy. In such a way that indirect measures steer the attention towards cues that are related to deception. Indirect measures, thus, bypass

held biases (i.e. relying on invalid cues) to steer individuals towards reliable cues (Hart et al., 2009; Street & Vadillo, 2016).

Direct and indirect measures of deception detection

Assessing the veracity of a speaker can be done directly and indirectly. Directly by asking to what extent the speaker is telling the truth and indirectly by asking to assess deception-related constructs (e.g. thinking hard) and converting those into either lie judgements or truth judgements. Street and Richardson (2015) explain that indirect measures steer the focus to useful cues, resulting in higher accuracy rates. However, indirect measures do not always outperform direct measures (Levine & Bond, 2014; Bond et al., 2015). A meta-analysis showed that studies which reached relatively high accuracy rates used indicators that have shown to be reliable indicators of deception (Bond et al., 2015). Studies, in which the direct measures performed better, used cues not being related to deception (Bond et al., 2015; Street & Vadillo, 2016; DePaulo et al., 2003). This explains why indirect measures do not always outperform direct measures: it is dependent on the reliability of the cues which are used. Previously researched indirect measures which have shown to outperform direct measures are cooperative, thinking hard, indifference, and audible immediacy (Bond et al., 2015). In addition, Hart et al. (2009) introduce behavioural change as an indirect measure. They theorize that no systematic cues exist which distinguish liars from truth tellers. Thus, the only valid indirect measure is assessing the occurrence of behavioural changes relative to initially shown behaviour.

People are poor lie detectors and by using indirect measures, people are steered towards valid cues, which will increase the veracity judgement accuracy. We therefore hypothesize that assessing the veracity of a speaker using the indirect measures, cooperative, thinking hard, indifference, audible immediacy, and behavioural change result in higher accuracy rates than direct measures (H1).

Familiarity and deception detection

Besides indirect measures, there is another important factor that is positively associated with accuracy, namely familiarity. Several studies show that familiarity increases veracity judgement accuracy (Reinhard et al., 2013; Reinhard et al., 2011; Zhou et al., 2013; Blair et al., 2010). In deception research, a distinction is often made between three types of familiarity. The first type is informational or contextual familiarity. Reasoned is that people, being familiar with the context in which speakers make certain claims, possess the possibility to verify statements by checking if they correspond with the already possessed knowledge about the context. In other words, verbal content (what is being said) can be compared with the knowledge someone

holds over the context. If this information is incongruent to what is already known, it is likely to assume that someone is lying and vice versa (Levine, 2015).

The second type of familiarity is relational familiarity (Burgoon et al., 1994; Zhou et al., 2013). The relational type refers to the extent to which two people are acquainted. It seems apparent to expect that relational familiarity increases accuracy since a person knows how the sender normally behaves. The observed behaviour can be referenced to the usual behaviour. However, relational familiarity showed to impede an accurate judgement (Zhou et al., 2013). This is explained by the truth-bias. The truth-bias applies to deception in general (e.g. Bond & DePaulo, 2006; Levine et al., 1999). Namely, people are more likely to judge messages as truthful rather than deceitful. But the more two people are familiar, the less likely they are to suspect someone of lying (Burgoon et al., 1994; Van Swol et al., 2012). Therefore, deception detection is likely to be impaired in case of high relational familiarity.

The third category is behavioural familiarity. This type refers to the degree of familiarity with deception detection in general. It also includes the knowledge a person possesses about significant cues to deception (Zhou et al., 2013; Burgoon et al., 1994). It is likely to expect that this type improves the deception detection accuracy, since knowing what cues are diagnostic for deceptive behaviour prevents relying on stereotypical beliefs (e.g. *'people in general tell the truth'*) (Best et al., 2018). We thus hypothesize that behavioural familiarity (i.e. valid preconceived ideas) improves veracity judgement accuracy (H2) and that relational familiarity impairs veracity judgement accuracy (H3).

Role type and deception detection

Next to familiarity, the role that someone has might also influence the accuracy of the detection of lies. Lies can be presented directly towards a person (while interacting) but also indirectly (by observing) from a third-person view. This difference might have implications for deception detection. The most prominent difference between these two roles is captured in the amount of cognitive load one is dealing with (Vrij & Mann, 2001). As an observer, the sole task is to conduct a veracity assessment of the information which is given by a person. The immediate receiver (i.e. interviewer) needs, besides assessing the veracity, to lead the conversation and anticipate signals. The immediate receiver needs to fulfil more tasks (i.e. multitasking) in comparison to the observer.

Multitasking is performing two or more tasks simultaneously (Salvucci & Taatgen, 2008). To a certain extent humans are powerful in multitasking as long as the activities become automatic processes (e.g. talking and walking) (Paivio & Desrochers, 1980). For other tasks,

multitasking is impaired, because of the limited cognitive processing capacity. Difficulties occur when tasks require the same resource (Salvucci & Taatgen, 2008; Phillips et al., 2008). When looking at emotion detection tasks, Phillips et al. (2008) found that performances were significantly impaired by a secondary task. At the same time, research showed that differences exist between individuals in their working memory capacity (Kane et al., 2007). Therefore, presumably, some individuals are likely to be more limited in a deception detection task than others while being a receiver. Yet overall, a process of selecting prior to encoding stimuli occurs, which might result in observing and processing less (potential) deception cues. In other words, receivers require more cognitive resources which cannot be devoted towards lie detection (Vrij & Mann, 2001).

Several studies compared accuracy rates in lie detection between conversational partners and observers (e.g. Buller et al., 1991; Stiff et al., 1992). Results showed that indeed observers had higher accuracy rates. We therefore hypothesize that observing leads to higher veracity judgement accuracy rates than receiving (i.e. being a direct conversational partner) (H4). However, Hartwig et al. (2004) did not find any differences in accuracy between police officers who conducted the interview themselves and observers who judged videotaped versions. Except that they compared the groups on different mediums, this in fact might have something to do with that receivers have the advantage of conversational involvement. Higher conversational involvement, and more specifically, the motivation to engage in effortful cognitive tasks, led to better discrimination of truthful and deceitful messages (Reinhard, 2010). This might be a result due to that information being processed via the central route, in case of effortful cognitive tasks, rather than via the peripheral route. The peripheral route relies namely more on stereotypical beliefs and shortcuts (Reinhard, 2010).

However, conversational involvement might influence accuracy in the opposite direction (i.e. more negatively) as well. Namely, being more involved in an interaction led to an increase in synchronously in nonverbal behaviour (Altmann et al., 2012). Mimicking also involves synchronous behaviour. Mimicking in turn leads to understanding or empathizing with the interaction partner (Stel et al., 2009). Therefore, receiving is presumably more likely to impede accuracy, than observing. This reasoning again substantiates our fourth hypothesis.

Medium type and deception detection

Hartwig et al. (2004) researched differences between observers and receivers, but at the same time they compared the groups on different media (face-to-face and videotaped versions). The media types, through which communication takes place, might also influence deception

detection. The media types that are referred to are interacting face-to-face (F2F) and via video call. When comparing these media, there might be some negative implications when communicating via a video in comparison to interacting F2F. First, a screen may lead to more distractions (e.g. pop-ups). But maybe, more importantly, the quality of the camera and internet connection could disturb the process substantially. As a result, more cues are missed due to either a lack of quality of material or stress. It thus is reasoned that F2F is the richest medium. The screen works less robustly due to resolution and sound quality for instance. The video medium might also be a reason why cues to deception are not or to a lesser extent shown.

This has to do with the distance of the modality. DePaulo et al. (1996) have shown that lying more often occurred in distant communication modalities (e.g. by telephone) than in more direct modalities (e.g. F2F). Communicating via video is experienced as more distant and this could have two opposite effects. First, liars experience a false sense of safety, resulting in censoring less of their behaviour, and thus that more cues appear via the video medium. On the other hand, the distance in the video might lead to 'distance safety', resulting in people feeling less distressed, and thus fewer cues appear. The authors explain it in concordance to the second option. Namely, lying is an act that causes, for most people, distress and therefore liars feel uncomfortable (DePaulo et al., 1996; DePaulo et al., 2003). Experiencing this 'distance safety' might lead to being more confident and comfortable during the act of lying and therefore showing off fewer cues. This might imply that receiving lies through a video – a more distant modality – leads to less accuracy in detecting lies, because of less prominent cues which are caused by a reduction of experienced distress in the deceiver. In this situation, the screen barrier might influence the behaviour of the liar and this in turn potentially influences deception detection. Therefore, we expect that using the F2F medium leads to higher veracity judgement accuracy than using the video call as medium (H5). It is also hypothesized that observing in combination with the F2F medium leads to the highest accuracy rates (H6).

Presentation modality and gaze behaviour

The deception presentation might influence gaze behaviour for the two earlier discussed modalities: presenting a lie directly at a conversational partner (immediate receiver) vs indirectly (observing) and presenting a lie via a F2F medium vs a video medium. First, the role type will be discussed with its expected implications for gaze behaviour, followed by the medium type.

A cognitive demanding task leads to higher eye movement rates, i.e. saccades (Ehrlichman & Micic, 2012). This is expected to apply more to the role of a direct

conversational partner than that of an observer, because this task is more cognitively demanding. These eye movements are not necessarily a function of visual input. They are called nonvisual eye movements since they also occur in the dark and during sleeping: they are found to be related to cognitive processes. Since the eye movements are expected to occur more frequently for the receiver and do not serve as visual input, the direct receiver might miss out on more nonverbal visual information. We therefore hypothesize that observers spent more time on the target in comparison to receivers (H7). Next to the effects of role type (i.e. observer/receiver), there might be an influence of the medium type on gaze behaviour.

Another difference between the media video call and F2F is the fact that the video call only reveals the top of the torso and the face whereas in a F2F situation the entire body can be perceived. One can logically conclude that for this reason observing or communicating via a video leads to more time spent on the face of the target in comparison to observing or communicating F2F (H8). Since discussed earlier that hand and arm movements might contain important nonverbal cues to deception, we predict that the time spent on the hands and arms of the target is positively associated with accuracy. Therefore, we finally hypothesize that, the more time spent on arms and hands, the higher the accuracy (H9).

Studying gaze behaviour is important. Of all senses, the eyes are the most dominant (Colavita, 1974). Meaning that humans rely more on visual stimuli in comparison to other sensory information (Sinnett et al., 2007). Linking this fact to deception detection, one could logically conclude that individuals base their veracity judgements more on what they see than on what they hear. In fact, Koppen et al. (2009) found a decrease in auditory sensitivity when visual stimuli were present simultaneously. This makes it so much more relevant to study gaze behaviour and eventually search for valid nonverbal cues since cues (perceived by the eye) have higher chances of being processed.

Current research

In general terms, this research aims to identify factors that could contribute to the improvement of deception detection accuracy. More specifically, we want to examine if there are differences between role (immediate receiver or observer in third-person view) and medium type (F2F or via a video) in veracity judgement accuracy and gaze behaviour. This study also examines the influence of relational and behavioural familiarity on veracity judgement accuracy. Finally, performance on direct and indirect measures will be examined. The following main research questions are formulated: *'What is the influence of the deception presentation modality and conversational involvement on veracity judgement accuracy?'* (RQ1). This

question researches the effects of the role and medium types and whether either one or a combination of the two factors elicits higher accuracy rates. The second research question is *'What is the impact of the deception presentation modality on gaze behaviour?'* (RQ2).

