

University of Tartu
Faculty of Social Sciences
Institute of Psychology

Krõõt Soomets

**THE EFFECT OF FEEDBACK-BASED TRAINING ON INTUITIVE DECEPTION
DETECTION ACCURACY**

Master's thesis

Supervisor: Annegrete Palu

Tartu 2022

The effect of feedback-based training on intuitive deception detection accuracy

Abstract

Literature has shown that human accuracy in deception detection is almost equal to chance level. Mediocre levels of deception detection are undesirable in everyday social situations but can be dangerous in the law enforcement and criminal context. The aim of the current study was to investigate the effects of feedback-based training during which participants were asked to make veracity judgments intuitively. The experiment consisted of a pre-training test, training, and a post-training test during which participants watched stimulus videos consisting of people telling truthful and deceitful stories. In addition to judging the videos' truthfulness, decision-making time was measured, and participants were asked to give confidence ratings for their answers. Participants ($N = 66$) made 1615 veracity judgments in total. Results did not find a significant increase in deception detection accuracy between pre- and post-training tests, however, truth detection accuracy decreased significantly. Confidence ratings predicted accuracy only in the post-training stage of the experiment, suggesting that participants learned to give more accurate estimates of their confidence after training. Decision-making time was not predictive of accuracy. The results demonstrated that brief feedback-based training is not effective in improving intuitive detection accuracy.

Keywords: deception detection accuracy; training; intuitive processing; confidence, decision-making time

Kohese tagasisidega treeningu mõju intuiitivsele valetamise tuvastamise täpsusele

Lühikokkuvõte

Varasemad uurimused on aga näidanud, et suudame korrektselt valesid tuvastada vaid umbes 54% juhtudest ehk vaevu juhusest paremini. Valetamise korrektne tuvastamine on kasulik igapäevaelus, kuid on eriti oluline õigusosalases kontekstis, kus korrektne tuvastus võib viia õigusrikkuja karistamiseni. Käesoleva uurimustöö eesmärk oli uurida kohese tagasisidega treeningu mõju inimese valede tuvastamise täpsusele, paludes katseisikul langetada otsuseid intuiitivselt. Eksperiment koosnes treeningueelne testist, treeningust ja treeningujärgne testist, millede jooksul vaatasid katseisikuid videoid lühijutte kõnelevatest inimestest. Lisaks tuvastustäpsusele uuriti ka kindlushinnangu ning otsuse langetamiseks kulunud aja seost nende tuvastamise täpsusega. Osalejad ($N=66$) langetasid kokku 1615 videote tõesust puudutavat otsust. Tulemused ei näidanud statistiliselt olulist tõusu valetamise tuvastamise täpsuses treeningueelse ja -järgse testi vahel. Seevastu leiti aga statistiliselt oluline langus tõdede tuvastamise täpsuses. Kindlushinnang ei ennustanud tuvastustäpsust treeningueelses katseosas, kuid tegi seda treeningujärgses osas, mis viitab asjaolule, et katseisikud õppisid oma otsuseid täpsemalt hindama. Otsuse langetamise aeg ei ennustanud tuvastamise täpsust ei treeningueelses ega -järgses osas. Tulemustest saab järeldada, et lühiajaline tagasisidel ning intuiitivsetel otsustel põhinev treening ei ole efektiivne parandamaks inimeste valetamise tuvastamise täpsust.

Märksõnad: valede tuvastamise täpsus; treening; intuiitivne infotöötlus; kindlushinnang; otsuse langetamise aeg

Lying is one of the most common forms of deception. According to Ekman (1997), lying can be differentiated from other forms of deception by two criteria. The first and most important criterion is the intention of the liar (or the sender). Ekman believes that lying can only be defined as a deliberate act of deception, not an accidental mistake when sharing information. Second important criterion in defining lying is the notion that the target (or the receiver) does not know about the liar's intention to deceive. The current study relies on Ekman's interpretation and defines lying as a conscious action with the purpose to deceive without the other person's knowledge.