Recent studies on deception detection focus on verbal cues and leave the nonverbal cues for the most part neglected (Vrij, 2019). This is a regrettable omission. Including eye-tracking as a tool might give new insights into the discussion of deception since eye-tracking data of individuals with high accuracy scores might identify new important areas of the body containing nonverbal cues related to deception. Therefore, this leaves the study with an exploratory part. *The last research question is explorative and will examine what the relation is between gaze behaviour and accuracy, regardless of the conditions (RQ3*). Another unique feature besides adding the eye-tracker is that we want to contribute to the discussion about researching receivers instead of senders. Most existing deception detection studies that include eye-tracking use this tool on the senders (e.g. Proudfoot et al., 2016; Wang et al., 2010). However, deception detection depends on the liar's ability to deceive and the receiver's ability to detect. Thus, approaching this topic from a different perspective by using eye-tracking on the receiver might shed new light on deception detection research. Results of this study might lead to recommendations of best practices regarding lie detection in the forensic field, such as investigative interviewing. Also, results can be relevant in a broader context as, for instance, now more social interactions take place in a digital setting (e.g. a job interview) due to COVID-19, but also due to an increase in technology usage in general.

Method

Participants

A total of seventy-three participants took part in the experiment. Three participants could not finish the experiment, since the eye-tracker did not work on these individuals. Two other participants were in the end also excluded, due to a low gaze sample (that is, percentage of correctly registered eye-movements). Hence, sixty-eight participants (44 female, 24 male) remained for this study (M_{age} = 22.88, SD_{age} = 4.45, range 17-48 years). Of the sample, 38.2% participated for Sona credits and were bachelor Psychology students of the University of Twente. The remaining 61.8% participated voluntarily from other studies or from outside the University. In addition, 57.4% had a Dutch nationality, 32.4% were German and the remaining 10.3% had other nationalities (Bulgarian, Greek, Indian, Latvian, Lithuanian, and Moldavian). Further, 39.7% had corrected vision whereas 60.3% had normal vision. Of the corrected vision group 17.6% wore glasses.

Design

A 2x2 between-subjects experiment was carried out. Participants were randomly allocated to one of the four conditions which resulted from a combination of medium type and role type (see Table 1): The receiver in the video condition (Vid/R), the observer in the video condition (Vid/O), the receiver in the face-to-face-condition (F2F/R), and the observer in the face-to-face condition (F2F/O). Participants in the Vid/R condition interacted with a person via a video call, therefore being an immediate receiver. Participants in de Vid/O condition observed a recorded video call. In the F2F/R condition, the participant interviewed the actor. In the F2F/O condition, two actors were observed while carrying out an interview.

Table 1

Experimental conditions

		Medium type	
		Video	Face-to-face
Role type	Receiver	Video/Receiver	Face-to-face/Receiver
	Observer	Video/Observer	Face-to-face/Observer

Procedure

The experiment was held on the campus of the University of Twente in one of the rooms belonging to the BMS lab. Participants either signed up via Sona on available timeslots or were manually scheduled when they were recruited from outside of the (bachelor) psychology department. In both cases, participants could read before the experiment, in quite general terms, what was to be expected of them. The instructions mentioned that the participants were asked to detect deception. Also, that during an investigative interview they would wear an eye-tracker. And lastly, that they had to classify statements made during the interview as either truthful or deceitful.

Prior to the beginning of the experiment Qualtrics assigned the participants randomly to one of the four conditions. Then, they were welcomed and were instructed to click through the questionnaire until it indicated that they should let the researcher know they finished reading. The questionnaire started with informed consent (see Appendix A), followed by a standardized instruction which differed slightly depending on conditions (see appendix B).

After the participants indicated to be finished with reading the case and task, the Tobi Pro glasses was installed. At this point, participants also received a scratch paper. After calibration of the glasses, the interview began. Receivers asked the pre-printed questions either via a video call or F2F, whereas observers could read them while the researcher asked the

questions to the actor. The actor formulated her answers spontaneously and, on the spot, to let either the lie or truth be as genuine as possible.

After either observing the interview or conducting the interview, the glasses were removed. Lastly, they were asked to continue with the questionnaire and fill out the items and were thanked for their participation. For the procedure as described applies that the steps were carried out by one individual at the time.

Materials and measures

To standardize the interaction as much as possible, detailed cases have been written for both the participant and actor in which it was outlined that a young girl was run over by a car and that the police was left with only one witness (see Appendix B and C). Additionally, a script was written consisting of questions for the criminal investigator – i.e. the participant (see Appendix D or F). The first two questions of the script served as a base rate for the truth, since the answers on these questions were given in the case (the time and location of the accident). The other questions were for instance with regard to what the witness was doing at that specific hour on the crime scene and asking for descriptions of the car and driver. There was no additional script for what the actor had to answer other than an indication of *'lie'* or *'tell the truth'*. There were five truthful and five lie statements. In addition, we strived to let the lies and truths be as genuine as possible by adding pictures to the case of the actor. Contradicting a case or story is less concrete than contradicting pictures at which the actor had to look during the interview (see Appendix E). Lastly, participants, regardless of their role (observer or receiver), had a scratch paper to their disposal with the pre-printed questions on which they could circle their instant judgement being either *'truth'* or *'lie'*. The scratch paper had also space for additional notes (see Appendix F).

Based on the interaction or observation veracity judgement accuracy (*dependent variable*) was measured with 60 items rated on a 7-point Likert scale (see Appendix G). The veracity judgement assessment was two-folded, namely, it was assessed directly and indirectly. Examples of items that measured, directly and indirectly, veracity judgement accuracy were: e.g. *'The target lied about the colour of the car'* (1= strongly disagree to 7 = strongly agree), and *'The target answered immediately after I formulated my question* (thinking hard)' (1= strongly disagree to 7 = strongly agree). This study used indirect measures which in other studies have been validated (Hart et al., 2009; Bond et al., 2015). Next, for each individual accuracy rates were calculated by dividing the correct number of items by the total number of items ($\frac{\text{correct number of items}}{\text{total number of items}}$). An item was classified as correct when it was either answered on

the 7-point scale with a 1,2 or 3, or when an item was answered with a 5, 6 or 7, depending on the case. More concretely, Lisa, for instance, had to lie about the model of the car. A question that assessed this statement was *'Lisa lied about the model of the car'*. If the participant answered this with 5, 6 or 7 (i.e. agreeing) than it was classified as correct, since the case indicated so. Moving on, an overall accuracy score came when adding up all the correct veracity items divided by the total number of items. The direct veracity score was confined by adding only the correct direct assessed items divided by the total number of direct items. The same goes for the indirect and the separate subtypes of indirect measures. Thus, every participant had a total of eight scores (expressed in percentages) on the distinct variants of assessed accuracy.

Gaze behaviour (*dependent variable and independent variable*) was measured with the Tobii Pro 2 glasses. More specifically, gaze behaviour was studied by fixations. A fixation is a cluster of gaze points that are close in space and time. Put differently, it means that the eyes are locked towards a specific area in the scene. These fixations can vary in duration, frequency (counts), visit counts, sequence, time to the first fixation, etc. (iMotions guide, 2017). This study limited itself by only including the total fixation duration, average fixation duration, fixation counts, and visit counts. The total fixation duration entails the total time which is spent on a specific area. The average fixation duration gives information over how long on average the fixations were. For reference, an average fixation lies between 100 and 300 milliseconds (iMotions guide, 2017). Moreover, fixation counts display the frequency of fixations within a specific area. Lastly, visit counts give the number of times when the eyes revisit an area of interest. This means, the higher this number, the more the eyes switched between predefined areas (Tobiipro.com, 2021). These areas are called Areas of interest (AOIs). For purposes of this study are several AOIs created on the face and body of the actor which were, entire face, forehead, eyes/brows, nose area, mouth area, chin, hands/arms, feet/legs, and entire target. For visualization see Appendix H.

The relation between the types of familiarity (*predictor variables*) and accuracy was also examined. Items that served to measure the degree of relational familiarity were *'I know the actor'* or *'I interacted with the actor before'* (1= strongly disagree to 7 = strongly agree). Behavioural familiarity (i.e. preconceived ideas) was measured with items regarding nonverbal behaviour which may or may not have shown to be related to deception. For instance, *'The target avoided eye contact'* or *'I believe that people who lie sit generally more still'*.

Data Analysis

The questionnaire data was directly exported from Qualtrics into IBM SPSS Statistics 26. The eye-tracking recordings were uploaded to the Tobii Pro Software. Then, each video was mapped onto the snapshots using the assistant mapping option and the Tobii I-VT fixation filter. As it turned out, these points deviated frequently from the video and were therefore manually remapped by the researcher. Then, the AOIs were created (see Appendix H). At this point, data could be derived from the AOIs and was exported to excel. From there it was imported into the already existing SPSS file.

To assess differences in accuracy (overall, direct, and indirect) between the four conditions, two-way ANOVAS were computed where role type and medium type served as independent variables. Next, one-way ANOVAS were used to investigate differences between medium or role type on gaze behaviour. Relations between gaze behaviour and accuracy were studied with simple and multiple regression analyses, where the time spent on areas of the face for instance served as predictor variables and direct accuracy as the dependent variable. Opting for direct accuracy as dependent variable, instead of indirect or a combination of both, is due to that this measure assesses robust accuracy. Indirect measures might not automatically correctly measure the accuracy. Next, differences between the direct and indirect measures were assessed using a paired sample *t*-test. Differences between the direct measure and the five subtypes of indirect measures were tested with a one-way repeated measure ANOVA. Finally, the influence of relational and behavioural familiarity on accuracy was examined using multiple regression. For all analyses, a significance level of .05 was used.

Results

Assumptions

Residual analysis was performed to test for the assumptions of the two-way ANOVA for each dependent variable. Outliers were assessed by inspection of a boxplot, normality was assessed using Shapiro-Wilk's normality test and homogeneity of variances was assessed by Levene's test. Identified outliers (1.5 box-lengths from the edge of the box) were kept into the analysis since its inclusion did not induce a big change in the results. Further, residuals for all dependent variables were normally distributed and there was homogeneity of variances.

The assumption of homogeneity of variances for the relative time spent on the entire target was violated; there was heterogeneity of variances, as assessed by Levene's test for

equality of variances (p = .021). Therefore, instead of a one-way ANOVA was a Welch's ANOVA computed.

Regression assumptions were checked by linearity by partial regression plots and a plot of studentized residuals against the predicted values. Independence of residuals was assessed by a Durbin-Watson statistic which should be close to 2. Homoscedasticity was assessed by visual inspection of a plot of studentized residuals versus unstandardized predicted values. Multicollinearity was assessed by tolerance values greater than 0.1. It was checked whether there were studentized deleted residuals greater than ±3 standard deviations, and whether there were leverage values greater than 0.2, and values for Cook's distance above 1. The assumption of normality was assessed by a Q-Q Plot. Assumptions for regression analyses were met.

Detection accuracy rates across conditions

The overall accuracy (undifferentiated between types of accuracy) scores ranged from 28.33% to 63.33% (regardless of condition). The overall direct assessed accuracy ranged from 20.00% to 80.00%. The overall indirect assessed accuracy ranged from 26.00% to 62.00%. Table 2 depicts means and standard deviations per condition for the overall, direct, and indirect veracity assessed types. Indirect accuracy was also measured with distinct categories. Table 3 shows the variation in accuracy rates per sub-types of the indirect measure.

Table 2

Detection accuracy per condition and per type of accuracy in percentages

	Overall	**Direct**	**Indirect**
	M (*SD)*	*M* (*SD*)	*M* (*SD*)
Video/Receiver	.48 (.07)	.47 (.15)	.44 (.06)
Video/Observer	.47 (.07)	.46 (.14)	.47 (.07)
F2F/Receiver	.46 (.09)	.44 (.12)	.46 (.09)
F2F/Observer	.45 (.09)	.43 (.15)	.45 (.09)
Total	.45 (.08)	.45 (.14)	.46 (.08)

Table 3

Types of indirect accuracy per condition in percentages

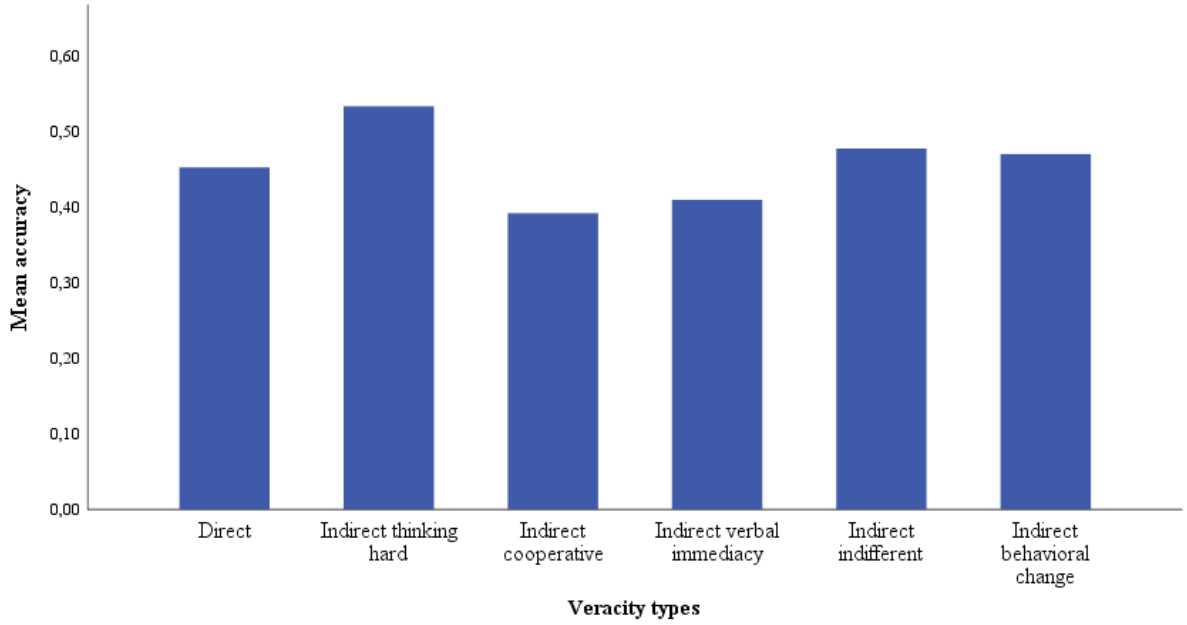
	Indirect Thinking Hard	**Indirect Cooperative**	**Indirect Verbal Immediacy**	**Indirect Indifferent**	**Indirect Behavioural Change**
	M (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)
Video/Receiver	.53 (.12)	.35 (.14)	.40 (.13)	.45 (.18)	.47 (.14)
Video/Observer	.55 (.14)	.38 (.14)	.44 (.15)	.49 (.15)	.49 (.16)
F2F/Receiver	.51 (.17)	.44 (.14)	.40 (.11)	.51 (.10)	.44 (.20)
F2F/Observer	.54 (.13)	.40 (.13)	.39 (.13)	.46 (.14)	.47 (.20)
Total	.53 (.14)	.39 (.14)	.41 (.13)	.48 (.14)	.47 (.17)

Indirect and direct measures

To test the hypothesis that indirect measures result in higher accuracy rates than direct measures (H1) a paired sample *t*-test was conducted. Veracity judgement served as a dependent variable whereas the two categorical related pairs (direct and indirect) served as the independent variable. Veracity judgement accuracy was found to be slightly higher using indirect measures (M = 45.56%, SD = 7.83%) in comparison to direct measures (M = 45.15%, SD = 14.08%). It led to an increase of 0.41%. However, there was no statistically significant difference found between the two types of veracity assessment, $t_{(67)}$ = -.257, p = 0.399. In other words, participants did not have significantly higher scores when deception detection was assessed indirectly than if it was assessed directly. Therefore, the first hypothesis is rejected. However, when differentiating between types of indirect measures, a significant difference was found. According to the repeated-measures ANOVA with a Greenhouse-Geisser correction a statistically significant difference between the direct veracity judgement accuracy and one of the different sub-types of indirect veracity judgement accuracy was found, F(4.424, 296.391) = 9.720, p <.01, *partial* η^2 = .127 (see Figure 1). Post hoc tests using the Bonferroni correction revealed that the mean veracity judgement accuracy for the indirect measure thinking hard led to an increase in the mean to 53.24% compared to an accuracy on the direct measure of 45.15%, which was significantly different (p = .011). Cooperative showed a trend towards statistical significance, p = .075. However, the effect was in the opposite direction than expected. Namely, accuracy assessed on cooperative was 39.12% whereas direct assessed accuracy was 45.15%. The other indirect assessed accuracy rates, including, verbal immediacy, indifference and behavioural change did not show a statistically significant difference in comparison to the direct veracity judgement measure, *ps* > .568.

Figure 1

Accuracy measures with corresponding accuracy rates



Role type and medium type on veracity judgement accuracy

Next, it was tested whether presentation modalities (i.e. conditions) had an influence on veracity judgement accuracy. It was hypothesized that observing in combination with the F2F medium led to the highest accuracy rates (H6). Therefore, a two-way analysis of variance (ANOVA) was conducted, where role type and medium type served as independent variables and overall veracity judgement as dependent variable. There was no statistically significant interaction effect found between the medium type and role type on the overall veracity judgement accuracy, $F(1,64) = .551$, $p = .461$, *partial* $\eta^2 = .009$. This means that the effect of the medium type on accuracy is not influenced by the role type or, equivalently, the effect of the role type is not influenced by the medium type. Therefore, H6 is rejected. Next it was hypothesized that observing leads to higher veracity judgement accuracy rates than receiving (H4). There was no main effect of the role type on the overall veracity judgement accuracy, $F(1, 64) = 0.079$, $p = .780$, *partial* $\eta^2 = .001$. This is inconsistent with the hypothesis and therefore it is rejected. It was also hypothesized that the F2F medium led to higher veracity judgement accuracy than the video call medium (H5). There was no statistically significant main effect of the medium type on the overall veracity judgement accuracy, $F(1, 64) = 0.082$, $p = .775$, *partial* $\eta^2 = .001$. Therefore, H5 is rejected.

Again, a two-way ANOVA was computed where role type and medium type were independent variables and direct and indirect (separately) were dependent variables. When

differentiating between the veracity judgement accuracy types, direct and indirect, again no significant interaction effects were found between role type and medium type, $F(1,64) = .001$, $p = .977$, *partial* $\eta^2 = < .001$ and $F(1,64) = .759$, $p = .387$, *partial* $\eta^2 = .012$, respectively. No statistically significant main effect of the medium type was found for direct, $F(1,64) = .852$, $p = .359$, *partial* $\eta^2 = < .013$ and indirect veracity judgement accuracy, $F(1,64) = < .001$, $p = .993$, *partial* $\eta 2 = < .001$. Also, no statistically significant main effect of the role type was found for direct, $F(1,64) = .096$, $p = .758$, *partial* $\eta^2 = .001$ and indirect veracity judgement accuracy $F(1,64) = .198$, $p = .658$, *partial* $\eta^2 = .003$.

Role type and medium type on gaze behaviour

To get an overview, Table 4 depicts means and standard deviations of the relative time that was spent on distinct areas of the target, per condition.

Table 4

Relative time spent on AOIs per condition in percentages

		Entire Lisa	**Entire face**	**Eyes and brows**	**Mouth**	**Arms and hands**	**Feet and legs**
		M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)
Video	Receiver	.44 (.10)	.40 (.12)	.14 (.15)	.09 (.07)	-	-
	Observer	.59 (.16)	.54 (.19)	.15 (.20)	.18 (.16)	-	-
F2F	Receiver	.38 (.14)	.25 (.16)	.05 (.05)	.07 (.06)	.00 (.00)	.01 (.01)
	Observer	.39 (.16)	.20 (.15)	.04 (.06)	.05 (.06)	.01 (.01)	.02 (.02)

It was earlier hypothesized that observers would spent more time on the target in comparison to receivers (H7). To test this a one-way Welch's ANOVA was computed. The relative time spent on the entire target was longer for the observers ($M = 48.49\%$, $SD = 18.95\%$) than for the receivers ($M = 40.97\%$, $SD = 12.14\%$), however the difference between the two groups was not statistically significant, Welch's $F(1,66) = 3.836$, $p = .054$, *partial* $\eta^2 = .055$. Therefore, H7 is rejected.

Additionally, a one-way ANOVA was computed to investigate whether observing or communicating via a video results in spending more time on the face of the target in comparison to observing or communicating F2F (H8). The relative time spent on the face of the target was longer for participants who communicated/observed via video ($M = 46.96\%$, $SD = 17.03\%$) than for the participants who communicated/observed F2F ($M = 22.20\%$, $SD = 15.54\%$). According to the one-way ANOVA, the relative time spent on the face was statistically significantly different between the two groups, $F(1,66) = 39.186$, $p = < .001$, *partial* $\eta^2 = .373$. Thus, H8 is accepted.

Post hoc analyses

There were no additional hypotheses formulated regarding differences in gaze behaviour for the distinct conditions. However, one additional analysis was computed in response to the result above. The results showed that participants in the F2F condition spent significantly less time on the face. One might wonder if this remaining time is spent on the rest of the body instead, since F2F had this option, contrary to the video. Therefore, differences in the relative time spent on the entire target are examined for participants in the video condition and F2F condition. The relative time spent on the entire target for the F2F condition was 38.29% (*SD* = 14.76%). Participants in the video condition, on the other hand, spent of the total time of the interview 50.95% (*SD* = 15.12%) on the entire target. This difference was statistically significant $F(1,66) = 12.203$, $p = .001$, *partial* $\eta^2 = .010$.

Relation gaze behaviour and accuracy

A simple linear regression was run to determine whether a positive relation existed between time spent on the hands and/or arms and the veracity judgement accuracy (H9). According to a simple linear regression model, the time spent on the hands and/or arms did significantly predict the accuracy rates, $F(1, 22) = 4.435$, $p = .047$, adj. $R^2 = .130$. Against initial assumptions, the relative time spent on the hands and/or arms were found to have significantly negative effects on accuracy rates, $B = -4.742$, $SE = 2.252$, $p = .047$. Even though the results are significant, the effect was negative instead of positive. Therefore, H9 is rejected.

In addition, more effects of gaze behaviour on direct accuracy were researched. In order to examine the effects of *the time which is spent* on distinct areas of the face on the (direct) veracity judgement accuracy, a multiple regression analysis was computed. Predictors of the veracity judgement accuracy were the relative time spent on the forehead, eyes and brows, nose area, mouth area, and chin. The multiple regression model did not statistically significantly predicted accuracy, $F(5,62) = 0.891$, $p = .492$, adj $R^2 = -.008$. However, only the time spent on the forehead was found to have significant effects on accuracy, $B = -1.555$, $SE = .770$, $p = .048$. The time spent on eyes and brows ($B = .045$, $SE = .144$, $p = .757$), nose ($B = -.174$, $SE = .331$, $p = .601$), mouth ($B = .033$, $SE = .194$, $p = .866$), and chin ($B = -.090$, $SE = .448$, $p = .841$) did not have a statistically significant effect on accuracy. Thus, the time which was spent on the distinct areas of the face did not affected the degree of accuracy, except for the time spent on the forehead. An increase in the time that is spent on the forehead is associated with a decrease in accuracy within this set of variables.