As lying can have many different objectives, literature has tried to systematise this by creating lie typologies. Cantarero and colleagues (2018) propose that lies can be categorized by their motivation (gain benefits vs avoid punishment) and the beneficiary (oneself vs others vs both). A similar typology has been suggested by Erat and Gneezy (2012), who mention four types of lies: altruistic white lies, Pareto white lies, selfish black lies and spiteful black lies. Both types of white lies are benign as they benefit either the receiver or both parties accordingly. Spiteful lies are ones that hurt both parties, whereas selfish lies, the most well-known type, profit the sender at the expense of the receiver. These detailed typologies demonstrate that lies differ in their nature depending on the goal they are supposed to help achieve.

It is evident that lying can be beneficial, and that these benefits can sometimes outweigh the immorality of the action, leading to deception. Although some state that lying is opted for only when other means of reaching an objective seem too difficult (Miller & Stiff, 1993, as cited in Cantarero et al., 2018), DePaulo et al (1996) demonstrate that lying is a common part of our social lives with people telling one lie per day on average. It is probable, that many of these lies are white lies and are not too detrimental to the involved parties. However, black lies can cause more damage, especially when considering the criminal context, where an undetected lie can result in a criminal's freedom or an innocent person's punishment. Since lying is shown to be common and possibly damaging, it brings people's ability to detect lies into question. It can be argued that our wish to avoid any harm should make us vigilant of lying and improve our ability to detect it. This idea has inspired the current study to explore deception detection and the possibility of improving it by training.

Deception Detection

The high prevalence of lying and the possible damage it may cause to the receiver should bring about an evolutionary pressure to advance in deception detection. However, despite the potential evolutionary pressure, research has shown that human deception

detection accuracy is 54%, which is almost equal to chance level (Bond & DePaulo, 2006). This detection rate may appear sufficient in everyday life, but seems dangerously low for law enforcement, where an undetected lie can lead to detrimental results. Multiple theories suggest explanations for our mediocre lie detection skills. The factors that may influence our deception detection ability include biases, specific situational factors, and individual differences.

Biases

In his Truth-Default Theory (TDT), Levine (2014) argues that our poor deception detection skills can be explained by our default assumption that people are honest. This truth-bias is believed to exist and persist due to most communications being honest in everyday life. The presumption of honesty is considered to play an important part in effective communication when socializing. Constant doubts about another's truthfulness would inhibit conversation and cooperation (Vrij, 2008). Furthermore, falsely accusing others of lying may have negative social repercussions (O'Sullivan, 2003). This truth-bias leads to the veracity effect, which describes the notion that people identify truthful statements more accurately than deceptive ones (Hartwig et al., 2004). Hence, TDT argues that due to the nature of our everyday communications, humans are more inclined to believe others than suspect deceit.

The opposite phenomenon is called the lie-bias, which is defined as a tendency to judge any received information as deceptive (McCornack & Levine, 1990). Lie-bias produces the reversed veracity effect, meaning the biased person is more accurate when judging lies, but may not be as successful in detecting truths (Schindler et al., 2021). Lie-bias holds a certain degree of suspicion, leading to the idea that police officers could be more lie-biased when communicating with suspects. Multiple studies have found this to be true (Meissner & Kassin, 2002; Masip & Herrero, 2017). This bias may lead officers (or people with similar professions) to always presume the suspect of being guilty. This can lead to police induced confessions, which could have devastating effects if the people involved are in fact, not guilty (Meissner & Kassin, 2002). Altogether, the existence of lie-bias tells us that some personal or situational factors may influence people to judge communications as deceptive more often than not.

Situational or Personal Factors

When discussing differences in lie detection accuracy, it is useful to investigate which personal or situational factors may be involved in causing the variance. This can help us understand what affects deception detection and whether this information can be used to improve people's ability to correctly distinguish lie from truth.