A similar procedure was used for examining the effect of *durations of fixations* (average fixation durations) on accuracy. A multiple regression analysis was run to predict accuracy from

the average fixation durations of the distinct areas being the forehead, eyes and brows, nose area, mouth area, and chin. The model did not statistically significantly predicted veracity judgement accuracy, $F(5, 62) = .892$, $p = .492$, adj $R^2 = -.008$. None of the predictors had a significant effect on accuracy, $p = > .599$. However, the predictor fixation duration forehead showed a trend, $p = .080$. See Appendix J for an comprehensive overview of slopes, standard errors and p-values.

Also, the effects of *fixation counts* on the AOIs, forehead, eyes and brows, nose area, mouth area, and chin were investigated. A multiple regression was run to predict accuracy from fixation counts. The model did not statistically significantly predict accuracy, $F(5,62) = 0.786$, p = .564, adj $R^2 = -.016$. None of the separate predictors (i.e. the distinct AOIs with the correspondent counts) had an effect on accuracy, $p = >.201$ (see Appendix J).

The number of times that the eyes went back to one of the AOIs (forehead, eyes and brows, nose area, mouth area, and chin) (i.e. visit counts) and its relation to accuracy in deception detection was also studied. The overall model did not statistically significantly predict accuracy, $F(5,62) = 1.319$, $p = .268$, $R^2 = .023$. However, the visit counts of the mouth were found to have a significant effect on accuracy, $B = .002$, $SE = .001$, $p = .047$. This shows that the more often the eyes switched back and forth to the mouth area the higher the accuracy. The other predictors did not have a statistically significant effect on accuracy, $p = >.103$ (see Appendix J).

Moreover, a simple linear regression was computed to investigate whether the relative time spent on the face in general (undifferentiated between areas) is related to accuracy. There was no effect of the time which is spent on the entire face of the target on accuracy, $F(1,66) = 0.005$, $p = .944$, adj $R^2 = -.015$. Thus, the time spent on the actors' face was not associated with accuracy.

Lastly, a multiple regression was run including only the people in the F2F medium who spent time looking at the target's arms/hands and or feet/legs. A combination of the predictors; relative time spent on the face, hands and/or arms, and feet and/or legs, and their influence on accuracy was tested. The model statistically significantly predicted accuracy, $F(3,16) = 5.303$, $p = .010$, adj $R^2 = .405$. Relative time spent on the hands/arms ($p = .009$) and feet/legs ($p = .004$) added statistically significantly to the prediction. Time spent on the face did not add to the prediction ($p = .641$). Regression coefficients and standard deviations can be found in Table 5 (below).

Table 5

Multiple regression results for veracity judgement accuracy for the relative time spent on areas of the body

Accuracy	*B*	*95 % CI for B*		*SE B*	*β*	R^2	ΔR^2
		LL	*UL*				
Model						.499	.405*
Constant	.443***	.33	.56	.06			
Face	-.081	-.44	.28	.17	.63		
Hands/Arms	-5.921**	-10.14	-1.71	1.99	-.61**		
Feet/Legs	3.991**	1.44	6.54	1.20	.63**		

Note. Model = 'Enter' method in SPSS Statistics; *B* = unstandardized regression coefficient; CI = confidence interval; *LL* = lower limit; *UL* = upper limit; *SE B* = standard error of the coefficient; β = standardized coefficient; R^2 = coefficient of determination; ΔR^2 = adjusted R^2. *$p < .05$. ** $p < .01$. *** $p < .001$.

Relational and behavioural familiarity on veracity judgement accuracy

Overall results indicate that the participants scored average on behavioural familiarity (M = 3.50, SD = 0.62), knowing that this construct was measured on a 7-point Likert-scale. Additionally, seven participants indicated to know the actor, whereas 61 did not know the actor.

A simple regression was computed to test H2 which stated that behavioural familiarity would increase accuracy. The regression model did not statistically significantly predicted accuracy rates $F(1, 66) = 0.531$, $p = .469$, adj. $R^2 = -.007$. Thus, behavioural familiarity did not improve nor impair accuracy and therefore is H2 rejected.

A similar procedure was followed to examine whether relational familiarity impaired accuracy (H3). The regression model did not statistically significantly predicted accuracy rates $F(1, 66) = 1.935$, $p = .169$, adj. $R^2 = .014$. In other words, relational familiarity had no influence on accuracy. Thus, H3 is also rejected.

Discussion

This research investigated how deception detection accuracy could be increased by studying the receiver instead of the sender. Put differently, it was explored which factors explained the variance in veracity judgement accuracy.

The overall accuracy rates, regardless of the manipulations, were consistent with the literature which repeatedly shows that people, in general, are not much better at detecting lies than chance level (Bond & DePaulo, 2006; Hartwig et al., 2004). However, when looking at the accuracy rates corresponding to the conditions, we failed to find interaction and main effects of the medium type and role type on veracity judgement accuracy. In other words, communicating or observing via a video versus F2F did neither improve nor impair accuracy rates (H5). The same applies to the role type; observing or leading the interview did not influence the accuracy rates (H4). Neither a combination of the two factors led to an increase in accuracy (H6).

There might be some explanations for these findings. First, an explanation for that F2F interaction did not yield higher accuracy rates than video interaction might be caused by us being more trained in video interaction. Recent articles try to explain the so-called 'Zoom Fatigue', due to the actuality around Covid-19 (e.g. Bailenson, 2021). It is reasoned that cognitive load in videoconferencing calls is substantially increased. In such a way, that nonverbal behaviour in sending – extra cues – and receiving relies much more on consciously processing. Consciously processing equals a higher cognitive load which creates the fatigue. Whereas in F2F interaction the complexity of nonverbal behaviour is processed unconsciously –i.e. effortless (Bailenson, 2021; Lee, 2020). One would logically expect that this increase in load would impair a secondary task, in this case, a deception detection task. However, after a full year of interaction via video calls, people might have become better at reading nonverbal behaviours via a video, and thus no difference in performance on accuracy between the two medium types occurred. People anno 2021 might developed manners in how to behave nonverbally in order to be well understood by trial and error during video interaction (e.g. an incident where speech content was misinterpreted occurred, so it is learned that emotions or expressions should be exaggerated in order to be well understood) (Lee & Wagner, 2002). Nevertheless, it is questionable if this can be ascribed to previous year of intensive computer-mediated interaction. This feeling for video interaction presumably started earlier, knowing that for instance MSN messenger was introduced in 1999 (Fang & Tang, 2017). Most participants in the sample were born in the 90's or later and so grew up with this medium type. The video medium might have interfered with the deception detection task for elderly and not for younger aged people.

Second, neither differences occurred between the type of roles. It is previously reasoned that the advantage of the observer is a lower cognitive load, due to the sole task of making veracity judgements. Besides that, the receiver needs to read and ask questions. Therefore,

veracity judgements would be impaired. However, it might be more nuanced; the role type observer has a disadvantage that might cancel out the mentioned advantage. That is for instance, observers are not in charge of the pace of the interview, in contrast to receivers. Therefore, load of the observers might increase while keeping up with the speed of the questions asked. In sum, the advantages of one role come with disadvantages. And the disadvantages of the other comes with advantages. These factors might cancel each other out, hence role type poses no effect.

There were differences in gaze behaviour for the distinct medium types; participants in the video condition gazed more at the face than in the F2F condition (H8). In a way, this makes sense, since in the F2F conditions people had also the opportunity to look at the body. But an additional analysis showed that this was not the case; of the total time looked participants in the video conditions significantly more at the entire target than people in the F2F conditions. Thus, this indicates that participants in the F2F conditions looked more often at other things in the scenery than participants in the video conditions. One might say, that in the F2F medium, people were more distracted by their environment than people using the video as a medium. One possible explanation for this occurrence is that in the video condition the eye gaze is at a close distance to the screen, making it harder to look around the screen. When communicating or observing F2F one is usually not at a 50-80 cm distance from the face (this is the case when communicating via a video call). Normally, this is only reserved for intimate relations (Meisels & Guardo, 1969). Let alone due to corona this F2F distance was even more increased, leaving thus more space to focus the gaze on peripheral objects.

In sum, the conditions, including video-receiver, video-observer, F2F-receiver, and F2F-observer, did not prove to elicited differences in accuracy. The conditions however did elicit differences in gaze behaviour. Combining these two findings seem to indicate that accuracy is not related to what has been perceived. Because differences in gaze (e.g. looking longer at the face in the video conditions) did not automatically translate to higher or lower accuracy rates. However, this analysis is not comprehensive, as such that only differences in gaze between conditions are studied. These might not necessarily yield a difference in accuracy, but another gaze behaviour would. Therefore, it was examined whether there was a relation between gaze behaviour and accuracy, regardless of the conditions.

Conclusions regarding relations between gaze behaviour and accuracy can be divided into two parts. First, conclusions regarding gaze behaviour on the face and accuracy will be discussed. Overall, looking at the face did not seem to have a relation with accuracy. In other words, looking at the face for longer or shorter periods of time did not impair nor improve accuracy rates. This is consistent with literature. Virtually most of the reviewed nonverbal cues

found on the face seem to be invalid (e.g. gaze aversion, smiles, head nodding, etc.) (DePaulo et al., 2003; Vrij et al., 2019). Another reason could be that if there are valid cues present in the face, they might be barely perceptible by the human eye (e.g. micro expressions).

Apart from studying the time spent on the entire face in relation to accuracy, there was also differentiated between areas of the face. Taking these factors (forehead, eyes and brows, nose, mouth, and chin) altogether a significant result occurred for one of the predictors. The relative time spent on the forehead was negatively associated with accuracy. Put differently, the more time was spent on the forehead, the less accurate participants became. This is different from earlier discussed findings that muscles in the upper face are less controllable and therefore leak more of the genuine emotion (Warren et al., 2009; Porter et al., 2012). Based on these papers one would expect higher accuracy rates by looking at the forehead instead. This negative association might also be due to wrong interpretations, or due to the shortness of muscle activity. It might also mean that cues are simply absent in the forehead.

Additionally, fixation durations connected to the distinct areas in the face did not predict accuracy. Neither did fixation counts. Visit counts, on the other hand, showed for the mouth area a significant effect, meaning that the more often the eyes switched back and forth to the mouth area the higher the accuracy. The literature does not show a theory that could explain the underlying factors that caused this finding. It would be interesting to see if additional research could replicate this finding and accordingly theorize about the relation to deception detection.

Second, next to the face, gaze behaviour of the body was also studied (in combination with the face) in relation to accuracy. An interesting finding occurred when including the lower body, upper body, and face in the multiple regression analysis. Time spent on the legs and feet positively predicted accuracy rates. Time spent on the hands and arms impaired accuracy (H9). The face again was within these predictors unrelated to accuracy. Theories regarding nonverbal behaviour in deception detection indeed have claimed that people were more accurate when looking at the lower part of the body (Ekman, 2003). The results of this study are in line with this finding. At the same time, it is inconsistent with other papers; feet and leg movements seem unrelated to accuracy, whereas fidgeting seemed positively related, and arms and hand movements negatively related to accuracy (DePaulo et al., 2003; Bond et al., 2015). It thus is inconclusive that paying attention to the feet and legs is related to accuracy. It is also remarkable that paying attention to the hands and arms impaired accuracy since participants had overall valid beliefs regarding those cues in relation to deception.

This research also studied the difference between the type of assessment and whether indirect measures would yield higher accuracy rates (H1). There was no difference in accuracy

between the direct and indirect measures. However, there was a significant difference between the direct measure and one subtype of the indirect measure, being thinking hard. So, participants became more accurate when asking *'did the suspect had to think hard when…*, instead of, *'did the suspect lie about…'*. Note, however, that despite the higher accuracy it still was not far above chance level. The not statistically significantly increase in accuracy might be because either the indirect constructs (i.e. cues) were not sufficiently operationalized, or that these indirect measures do not rely on valid cues to deception. However, regarding the latter, Bond et al. (2015) found that these four indirect measures did outperform direct measures (see also DePaulo et al., 2003). At the same time, they mentioned that these measures are not well studied. This study made attempts to examine performance in accuracy on both measures. However, a lack of theorization of the distinct concepts may have resulted in invalid items. Lastly, the hypotheses did not include truth and lie accuracy but we have indications that performance on direct and indirect measures might be different when dividing accuracy into truth and lie accuracy (see Appendix I).