One personal factor to consider is the amount of experience that people have had with deception and its detection. Here, the logical cohort to investigate are police officers and other vocations where deception detection is a part of the job description (e.g., airport personnel, security, medical professionals) (Leach et al., 2004). Some indication of experience possibly increasing deception detection accuracy comes from Whelan et al. (2015) who showed a 72% accuracy in police officers. High detection scores were also shown for law-enforcement groups and clinical psychologists by Ekman et al. (1999). However, it is important to note that similar professions have been investigated in other studies, where high detection accuracies were not found (Bond & DePaulo, 2006). This variance in results may result from multitude of reasons. Firstly, a specific amount of experience (e.g. certain number of years) might be necessary for it to significantly affect deception detection accuracy, which may not have always been achieved by the samples under investigation. Secondly, none of the studies can fully confirm that experience is the critical personal factor that predicts detection success. This means, other factors that influence detection may have existed in the sample groups of some studies and not in the others, without us having the knowledge of it. In short, experience is a personal factor that may be a useful indicator of detection success but is not definitive. The current study aims to add information to this research field by investigating whether gaining experience during a deception detection training programme improves detection accuracy.

In addition to and related with experience is the effect that confidence may have on deception detection. Often, the more experience we gain with doing something, the more confident we become at doing it. In deception detection context, literature has continued its investigation on people working in law enforcement or similar fields. An investigation into law enforcement officers' and students' ability to detect children's lies showed that officers' confidence during veracity judgements was significantly higher, however they were not more accurate in their judgements (Leach et al., 2004). In another study, DePaulo and colleagues (1997) also found that high confidence is not associated with higher accuracy in deception detection tasks. These results demonstrate that confidence does not always positively correlate with accuracy, as it is always possible to be confidently incorrect. The current study will contribute to this research by investigating confidence-accuracy interactions in a deception detection training programme context.

In addition to personal aspects, it is also important to discuss situational factors that may influence deception detection success. One of such factors is the medium of communication between the sender and the receiver. Instinctively, detection should be easiest

during a face-to-face communication, as some information is not shared when communicating only via medium (e.g. inability to see facial expressions on a phone call). As a face-to-face interactive communication is hard to produce and control in a study setting, an audio-visual stimulus is the next best option. Research has found that deception detection accuracy is higher, when the stimulus is audio-visual, compared to pure visual or pure audio stimulus (Sweeney & Ceci, 2014). Thus, the current study has opted for audio-visual stimulus material as well. In addition to simplifying the experiment setup, the use of audio-visual stimuli is also reflective of today's world where a large portion of media is consumed in this form.

Cues of deception

In addition to the effects that biases and abovementioned factors may have on deception detection, the main influencer is arguably our belief that liars act inherently differently from honest people. The discussion over behavioural cues that indicate deception has been long and full of controversy. Studies have investigated the reliability and validity of many behavioural cues and found that many lack diagnostic value when it comes to deception detection, meaning attending to these cues raise deception detection accuracy significantly at all (Levine et al., 2005; DePaulo et al., 2003). However, it is important to discuss the extensive research conducted on the topic, as the current study aims to lessen biases that may be influenced by previous knowledge of behavioural cues of deception.

Over the years, people have developed certain beliefs about which behaviours liars demonstrate that allow us to recognise deception. For example, nervous fidgeting and looking to the right while speaking are some examples of behaviours that the general population judges suspicious and possibly indicative of lying (Gillet & Lee, 2019). Scientific research has also previously proposed an extensive list of possible deceptive cues: Elaad (2009) lists pupil dilation, fewer hand movements, fewer illustrations, more negative statements, and higher pitched voice as valuable deception cues. However, the author states that gaze aversion and frequent head movements and shifts in whole body position are not of diagnostic value when it comes to deception detection. Levine and colleagues (2005) list indicators like longer response latency, more adaptors, more pauses, and speech errors. The large number of proposed behavioural cues raises the question whether all of them can truly indicate lying.

A few reasons have been proposed as to why concentrating on behavioural cues is not helpful in a deception detection scenario. Firstly, it is possible that deception detection suffers due to the mere number of possible cues to consider. As no single cue exists that would

reliably predict deception, we are forced to analyse many cues simultaneously (provided that they even exist or are displayed). Trying to spot all deceptive cues and then incorporate them into an analysis that results in a successful deception detection is arguably a difficult task to undertake, especially when energy or time resources are scarce. Additionally, as there is no single cue that predicts deception, there is also no certain combination of all the mentioned cues that would always inform us of deception (Levine et al., 1999).