Lastly of all the findings, behavioural and relational familiarity did not significantly seem to improve nor to impair accuracy (H2 and H3). Regarding the behavioural familiarity, it might have been the case that in hindsight participants knew what valid cues were when seeing them in the questions, but previously during the interview, no attention was paid to them. This would explain the non-existence of a positive relation. Relational familiarity was neither related to accuracy. This might be due to that only six people of the entire sample seem to have indicated to know the actor and of these people, they might not have been very closely related.

Limitations

To set the results of this study in perspective, it is important to also look at some limitations of the study. First, a condition for finding variation in accuracy scores is that cues in the first place should have been present. Despite attempts, the situation might not have been realistic enough to elicit pressure for the actor to be as perceived as honest as possible. Therefore, cues might not have been pronounced enough and thus there were no differences (DePaulo, 2003). Despite this potential limitation, accuracy rates were in line with the rates which are repeatedly found in the literature .

A second limitation of this study was that the interview was a role-play which was repeated up to at least 74 times. This repetition has likely resulted in a solid story of the actor who came into a certain cadence. This might have led to the occurrence of fewer cues in the end, compared to the beginning. However, it is likely to assume that in real life, guilty suspects

also prepare and practice a credible story. Yet, probably not up to 74 times and virtually the same every iteration.

Moreover, the features of the actor might have influenced the results. Deception detection is dependent on two aspects; the ability of the liar to deceive and the ability of the detector to detect (Bond & DePaulo, 2008). The actor had a history in theatre, and this could have resulted in her having higher abilities to deceive, compared to an average person. Also, the appearance of the actor could have led to a truth-bias. Papers exist which research the relation between attractiveness and trustworthiness for example (Jaeger et al., 2018).

Also, deception detection is a challenging task on its own. What it might have made it more challenging for the participants is making some claims truthful and others deceitful within one story. An improvement for following studies would be to either consistently lie or tell the truth. This way, if systematic cues exist, cues would be more likely to be detected if they are consistent within one interview. The researcher namely repeatedly got back from the participants that it was an incredibly hard task. However, a disadvantage when changing this aspect is when people are unsure about their classification of either truthful or deceitful, one has a chance of 50% in succeeding. This is eliminated using the current design.

A third limitation that should be mentioned is the fact that participants might not have remembered the exact behaviour or situation to which the questions in the questionnaire referred to. This could have affected the reliability of the indirect items. A veracity judgement had to be made for each question asked in the interview. The interview consisted of 12 questions. We anticipated that it is quite hard to remember every initial judgement after a response from the actor, by giving participants a scratch paper with the preprinted questions (see Appendix F). After each question the participant could circle around their immediate judgement, being either a classification of *'truth'* or *'lie'*. Later in the questionnaire, they could look at their initial thoughts or judgements regarding certain statements. However, the questionnaire asked many other – indirect – questions, for instance, *'did the suspect had to think hard when answering this question'*. These were not preprinted on the paper. We often got back from the participants, that they did not remember the detailed behaviours that might have been triggered after a question. This could have impacted the results in such a way that if behaviours that needed to be assessed are not remembered that they are not measured valid. A recommendation for this matter will be mentioned later.

Lastly, the quality (i.e. validity) of the questionnaire could have influenced the results. Namely, if the items do not measure the construct correctly, it naturally becomes impossible to verify the higher performance of these four indirect measures. It was out of the scope of this

paper to assess the validity of the separate scales or constructs. Yet, this is an important step that must be followed before assessing the performance of indirect measures. Nevertheless, items were inspired by previous research.

It will always remain a challenge to study deception detection. But despite these limitations there were also some strengths of the design. This study attempted to make the case as realistic as possible by providing detailed cases in which the participant and actor could empathize with. The topic was also deliberately chosen (witness/accomplice in car accident) in order to increase the stakes and let the actor feel as much distressed to the extent possible. A less serious case might not evoke any sense of distress when imaging oneself in a role when it is about a less serious crime or accusation (e.g. cheating on an exam). Also, aspects of the case were visualized with the underlying reason that contradicting pictures is more concrete then contradicting a written story (see Appendix E). This might have had an effect that potential cues to deceit would be more present in this experimental setting. In response to these limitations follow recommendations regarding future research.

Future research

This study showed that looking at the face is unrelated to accuracy. Thus, one can conclude that either systematic nonverbal cues to deception in the face do simply not exist or it could mean that receivers do not recognize nonverbal cues to deception *(weak objective cue hypothesis)*. This research is unable to determine what the most likely explanation for this occurring result is. However future research could come with more conclusive results by letting a computer studying the face instead of relying on human perceptions. More concretely, the Imotions software can ascertain which emotions and facial expressions are linked to deception. This way a computer can be conclusive in determining whether systematic cues exist at all, and if so, it is likely to assume that humans do not detect or recognize them.

It is also recommended that future research should further explore the effects of role type and medium type, by isolating the constructs that come with the role or medium type. Meant is that for instance observing comes with less cognitive load, lower conversational involvement, etc. In this study, there was no effect, but if there was, it is hard to appraise it to the factor that led to the increase or decrease in accuracy. Also, in doing so, maybe an effect will appear and the one factor cancelling out the other is no longer an issue. Additional research can also explore the effects of video versus F2F interaction on emotion/deception detection tasks. This research can investigate whether elderly people have indeed an impaired performance in emotion or deception detection tasks compared to young people via video interaction. And if so, conclusions can be drawn whether the earlier reasoning can be substantiated that video

interaction for young people does not interfere with the communication quality, unlike for elderly.

Moreover, to eliminate the earlier mentioned potential effects of repetition and features of the actor, it is suggested that future research adopts a slightly different method. The experimental setting should make use of pairs of students. The 'interviewer' has information over the 'suspect' regarding a case. The suspect has either a truthful story or a false story. The suspect should convince the other of their innocence. The interviewer should come up with questions themselves. If they succeed, they get an extra reward. Doing an experiment this way has some benefits over the current design. First, it is more of a game where students could be more motivated to convince others of their innocence. Also, an additional reward would help a lot to motivate the suspect. Moreover, each student is questioned once, and therefore it will not become a memorized story that they will tell, but a rather more spontaneous response to the questions. Also, in this way potential influences of characteristics of the actor are ruled out. A downside however is less standardization.

Lastly, a previously mentioned limitation was that specific behaviours which needed to be assessed after each question were not remembered (e.g. *'Did the witness had to think hard?'*, or *'Did the witness seem indifferent?'*). To anticipate on this matter, one could apply a slightly different method, namely using a think-aloud procedure. During the investigative interview, the researcher should interrupt after each question and ask for instance *'indicate on a scale from one to ten how hard he/she had to think about her answer'* and notate this. This procedure should continue till the end of the interview. A slightly different option could be that participants watch their recording of the glasses back, during answering the items. This way, it becomes a more retrospective study. If valid items are used, both ways could substantially increase the reliability of the performance of indirect measures versus direct measures.

Implications

The findings of this research contributed to the existing literature as such, that it again showed that people are poor lie detectors. But more importantly, it found that looking at the arms and legs significantly predicted accuracy. This seems to support the existence of valid nonverbal cues to deception. Finding support for nonverbal cues in deception detection using an eye-tracker might stimulate the deception detection discipline to refocus on the nonverbal aspect of deception detection.

The findings might eventually also be important for the field wherein deception detection is relevant, such as the criminal investigation department of the – Dutch – police.

Practitioners have shown to hold invalid beliefs when it comes to deception detection (Denault & Dunbar, 2019; Reinhard et al., 2013). Therefore, doing research into ways to improve accuracy is of great importance. Consequently, findings of research can provide content for future training programs for officers. Eventually, this might prevent unjust suspicion or conviction for innocent suspects. This research made a step in contributing to this objective.

This study found that looking at the feet and legs improved veracity judgement accuracy. However, interactions take increasingly more place via video calls. A video call only shows the upper body, leaving the feet and legs hidden from the conversational partner. This medium thus has a potential deficiency that might impair veracity judgement accuracy. One could consider instead capturing the entire person on camera, for instance, during a job interview or an oral test, in order to let communication through this medium be as rich to the extent possible.

Conclusion

Several approaches were used to investigate ways to improve veracity judgement accuracy. In conclusion, role type and medium type did not pose an effect on veracity judgement accuracy. Neither did the indirect measures when comparing them to the direct measures. Further, relational and behavioural familiarity proved not to improve nor to impair accuracy in deception detection. The role type and medium type did elicit differences in gaze behaviour. This however did not translate to differences in accuracy and seemed to indicate that accuracy is not dependent on what has been perceived by the eye. Additional analyses indicated however differently. Relations were found between the time spent on hands and/or arms, feet and/or legs, and the forehead on accuracy. The more time spent on the hands and/or arms and forehead the lower the accuracy and the more time spent on the feet and/or legs, the higher the accuracy.

References

Altmann, U., Oertel, C., & Campbell, N. (2012). Conversational involvement and synchronous nonverbal behaviour. *In Cognitive Behavioural Systems* (pp. 343-352). Springer, Berlin, Heidelberg.

Bailenson, J. N. (2021). Nonverbal overload: A theoretical argument for the causes of Zoom fatigue. *Technology, Mind, and Behaviour, 2*(1). doi: 10.1037/tmb0000030

Best, G., Hodgeon, J., & Street, C. N. (2018). How contemporary theory informs lie detection accuracy and bias. *Crime Security and Society, 1*(2), 30-36.

Blair, J. P., Levine, T. R., & Shaw, A. S. (2010). Content in context improves deception detection accuracy. *Human Communication Research, 36*(3), 423–442. https://doi.org/10.1111/j.1468-2958.2010.01382.x

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

Bond, C. F., & DePaulo, B. M. (2008). Individual differences in judging deception: Accuracy and bias. *Psychological bulletin, 134*(4), 477-492. https://doi.org/10.1037/0033-2909.134.4.477

Bond, C. F., Levine, T. R., & Hartwig, M. (2015). New findings in nonverbal lie detection. Deception detection: Current challenges and new directions, 37-58. doi:10.1002/9781118510001

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

Burgoon, J. K., Buller, D. B., Ebesu, A. S., & Rockwell, P. (1994). Interpersonal deception: V. Accuracy in deception detection. *Communications Monographs, 61*(4), 303-325. doi:10.1080/03637759409376340

Colavita, F. B. (1974). Human sensory dominance. *Perception & Psychophysics, 16*(2), 409-412. https://doi.org/10.3758/BF03203962

Denault, V., & Dunbar, N. E. (2019). Credibility assessment and deception detection in courtrooms: Hazards and challenges for scholars and legal practitioners. In The Palgrave handbook of deceptive communication, 915-935. Palgrave Macmillan, Cham.

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), 979-995. https://doi.org/10.1037/0022-3514.70.5.979

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

Ekman, P. (2003). Darwin, deception, and facial expression. *Annals of the New York Academy of Sciences, 1000*(1), 205-221. doi:10.1196/annals.1280.010

Ekman, P., & Friesen, W. V. (1969). Nonverbal leakage and clues to deception. *Psychiatry, 32*(1), 88-106. doi: 10.1080/00332747.1969.11023575

Ekman, P., & Friesen, W. V. (1982). Felt, false, and miserable smiles. *Journal of nonverbal behaviour, 6*(4), 238-252.