Besides the large number of possible detection cues and the added complexity of combining them for optimal deception detection, we may be relying on the wrong cues after all. For example, the previously mentioned higher pitched voice and pauses and errors in speech may come about with nervousness or some other emotion. In a deception detection situation, it is often assumed that anxiety reflects the guilt and fear that the liar feels when trying to mislead the other. However, a truthful person may also show signs of anxiety for many different reasons (e.g. being in a stressful situation and being suspected of something) (Bogaard et al., 2016). Simultaneously, not all people are necessarily nervous or feel guilt when lying, rendering these cues useless in the context of deception detection. Thus, the cues we may be inclined to search for in a liar may mislead us or not be there at all.

The noticeability of the behavioural cues may be another reason why deception detection often fails. Hartwig and Bond (2011) propose the weak objective cue hypothesis, which as the name states, describes the idea that the real deception detection cues are too minute for us to notice. The weak objective cue hypothesis becomes even more important when we consider the possible role of practicing one's lies. It can be argued that the already minute cues become even less detectable once the sender has practiced the lie. The idea that practice affects the behavioural cues we display relates back to the idea that not all liars' display signs of guilt or nervousness as rehearsing one's lies may lessen nervousness or even eliminate such behavioural cues.

This variance in literature pertaining behavioural cues of deception led Vrij (2008) to write an article in which he strongly recommends a change in police practice. Accumulating different results in his work, Vrij states that behavioural cues of deceptions do not carry diagnostic value and should not be used. Instead, he suggests concentrating only on verbal cues which reveal contradictions in the suspect's statements. Although verbal content cues have been shown to be better predictors of deception (Levine, 2015), they are difficult to detect as they often rely on previous knowledge and communication with the possible sender. This is because detecting inconsistencies in the sender's story often relies on previous information and the chance to interact with them. As such information is not always

available, some researchers have started to concentrate on forms of deception detection that are free from active cue detection.

Intuition

Our shortcomings in information processing and the many inconsistencies within deception cue literature inspired the emergence of another perspective in this area – the role of intuition in detecting lies. APA Dictionary of Psychology defines intuition as ‘*immediate insight or perception, as contrasted with conscious reasoning or reflection.*’ (American Psychological Association, n.d). In the context of deception detection, intuitive processing would thus mean making the veracity judgment quickly and without deep conscious processing. The theoretical basis supporting the benefits of intuitive processing states that purposeful systematic processing of multiple cues (behavioural or verbal) is potentially too difficult for us. It may hinder the decision process and accuracy due to the overflow of information and an inability to fully analyse and integrate said cues (Zhou, Zhang & Sung, 2013; Reinhard et al., 2013). In other words, trying to concentrate on multiple cues at once may result in very shallow processing of said cues and this inadequate processing may hinder decision-making and make people doubt themselves. Thus, research has changed direction from investigating cues to the role of intuition in deception detection.

Multiple techniques have been used to bring about intuitive processing in participants during experiments. For example, Albrechtsen and colleagues (2009) opted for cutting stimulus videos into 5 second clips (thin-slicing), which they argued are short enough and contain minimal information making systematic processing so difficult, that participants have to answer intuitively. Albrechtsen et al. (2009) also conducted a second experiment during this study, in which participants’ intuition was encouraged by asking them to do a concurrent memory task while watching (this time full-length) stimulus videos. Reinhard et al (2013) used a somewhat similar technique. However, instead of a concurrent task, the experimental group was asked to work on a difficult puzzle for 3 minutes after watching all the recordings. The main idea behind the use of concurrent and directly subsequent tasks is to disrupt subjects’ processing of the stimulus information either at the time of viewing it or directly after. Ten Brinke et al (2014) used a complicated experimental design that included stimulus mock-crime videos and implicit association tasks. The results showed that indirect measures of deception detection are superior to direct measures, hinting towards the idea that unconscious or intuitive deception detection may be more accurate than cue detection. It is apparent that research on intuition in deception detection is novel and has accordingly produced many different experimental methods.