Ehrlichman, H., & Micic, D. (2012). Why do people move their eyes when they think?. *Current Directions in Psychological Science, 21*(2), 96-100. https://doi.org/10.1177/0963721412436810

Fang, Y. H., & Tang, K. (2017). Involuntary migration in cyberspaces: The case of MSN messenger discontinuation. *Telematics and Informatics, 34*(1), 177-193. https://doi.org/10.1016/j.tele.2016.05.004

Gullberg, M., & Holmqvist, K. (2001). Visual attention towards gestures in face-to-face interaction vs. on screen. *In International Gesture Workshop*, 206-214. Springer, Berlin, Heidelberg.

Hart, C. L., Fillmore, D. G., & Griffith, J. D. (2009). Indirect detection of deception: Looking for change. *Current Research in Social Psychology, 14*(9), 134-142. Retrieved from: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.364.1259&rep=rep1&type=pdf

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. doi:10.1177/1098611104264748

iMotions guide. (2017). Eye tracking: the complete pocket guide.

Jaeger, B., Sleegers, W. W., Evans, A. M., Stel, M., & van Beest, I. (2019). The effects of facial attractiveness and trustworthiness in online peer-to-peer markets. *Journal of economic psychology, 75*, 102125. https://doi.org/10.1016/j.joep.2018.11.004

Kane, M. J., Brown, L. H., McVay, J. C., Silvia, P. J., Myin-Germeys, I., & Kwapil, T. R.

(2007). For whom the mind wanders, and when: An experience-sampling study of working memory and executive control in daily life. *Psychological science*, *18*(7), 614-621. https://doi.org/10.1111/j.1467-9280.2007.01948.x

Koppen, C., Levitan, C. A., and Spence, C. (2009). A signal detection study of the Colavita visual dominance effect. *Exp. Brain Res. 196*, 353–360. doi: 10.1007/s00221-009-1853-y

Lee, J. (2020). A neuropsychological exploration of Zoom fatigue. Psychiatric Times.

Lee, V., & Wagner, H. (2002). The effect of social presence on the facial and verbal expression of emotion and the interrelationships among emotion components. *Journal of Nonverbal Behavior, 26*(1), 3-25. https://doi.org/10.1023/A:1014479919684

Levine, T. R. (2015). New and improved accuracy findings in deception detection research. *Current Opinion in Psychology (6*), 1-5. doi:10.1016/j.copsyc.2015.03.003

Levine, T. R., & Bond, C. F. (2014). Direct and indirect measures of lie detection tell the same story: A reply to ten Brinke, Stimson, and Carney (2014). *Psychological science, 25*(10), 1960-1961. doi: 10.1177/0956797614536740

Levine, T. R., Park, H. S., & McCornack, S. A. (1999). Accuracy in detecting truths and lies: Documenting the "veracity effect". *Communications Monographs, 66*(2), 125-144. doi: 10.1080/03637759909376468

Mann, S., & Vrij, A. (2006). Police officers' judgements of veracity, tenseness, cognitive load and attempted behavioural control in real-life police interviews. *Psychology, Crime and Law*, *12*(3), 307–319. https://doi.org/10.1080/10683160600558444

Meisels, M., & Guardo, C. J. (1969). Development of personal space schemata. *Child Development*, 40(4), 1167-1178. https://doi.org/10.2307/1127021

Paivio, A., & Desrochers, A. (1980). A dual-coding approach to bilingual memory. *Canadian Journal of Psychology/Revue canadienne de psychologie, 34*(4), 388-399. https://doi.org/10.1037/h0081101

Phillips, L. H., Channon, S., Tunstall, M., Hedenstrom, A., & Lyons, K. (2008). The role of working memory in decoding emotions. *Emotion, 8*(2), 184–191. https://doi.org/10.1037/1528-3542.8.2.184

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 Behaviour, 36*(1), 23-37. doi:10.1007/s10919-011-0120-7

Proudfoot, J. G., Jenkins, J. L., Burgoon, J. K., & Nunamaker Jr, J. F. (2016). More than

meets the eye: How oculometric behaviors evolve over the course of automated deception detection interactions. *Journal of Management Information Systems, 33*(2), 332-360. https://doi.org/10.1080/07421222.2016.1205929

Reinhard, M. A. (2010). Need for cognition and the process of lie detection. *Journal of Experimental Social Psychology, 46*(6), 961-971. https://doi.org/10.1016/j.jesp.2010.06.002

Reinhard, M. A., Scharmach, M., & Müller, P. (2013). It's not what you are, it's what you know: Experience, beliefs, and the detection of deception in employment interviews. *Journal of Applied Social Psychology, 43*(3), 467-479. https://doi.org/10.1111/j.1559-1816.2013.01011.x

Reinhard, M. A., Sporer, S. L., & Scharmach, M. (2013). Perceived familiarity with a judgmental situation improves lie detection ability. *Swiss Journal of Psychology, 72*(1), 53-61. doi:10.1024/1421-0185/a000098

Reinhard, M. A., Sporer, S. L., Scharmach, M., & Marksteiner, T. (2011). Listening, not watching: Situational familiarity and the ability to detect deception. *Journal of Personality and Social Psychology, 101*(3), 467-484. https://doi.org/10.1037/a0023726

Salvucci, D. D., & Taatgen, N. A. (2008). Threaded Cognition: An Integrated Theory of Concurrent Multitasking Dario. *Psychological Review, 115*(1), 101–130. https://doi.org/10.1037/0033-295X.115.1.101

Sinnett, S., Spence, C., & Soto-Faraco, S. (2007). Visual dominance and attention: The Colavita effect revisited. *Perception & Psychophysics, 69*(5), 673-686. https://doi.org/10.3758/BF03193770

Sporer, S. L., & Schwandt, B. (2007). Moderators of nonverbal indicators of deception: A meta-analytic synthesis. *Psychology, Public Policy, and Law, 13*(1), 1-34. doi: 10.1037/1076-8971.13.1.1

Stel, M., & van Dijk, E. (2018). When do we see that others misrepresent how they feel? detecting deception from emotional faces with direct and indirect measures. *Social influence, 13*(3), 137-149. doi:10.1080/15534510.2018.1473290

Stel, M., Van Dijk, E., & Olivier, E. (2009). You want to know the truth? Then don't mimic!. *Psychological science, 20*(6), 693-699. https://doi.org/10.1111/j.1467-9280.2009.02350.x

Stiff, J. B., Kim, H. J., & Ramesh, C. N. (1992). Truth biases and aroused suspicion in

relational deception. *Communication Research, 19*(3), 326-345. https://doi.org/10.1177/009365092019003002

Street, C. N., & Richardson, D. C. (2015). The focal account: Indirect lie detection need not access unconscious, implicit knowledge. Journal of Experimental Psychology: Applied, 21(4), 342-355. Retrieved from: http://eprints.hud.ac.uk/id/eprint/25710/1/StreetRichardson_Accepted_FocalAccount_Suppl.pdf

Street, C. N., & Vadillo, M. A. (2016). Can the unconscious boost lie-detection accuracy?. Current Directions in Psychological Science, 25(4), 246-250. doi:10.1177/0963721416656348

Tobiipro.com. (2021). *Understanding Pro Lab's eye tracking metrics.* Retrieved from: https://www.tobiipro.com/learn-and-support/learn/steps-in-an-eye-tracking-study/data/understanding-Pro-Labs-eye-tracking-metrics/

Van Swol, L. M., Malhotra, D., & Braun, M. T. (2012). Deception and its detection: Effects of monetary incentives and personal relationship history. *Communication Research, 39*(2), 217-238. https://doi.org/10.1177/0093650210396868

Vrij, A. (2019). Deception and truth detection when analyzing nonverbal and verbal cues. *Applied Cognitive Psychology*, *33*(2), 160-167. https://doi.org/10.1002/acp.3457

Vrij, A., Akehurst, L., Soukara, S., & Bull, R. (2004). Detecting deceit via analyses of verbal and nonverbal behaviour in children and adults. *Human communication research, 30*(1), 8-41. doi:10.1111/j.1468-2958.2004.tb00723.x

Vrij, A., Hartwig, M., & Granhag, P. A. (2019). Reading lies: Nonverbal communication and deception. *Annual review of psychology,* 70, 295-317. https://doi.org/10.1146/annurev-psych-010418-103135

Vrij, A., & Mann, S. (2001). Telling and detecting lies in a high-stake situation: the case of a convicted murderer. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *15*(2), 187-203. https://doi.org/10.1002/1099-0720(200103/04)15:2<187::AID-ACP696>3.0.CO;2-A

Wang, J. T. Y., Spezio, M., & Camerer, C. F. (2010). Pinocchio's pupil: using eyetracking and pupil dilation to understand truth telling and deception in sender-receiver games. *American Economic Review, 100*(3), 984-1007. doi:10.1257/aer.100.3.984

Warren, G., Schertler, E., & Bull, P. (2009). Detecting deception from emotional and

unemotional cues. Journal of Nonverbal Behaviour, 33(1), 59-69. doi: 10.1007/s10919-008-0057-7

Zhou, L., Zhang, D., & Sung, Y. W. (2013). The effects of group factors on deception detection performance. *Small group research, 44*(3), 272-297. https://doi.org/10.1177/1046496413484178

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

Appendices

Appendix A: Informed consent

Thank you for taking part in this experiment. Before the experiment will start you are asked to read the following carefully.

The data that will be collected during this experiment will be used for my master thesis project and following projects which will study deception detection at the University of Twente. This study is interested in the general ability of deception detection among people. Participation in this study is on a voluntary basis. You have the right to withdraw from this study at any point without giving an explanation or justification. In addition, during this study, your anonymity will be ensured. The data will be processed anonymously and cannot be linked to your identity. If you have any questions you can ask them right now or email j.m.schimmel@student.utwente.nl

Lastly, if you have any complaints about this experiment you can email to ethicscommittee-bms@utwente.nl

I participate in this study on a voluntary basis.

- Yes

I am aware that I can withdraw from this study at any point without explaining.

- Yes

I agree that my anonymous data will be used for this master thesis project and following projects which will study deception detection.

- Yes

Appendix B: Scenarios and instructions per condition for participants

1. Video/Receiver

The case

You work as a police officer for the criminal investigation department (recherche). You have been put on a case where a girl (16) is hit by a car and the suspect drove off. As a tactical criminal investigator, it is your job to collect as much information about a certain crime as possible. Therefore, you question victims and witnesses to get a better understanding of what exactly happened. You also search for camera footage of the area wherein a criminal act took place.

You are at the very beginning of this investigation. At this point, you don't have many leads. The car accident happened on a deserted country road, the Lappenweg, outside of Enschede, on Friday night around midnight, with no cameras any close. You can't question residents in the area since within a radius of 2 km there are no houses. The victim herself was unconscious and did not see the car coming, since she was hit from the back. So, the only starting point for you now is questioning the witness who called 112.

Therefore, today you will have an investigative interview with Lisa who says she witnessed the car accident. Unfortunately, the witness is tested positive on the PCR test and must be quarantined. You cannot wait until the witness is out of quarantine. Therefore, you decided to interview the witness via a video call.

Instructions

It is your task to have the interview with Lisa, by asking the questions the researcher will give to you. Also, you have to classify the statements of Lisa as either truthful or deceitful (later in the questionnaire). During the interview, you can circle 'truth' or 'lie', after each question on a scratch paper. Try to make, after each question, an estimation of whether Lisa answers with a lie or with the truth.

Let the researcher know you've finished reading. She will now reach you the questions you need to ask during the interview.