Although the previously mentioned theories of systematic vs intuitive processing support the use of such methodologies, they come with shortcomings. For example, Reinhard et al (2013) failed to mention how much time was given to the participants to make their veracity decisions after the puzzle task. If participants were given infinite time then it is possible that participants slipped into deliberate processing instead of answering intuitively. In addition to problematic experimental designs (or insufficient descriptions of them), a study can be equally invalid and unreliable if the data analysis is incorrect. This is what happened with the work of Ten Brinke et al (2014) whose study was later strongly criticized for its possibly flawed data analysis (Franz & von Luxburg, 2015). Due to these flaws, the current study does not replicate these methodologies and tries to add to the intuition literature with a different approach.

Training deception detection

Investigations into the multitude of ways we can approach deception detection and our accuracy in this area has brought into question whether it is a skill that can be improved by training. The most discernible benefit of increased detection accuracy would be seen in the criminal context, where it could aid investigations become more efficient and bring about fair verdicts more often. The idea behind training deception detection relates back to the previously mentioned findings that people of some professions are sometimes better at deception detection than average (Ekman et al., 1999; Whelan et al., 2014). It is possible that the main difference between these professionals and laypeople is the amount of experience and feedback with deception detection that they have gained in their vocation, implying that by giving people more experience in this activity, it is possible to improve their accuracy.

Many deception detection training studies concentrate on teaching participants about behavioural cues of deception and investigating whether that improves their accuracy. Three decades ago, Turck and Miller (1990) conducted a study during which the experimental group received a 30-minute training on cues of deception, many of them being behavioural cues. The results showed that the training group's accuracy did improve. In addition to measuring participants' accuracy over training and time, the authors also measured confidence and found that training did not lead to overconfidence. George et al. (2003) created a similar experiment, during which subjects participated in lectures about deception detection, its cues and heuristics in decision making. The authors fail to mention which cues and heuristics subjects learned about, but the results showed that lecture-based training on deception cues did not improve subjects' accuracy in a later deception detection task. Porter et al., (2010) also used lecturing as part of their three-hour training program but improved it

by adding feedback exercises to it as well. The results showed an improvement in total accuracy rates, which indicates that practice detection exercises might be a key part of a successful training program.

As mentioned above, giving participants feedback on their accuracy during a training session may be one of the crucial factors that raises its effectiveness. Feedback's possible efficacy is built on the premise that our currently seen low deception detection rates are due to lack of "training" in real life (Porter et al., 2000). This is not to say that humans are incapable of learning from their mistakes. Rather, the problem lies in the fact that when it comes to everyday life, we rarely get to know if we were lied to. Hence, we fail to learn to detect lies. Even if one supposedly detects a liar, the feedback to know whether the judgment was correct may come too late to teach the receiver to make the same judgment in similar future occasions (Vrij & Semin, 1996).

The effect of feedback on detection accuracy was investigated by Porter and colleagues (2000). They conducted a study which measured and compared the effectiveness of two training programs. One of these programs included informing subjects about deception cues and giving them feedback on their veracity judgment accuracy. The second program only included feedback without any information about deception cues. The results indicated that accuracy improved approximately equally well in both training groups when compared to the control group, supporting the notion that feedback only might be enough to train detection. This is an important finding considering the controversy in deception cue research. Furthermore, a training program that includes only feedback is significantly easier to produce and administer to larger groups of people. Taking these results into consideration, the current study does not inform participants on any sort of deception cues and instead focuses on giving feedback on their judgments during training.

Overall, deception detection training studies show a small to medium training effect (Hauch et al., 2016). The variance of results seems to reflect the previously mentioned contradicting literature pertaining behavioural cues in deception. Due to this, the validity and reliability of training programs that teach individuals of these behavioural cues are also uncertain, especially when it has been suggested that training individuals to only look for certain cues causes biased information seeking and processing, leading to lie bias and inaccurate judgments (Masip et al., 2009). The most successful form of training is based on verbal content cues (Hauch et al., 2016), which supports Vrij's (2008) proposal that police should only rely on verbal cues. However, there is always a possibility, that the sender makes no mistakes in their verbal communication (e.g. they have practiced the lie) in which case

there are no relevant cues. Hence, the study at hand is interested in adjoining the concept of intuitive processing and deception detection training which includes feedback to investigate its effect on deception detection accuracy.