2. Video/Observer

The case

You work as a police officer for the criminal investigation department (recherche). You have been put on a case where a girl (16) is hit by a car and the suspect drove off. As a tactical criminal investigator, it is your job to collect as much information about a certain crime as possible. Therefore, you question victims and witnesses to get a better understanding of what exactly happened. You also search for camera footage of the area wherein a criminal act took place.

You are at the very beginning of this investigation. At this point, you don’t have many leads. The car accident happened on a deserted country road, the Lappenweg, outside of Enschede, on Friday night around midnight, with no cameras any close. You can't question residents in the area since within a radius of 2 km there are no houses. The victim herself was unconscious and did not see the car coming, since she was hit from the back. So, the only starting point now is questioning the witness who called 112.

Therefore, a colleague out of your team had an investigative interview with Lisa who says she witnessed the car accident. The interview went via a video call and is recorded. You are about to watch the interview.

Instructions

It is your task to observe the interview with Lisa. While doing so, you have to classify the statements of Lisa as either truthful or deceitful (later in the questionnaire). During the interview, you can circle ‘truth’ or ‘lie’, after each question on a scratch paper. Try to make, after each question, an estimation of whether Lisa answers with a lie or with the truth.

Let the researcher know you've finished reading.

3. Face-to-Face/Receiver

The case

You work as a police officer for the criminal investigation department (recherche). You have been put on a case where a girl (16) is hit by a car and the suspect drove off. As a tactical criminal investigator, it is your job to collect as much information about a certain crime as possible. Therefore, you question victims and witnesses to get a better understanding of what exactly happened. You also search for camera footage of the area wherein a criminal act took place.

You are at the very beginning of this investigation. At this point, you don't have many leads. The car accident happened on a deserted country road, the Lappenweg, outside of Enschede, on Friday night around midnight, with no cameras any close. You can't question residents in the area since within a radius of 2 km there are no houses. The victim herself was unconscious and did not see the car coming, since she was hit from the back. So, the only starting point for you now is questioning the witness who called 112.

Therefore, today you have an investigative interview with Lisa who says she witnessed the car accident.

Instructions

It is your task to have the interview with Lisa, by asking the questions the researcher will give to you. Also, you have to classify the statements of Lisa as either truthful or deceitful (later in the questionnaire). During the interview, you can circle 'truth' or 'lie', after each question on a scratch paper. Try to make, after each question, an estimation of whether Lisa answers with a lie or with the truth.

Let the researcher know you've finished reading. She will now reach you the questions you need to ask during the interview.

4. Face-to-Face/Observer

The case

You work as a police officer for the criminal investigation department (recherche). You have been put on a case where a girl (16) is hit by a car and the suspect drove off. As a tactical criminal investigator, it is your job to collect as much information about a certain crime as possible. Therefore, you question victims and witnesses to get a better understanding of what exactly happened. You also search for camera footage of the area wherein a criminal act took place.

You are at the very beginning of this investigation. At this point, you don’t have many leads. The car accident happened on a deserted country road, the Lappenweg, outside of Enschede, on Friday night around midnight, with no cameras any close. You can't question residents in the area since within a radius of 2 km there are no houses. The victim herself was unconscious and did not see the car coming, since she was hit from the back. So, the only starting point for you now is questioning the witness who called 112.

Therefore, today your colleague will have an investigative interview with Lisa who says she witnessed the car accident.

Instructions

It is your task to observe the interview with Lisa. While doing so, you have to classify the statements of Lisa as either truthful or deceitful (later in the questionnaire). During the interview, you can circle ‘truth’ or ‘lie’, after each question on a scratch paper. Try to make, after each question, an estimation of whether Lisa answers with a lie or with the truth.

Let the researcher know you've finished reading.

Appendix C: Scenario actor

Case actor

You and your best friend (Tom) just had a few drinks at another friend's house. At 00.00 o'clock you and your friend decide to go home. Actually, you both drank a bit too much than is allowed, since your friend still needs to drive back by car. But since it is just a 10 minute drive you're not too worried. On your way back, unexpectantly, a biker appears. It is too late to avoid the biker and the cyclist gets hit by the car. You both are shocked and get out of the car. A relatively young girl (16) lies unconsciously next to her bike. Your friend is so stressed and cannot think clearly. Without saying anything, he walks back to the car and leaves the scene at high speed, while you are still standing there. Now you panic even more. You realize
that Tom and you are in big, big trouble. For one, drinking too much alcohol while driving, and two, for leaving the scene.

You know what is best to do but don't want any trouble at the same time. Therefore, you decide first to call 112, since the girl definitely needs medical help. For when the police arrive you decide not to tell the whole truth. For starters, you were just walking by. This makes you a witness instead of a partner in crime. You don't want to double-cross your friend either. Therefore, you decide to tell partly the truth, just enough to not get your friend involved. At the same time, you want to stay close to the truth since a CCTV might catch him on camera somewhere after he drove off. If you give a complete opposite description, that makes you suspicious.

On the printed pictures which the researcher gave you, you can see the location, the car of your friend and your friend himself. The car of your friend is a rather old and rare model/type but a common brand. In order to prevent that the search narrows quickly down to your friend (and you), you decide to lie about only the model. The color is also remarkable, namely the car is mint green. You don't want to say this, since only a few cars have this
color. Instead, you'll say another color. You know for sure that the police are going to ask about a description of the person in the car. You decide not to lie about the gender because this act is less typical for a woman and your story must look credible. You will lie about his height and will tell the truth about his age.

You are summoned for a witness interview. Today you will be interviewed.

Appendix D: Script for participants and actor

-	Hello Lisa, good that you are here. How are you?
	Neutral
1.	Ah okay. Well, I believe it is clear why you are here. You witnessed the accident and I'd like to ask you some questions. First, where did happen?
	Truth → picture 1
2.	At what time did it happen?
	Truth → picture 1
3.	Can you tell me from the beginning what happened?
	Lie
4.	What did you do at this time on the Lappenweg?
	Lie
5.	Okay, tell me more about the car. What was the color of the car?
	Lie → picture 2
6.	What brand was the car?
	Truth→ Volkswagen golf, picture 2
7.	Was it a new or old model?
	Lie
8.	Do you remember the license plate or parts of the license plate?
	Truth
9.	Okay, enough about the car for now. Can you describe the driver?
	Lie Don't mention age yet (=next question)
10.	What about his age, can you estimate his age?
	Truth
11.	What was the color of his hair?
	Truth
12.	Okay. The car hits the girl and you ran towards it. What happened then?
	Truth
-	That was it for now, thank you for your time!
	Neutral

Appendix E: Pictures actor

1. Location

Street: Lappenweg

City: outside the center of Enschede

Time: 00.00 o'clock

2. The car:

Volkswagen

Golf

3. The driver

Length:

+/- 1.90 m

Appendix F: Scratch paper for participants

-	Hello Lisa, good that you are here. How are you?

1.	Ah okay. Well, I believe it is clear why you are here. You witnessed the accident and I'd like to ask you some questions. First, where did it happen?

TRUTH LIE

2.	At what time did it happen?

TRUTH LIE

3.	Can you tell me from the beginning what happened?

TRUTH LIE

4.	What did you do at this time on the Lappenweg?

TRUTH LIE

5. Okay, tell me more about the car. What was the color of the car?

TRUTH LIE

6. What brand was the car?

TRUTH LIE

7. Was it a new or old model?

TRUTH LIE

8. Do you remember the license plate or parts of the license plate?

TRUTH LIE

9. Okay, enough about the car for now. Can you describe the driver?

TRUTH LIE

10. What about his age, can you estimate his age?

TRUTH LIE

11. What was the color of his hair?

TRUTH LIE

12. Okay. The car hits the girl and you ran towards it. What happened then?

TRUTH LIE

That was it for now, thank you for your time!

Appendix G: Questionnaire

INFO

As the last part of the experiment, you are asked to fill out some questions. The survey starts with some demographic-related questions. Then, you are asked to fill out some deception detection-related questions. These questions are rated on a 7-point scale. This will take you approximately 15-20 minutes.

DEMOGRAPHICS

What is your gender?

- Male (1)
- Female (2)
- Other (3)

What is your age?

[Text entry]

What is the highest degree or level of education you have completed?

- University level or higher
- University of Applied Science (HBO)
- Vocational Education (MBO)
- Secondary school (Middelbare school)
- Other

Are you a student?

- Yes
- No

What do you study?

[Text entry]

What is your nationality?

- Dutch
- German
- Other [Text entry]

Do you have corrected vision?

- Yes
- No

I wear

- Contacts
- Glasses
- Other [Text entry]

VERACITY JUDGEMENT ACCURACY ASSESSMENT

INFO

The following questions are regarding the first question asked: ‘Where did it happen?’

DVJ_1 Lisa told the right location where the accident happened.

TH_1 Lisa had to think hard before formulating her answer.

ID_1 Lisa seemed not interested.

VI_1 Lisa seemed to distance herself from her answer.

CO_1 Lisa seemed cooperative.

BC_1 Lisa’s behaviour changed after this question was asked.

INFO

The following questions are regarding the second question asked: ‘At what time did it happen?’

DVJ_2 Lisa lied about the time of the accident.

ID_2 Lisa seemed proactive

TH_2 Lisa answered immediately, without thinking long.

BC_2 I did not notice a change in Lisa’s behaviour after this question was asked.

VI_2 Lisa seemed to distance herself from her answer.

CO_2 Lisa seemed cooperative.

INFO

The following questions are regarding the 3rd question asked: ‘Can you tell me from the beginning what happened?’

DVJ_3 Lisa told the truth when she explained what happened.

VI_3 Lisa’s answer was personal.

ID_3 Lisa seemed uninterested.

TH_3 Lisa's answer was fluently formulated.

CO_3 Lisa seemed not very cooperative.

BC_3 I noticed a change in Lisa’s behaviour after this question was asked.

INFO

The following questions are regarding the 4th question asked: ‘What did you do at this time on the Lappenweg?’

DVJ_4 Lisa lied when she explained what she was doing at that time on the Lappenweg.

BC_4 Lisa behaved differently compared to the beginning of the interview.

ID_4 Lisa seemed enthusiastic.

TH_4 Lisa had to think hard after this question was asked.

CO_4 Lisa seemed willing to help.

VI_4 Lisa seemed to distance herself from her answer.

INFO

The following questions are regarding the 5th question asked: ‘Okay, tell me more about the car. What was the color of the car?’

DVJ_5 Lisa told the truth about the color of the car.

ID_5 Lisa seemed indifferent.

TH_5 Lisa had to think hard before answering.

VI_5 Lisa seemed to distance herself from her answer.

BC_5 Lisa’s behaviour did not change compared to the beginning of the interview.

CO_5 Lisa seemed not willing to help.

INFO

The following questions are regarding the 6th question asked: ‘What brand was the car?’

DVJ_6 Lisa lied about the brand of the car.

ID_6 Lisa seemed proactive.

TH_6 Lisa answered immediately, without thinking long.

BC_6 I noticed a change in Lisa’s behaviour after this question was asked.

VI_6 Lisa seemed to distance herself from her answer.

CO_6 Lisa seemed cooperative.

INFO

The following questions are regarding the 7th question asked: ‘Was it a new or old model?’

DVJ_7 Lisa lied about the model.

VI_7 Lisa distanced herself from her answer.

ID_7 Lisa seemed uninterested.

TH_7 Lisa seemed to think carefully about formulating her answer.

CO_7 Lisa seemed not very cooperative.

BC_7 I noticed a change in Lisa’s behaviour after this question was asked.

INFO

The following questions are regarding the 8th question asked: ‘Do you remember the license plate or parts of the license plate?’

DVJ_8 Lisa lied about not knowing the license plate.

BC_8 Lisa behaved differently compared to the question before.

CO_8 Lisa seemed willing to help.