The current study

Literature on deception detection has largely concentrated on the many possible cues of deception and whether teaching these can be used to improve detection accuracy. However, research in this field has produced mixed results. The often contradicting findings have inspired a new direction of research that is arguably more simple in its content. Instead of trying to distinguish the best cues of deception, studies are investigating the role of intuition in deception detection. The study at hand aims to add value to this research area by investigating the effectiveness of a deception detection training programme which aims to lessen the effect of biases, induces intuitive processing, and provides participants feedback on their progress. Aspects that may affect the reliability and validity of this study (and other deception detection research) are discussed below.

One critical factor in deception detection research is the act of deception itself. It can be argued that our ability to spot lies is dependent on the liar and the lie told. One factor to consider and that can be taken under control if necessary is the liars' gender. Literature has demonstrated mixed results when it comes to gender differences in lying. For example, Capraro and colleagues (2018) showed that men are more likely to lie than women. Erat and Gneezy (2012) found similar, but more nuanced results, where men are more likely to tell Pareto white lies and selfish black lies, while women have the tendency to tell more altruistic white lies. Yet, Bizziou-van-Pol et al. (2015) found the opposite. In their study, men were more likely to tell an altruistic white lie than women. Furthermore, Lohse and Qari (2021) discuss how lying behaviour may be affected differently for men and women, based on whether deception takes place face-to-face or not (e.g., via audio). According to the authors, a face-to-face context may increase a person's awareness of their social image and how lying may affect it. The variety of findings show that gender differences in lying might exist, but seem to be dependent on the context, the reward and/or the lie itself. In the current study, all the senders are telling lies and truths in the same context, on the same topics and with no reward or punishment. Thus, it was not deemed important to balance the senders by gender in this experiment.

In addition to the characteristics of the liar, the content and type of lie may also influence our success of detecting it. Besides the black lies vs white lies typology mentioned above, researchers have argued about the importance of difference between high-stake and

low-stake lies. For example, Whelan et al (2014) argue that the low accuracy rates demonstrated in some of the deception detection literature may be an artefact of the experimental designs, which usually include detecting low-stake lies. This means that the consequence of finding out the truth would not be too detrimental, which is why the sender does not show enough deceptive cues (if they exist). It can be argued that criminal offenders encounter more high-stake lies than the average laymen. Due to their situational circumstances, differentiating between truths and lies may sometimes even be a life-or-death question. However, a recent study showed that offenders did not perform better in deception detection compared to non-offenders (Schindler et al., 2021). This suggests that the low-stake vs high-stake lies differentiation may not be of importance when it comes to detection accuracy. Although the current study includes so-called low-stake lies, it can be argued that the same lies could be high-stake in a different context.

In addition to the high-stake vs low-stake lie argument, it is important to ensure that the truths and lies participants receive are of equal difficulty. Arguably, a lie that includes a true story with a small detail change (e.g. lying about one's shirt colour) seems qualitatively different from a story that is completely made up. Thus, the current study proposed questions based on the Estonian Code of Criminal Procedure § 62 to help equalize the content and difficulty of the truths and lies (see Method). As the objective is to distinguish whether the training program is successful, all stories are told in the same setting and context.

The aim of this research paper is to investigate the effectiveness of a deception detection training with instant feedback. The following is hypothesised:

H1: Deception detection training improves lie detection accuracy

H2: Confidence ratings are not predictive of detection accuracy.

The study also poses two research questions. According to Hauch et al. (2016), it is crucial for a training program to measure both the accuracy of detecting lies and the accuracy of detecting truths. Hence, the current study asks the following:

RQ1: What effect does the training program have on truth accuracy specifically?

Furthermore, no literature currently exists on the association between decision-making times and deception detection accuracy (e.g. the faster the veracity judgment the more/less accurate one is). Thus, the current study will exploratively investigate the role of decision-times in veracity judgments' accuracies:

Explorative RQ: Does decision-making time predict detection accuracy?

Method

Sample

Participants for this study were recruited using convenience sampling and advertisements were posted on social media and appropriate University of Tartu mailing lists. Participants had to be over the age of 18 and speak Estonian as their native language.