ID_8 Lisa seemed proactive.

TH_8 Lisa answered immediately, without thinking long.

VI_8 Lisa seemed to distance herself from her answer.

INFO

The following questions are regarding the 9th question asked: ‘Okay, enough about the car for now. Can you describe the driver?’

DVJ_9 Lisa gave the wrong description of the driver.

ID_9 Lisa seemed indifferent.

TH_9 Lisa did not have to think hard before answering.

VI_9 Lisa's answer was personal.

BC_9 Lisa’s behaviour changed compared to the question before.

CO_9 Lisa seemed cooperative.

INFO

The following questions are regarding the 10th question asked: ‘What about his age, can you estimate his age?’

DVJ_10 Lisa lied about the age of the driver.

ID_10 Lisa seemed proactive.

TH_10 Lisa answered quickly, without thinking long.

BC_10 I noticed a change in Lisa’s behaviour after this question was asked.

VI_10 Lisa distanced herself from her answer.

CO_10 Lisa seemed not very cooperative.

INFO

The following questions are regarding the 11th question asked: ‘What was the color of his hair?’

DVJ_11 Lisa told the truth about the driver's hair color.

VI_11 Lisa distanced herself from her answer.

ID_11 Lisa seemed uninterested.

TH_11 Lisa answered immediately, without thinking long.

CO_11 Lisa seemed not very helpful.

BC_11 I noticed a change in Lisa’s behaviour after this question was asked.

INFO

The following questions are regarding the 12th question asked: ‘Okay. The car hits the girl and you ran towards it. What happened then?’

DVJ_12 Lisa lied in her narrative.

BC_12 Lisa behaved the same compared to the question before.

CO_12 Lisa seemed willing to help.

ID_12 Lisa seemed indifferent.

TH_12 Lisa had to think hard before answering.

VI_12 Lisa’s answer was personal.

- Strongly Disagree (1)
- Disagree (2)
- Somewhat Disagree (3)
- Neutral (4)
- Somewhat Agree (5)
- Agree (6)
- Strongly Agree (7)

BEHAVIOURAL FAMILIAIRTY

PRI_1 (preconceived ideas) I believe that people who lie avoid eye contact.

PRI_2 I believe that people who lie keep their hands still.

PRI_3 I believe that people who lie fidget more with their fingers.

PRI_4 I believe that people who lie look in many directions and to a lesser extent at the conversational partner.

PRI_5 I believe that people who lie blink more than they should naturally do.

PRI_6 People who lie sit in general more still.

PRI_7 I believe that people who lie use more illustrators (hand gestures) that accompany speech.

PRI_8 I believe that people who lie stare you more continuously in the eye.

PRI_9 I believe that people who lie move their feet and legs more than someone who tells the truth.

- Strongly Disagree (1)
- Disagree (2)
- Somewhat Disagree (3)
- Neutral (4)
- Somewhat Agree (5)
- Agree (6)
- Strongly Agree (7)

CONTEXTUAL FAMILIARITY

CF_1 (contextual familiarity) The description of the scenario influenced my perception of the situation/ influenced my judgement regarding Lisa.

- Strongly Disagree (1)
- Disagree (2)
- Somewhat Disagree (3)
- Neutral (4)
- Somewhat Agree (5)
- Agree (6)
- Strongly Agree (7)

RELATIONAL FAMILIAIRTY

RF_1 (relational familiarity) I know the actor.

RF_2 I interacted with the actor before.

- Strongly Disagree (1)
- Disagree (2)
- Somewhat Disagree (3)
- Neutral (4)
- Somewhat Agree (5)
- Agree (6)
- Strongly Agree (7)

OTHER (Branched if participants were in one of the video conditions)

Name_1 Did you notice the name Lea in the video call?

- Yes (1)
- No (2)

Name_2 Did it confuse you that two different names were used?

- Not at all (1)
- A little (2)
- Very much (3)

DEBRIEFING

Thank you for taking the time to participate in this study. As mentioned earlier, this study researches deception detection. We used four different ways (conditions) in which the interview is computed. In the first condition, participants had an online interview with Lisa. In the second condition, participants were instructed to observe an interview between Lisa and a colleague which was recorded earlier. In the third condition, participants had to interview Lisa face-to-face. In the fourth condition, participants had to observe the interview between Lisa and a colleague, while being present in the same room. In sum, the conditions varied in either observing or receiving lies and in either having the interview face-to-face or via a video.

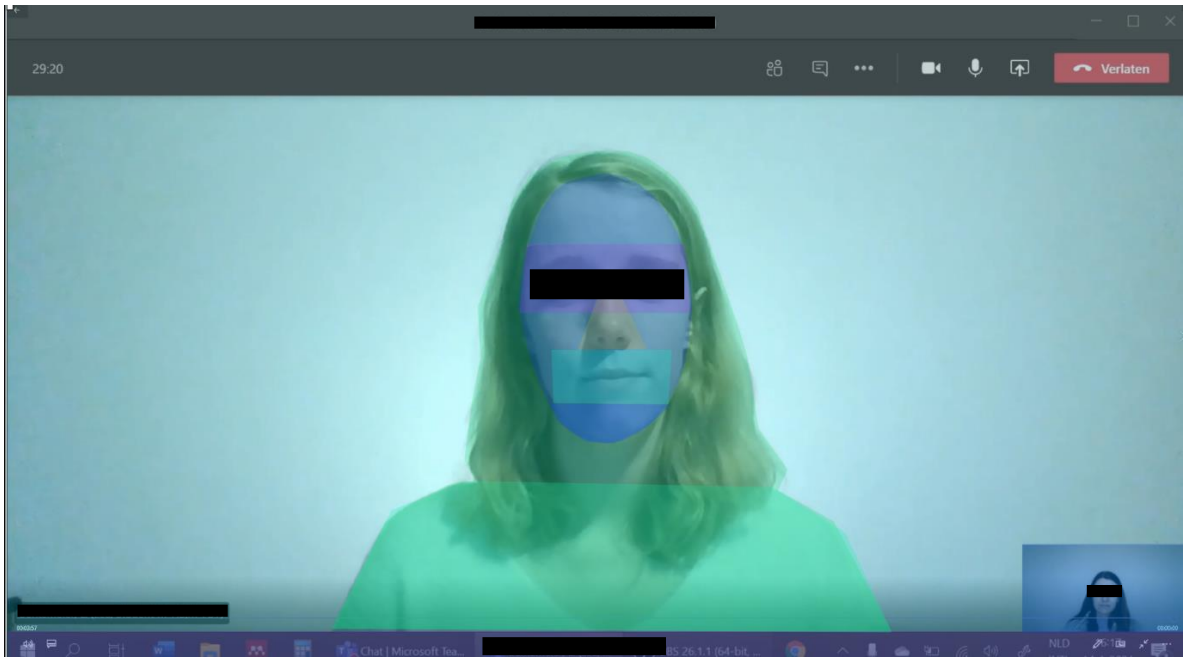
[You were in the Video/Receiver condition.] [You were in the Video/Observer condition.] [You were in the F2F/Receiver condition.] [You were in the F2F/Observer condition.]

We now want to compare potential differences between the four conditions in the degree of accuracy in distinguishing truths from lies. For instance, it might be that receivers of lies are less accurate in comparison to observers since receivers use more cognitive capacity which cannot be devoted to deception detection because they need to lead the interview. We will also compare the four conditions in gaze behaviour.

You can email j.m.schimmel@student.utwente.nl if you are interested in the results of this study.

Appendix H: Visualization of AOIs on target

Areas of interest for the Video/Observer condition



Areas of interest for the Face-to-Face/Receiver condition

Appendix I: Truth and lie accuracy

	Groups									
	Video/Receiver		Video/Observer		F2F/Receiver		F2F/Observer			
	Truth *M (SD)*	Lie *M (SD)*	Truth *M (SD)*	Lie *M (SD)*	Truth *M (SD)*	Lie *M (SD)*	Truth *M (SD)*	Lie *M (SD)*	Overall truth *M (SD)*	Overall lie *M (SD)*
Direct	.56 (.25)	.39 (.24)	.48 (.28)	.45 (.15)	.46 (.23)	.42 (.25)	.51 (.19)	.35 (.30)	.50 (.24)	.40 (.24)
Indirect	.56 (.13)	.33 (.09)	.55 (.16)	.39 (.10)	.60 (.16)	.32 (.15)	.58 (.15)	.32 (.14)	.57 (.15)	.34 (.12)
Overall	.56 (.14)	.34 (.10)	.54 (.16)	.40 (.10)	.57 (.16)	.34 (.15)	.57 (.14)	.33 (.15)	.56 (.14)	.35 (.13)

Appendix J: Slopes, standard errors and p-values for predictors Multiple Regression

Multiple regression predictors fixation durations on the distinct areas of the face

Accuracy	*B*	*95 % CI for B*		*SE B*	*β*	*p*	R^2	ΔR^2
		LL	*UL*					
Model							0.067	-0.008
Constant	0.458	0.400	0.515	0.029		0.000		
Forehead	-0.067	-0.141	0.008	0.037	-0.279	0.080		
Eyes and brows	0.030	-0.084	0.145	0.057	0.102	0.599		
Nose area	-0.001	-0.072	0.071	0.036	-0.004	0.985		
Mouth area	-0.005	-0.060	0.050	0.028	-0.030	0.858		
Chin	0.012	-0.066	0.091	0.039	0.047	0.754		

Note. Model = 'Enter' method in SPSS Statistics; *B* = unstandardized regression coefficient; CI = confidence interval; *LL* = lower limit; *UL* = upper limit; *SE B* = standard error of the coefficient; β = standardized coefficient; R^2 = coefficient of determination; ΔR^2 = adjusted R^2.

Multiple regression predictors fixation counts on the distinct areas of the face

Accuracy	*B*	*95 % CI for B*		*SE B*	*β*	*p*	R^2	ΔR^2
		LL	*UL*					
Model							0.060	-0.016
Constant	0.460	0.402	0.519	0.029		0.000		
Forehead	-0.003	-0.007	0.002	0.002	-0.180	0.201		
Eyes and brows	3.935E-05	-0.001	0.001	0.000	0.021	0.886		
Nose area	-3.618E-05	-0.001	0.001	0.001	-0.009	0.954		
Mouth area	0.001	0.000	0.001	0.000	0.214	0.256		
Chin	-0.001	-0.005	0.002	0.002	-0.162	0.346		

Note. Model = 'Enter' method in SPSS Statistics; *B* = unstandardized regression coefficient; CI = confidence interval; *LL* = lower limit; *UL* = upper limit; *SE B* = standard error of the coefficient; β = standardized coefficient; R^2 = coefficient of determination; ΔR^2 = adjusted R^2.

Multiple regression predictors visit counts on the distinct areas of the face

Accuracy	*B*	*95 % CI for B*		*SE B*	*β*	*p*	R^2	ΔR^2
		LL	*UL*					
Model							0.096	0.023
Constant	0.458	0.400	0.516	0.029		0.000		
Forehead	-0.001	-0.007	0.006	0.003	-0.036	0.816		
Eyes and brows	-0.001	-0.004	0.001	0.001	-0.270	0.177		
Nose area	2.456E-05	-0.001	0.001	0.001	0.005	0.972		
Mouth area	0.002	0.000	0.004	0.001	0.451	0.047		
Chin	-0.003	-0.007	0.001	0.002	-0.308	0.103		

Note. Model = 'Enter' method in SPSS Statistics; *B* = unstandardized regression coefficient; CI = confidence interval; *LL* = lower limit; *UL* = upper limit; *SE B* = standard error of the coefficient; β = standardized coefficient; R^2 = coefficient of determination; ΔR^2 = adjusted R^2.