A total of 80 individuals began the experiment, out of which 67 people completed it. From the 13 people that did not complete the experiment, 6 gave only their demographic information and read the instructions, whereas the remaining 7 had also proceeded to watch some of the stimulus videos. Thus, their data was not usable in the current study.

From the 67 participants who completed the study, 7 were acquaintances with someone from the stimulus videos. Out of these 7 participants, one was fully excluded as their post-training test consisted of only videos of people they knew, rendering it impossible to analyse the effectiveness of the training. The other 6 subjects were included but data of the videos with their acquaintances were removed. This creates a more informative analysis and is more ethical as the effort of these participants is not fully lost. The final sample size was $N = 66$. The mean age of the participants was 32.03 years ($SD = 11.69$). The sample consisted of 51 women, 13 men, and 2 people reporting 'Other' for their gender.

Participation in the study was voluntary. All subjects were informed of the study's purpose and conditions. Subjects could only proceed with the experiment once they had given their consent to participate. All participants were offered to enter a draw to win a 20€ (x2) gift card. Furthermore, undergraduate psychology students were offered course credits for their participation.

Materials

The stimuli consisted of 50 video clips in which five people (2 men, 3 women) truthfully and deceitfully described personal stories. People who were filmed for the stimuli were native Estonian speakers with the mean age of 31.6 years ($SD = 6.66$) at the time of filming. The people recruited for this task had to be over the age of 18 and were not allowed to later participate in the experiment. For clarity purposes and easier differentiation from subjects participating in the experiment later, the people in the stimulus videos are called 'actors' from here onward. However, it is important to note that none of the participants were professional actors nor did they practice for this activity.

The videos were filmed using the Logitech ConferenceCam Connect camera and during the experiment they were presented at the size of 732x412 pixels and were centred on the screen. All filming was done in the same location and actors were asked to wear a black

T-shirt to keep the stimulus videos as visually consistent as possible (Figure 1). In addition to minimising distractors, using neutral T-shirts can also help to lessen any prejudices about the actors (e.g. no slogans or brands reflecting their values or beliefs).

Figure 1

Frame setup for the stimulus videos



Each actor was asked to tell one truthful and one deceitful personal story on each of these five topics: travelling, graduations, competitions, injuries, and entertainment events. Actors were not given time to practice their stories and the first clips to be filmed were used in the study. The questions used to guide the filming of stimulus videos were based on the Estonian Code of Criminal Procedure § 62, which states that time, place, and manner of commission of criminal offence are subjects of proof (Code of Criminal Procedure, 2022). When telling a lie, participants were asked to lie about all three mentioned aspects: where the event took place, when it took place and what did the subject do during this event. This helped to ensure that lies and truths were similar in terms of the number of aspects mentioned. The videos were between 12 and 75 seconds long, with the average length of 41.3 seconds ($SD = 16.65$). Independent samples t-test did not find a significant difference between the length of truthful and deceitful videos ($t(48) = 1.203, p = 0.235$).

Stimulus validation

The stimulus videos were validated to ensure that all videos were similar in their difficulty to detect their veracity. 20 volunteers (10 male, 10 female) with the mean age of 24.5 years ($SD = 1.76$) participated in the validation process. The validation process took approximately 45 minutes, during which the participants gave their veracity judgment to all 50 stimulus videos. Participation was voluntary and was based on the volunteer's informed consent. Participants had the opportunity to rest mid-way through the process. No feedback about the correctness of their veracity judgments were given.

Participants were not offered any rewards for their participation in the stimulus validation part of the experiment. These volunteers were not able to participate in the experiment as they were familiarised with the stimuli during validation.

Five videos were detected as having the correct/incorrect veracity judgment ratio over 25%:75% (in either direction). However, to keep the balance of the five scenarios and between truthful and deceitful videos, all videos were kept in the experiment. After the experiment, a preliminary analysis was performed before any further analyses, to see whether the inclusion of these five videos significantly affected the results (see Results – Preliminary analysis).

Study design

The current study utilised a within-subjects design and the experiment was divided into three consecutive parts: a pre-training test, a training session, and a post-training test. Receiving feedback during the training was the independent variable in this study. The dependent variables were the participants' truth and lie detection accuracy rates, decision-making time, and the confidence rating.

Procedure

The web-based experiment was created and performed on the Labvanced platform (Finger et al., 2017). Only computers (no tablets or phones) could be used to participate.

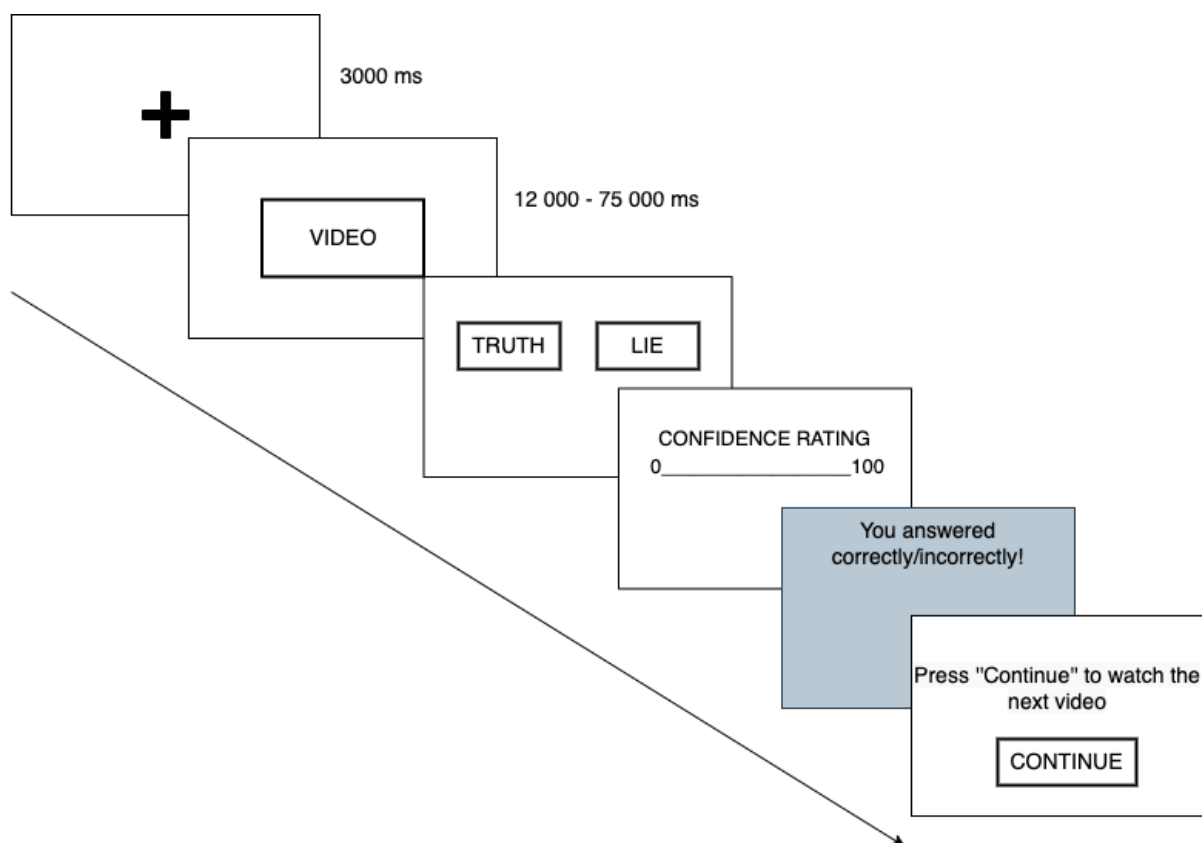
At the beginning of the experiment, participants were asked to sign a consent form and fill in a questionnaire regarding their demographic information (sex and age). Before starting, subjects were given some information pertaining to lying and lie detection. This included information about the general accuracy in deception detection seen in scientific studies. Participants were also told that no reliable verbal or non-verbal cues exist that reliably predict deception. Subjects were asked to make their veracity judgments as quickly as possible and intuitively. This step was included to lessen the possible biasing effects of previous knowledge and attitudes towards lying and lie detection. After the instructions, a

short video was shown, in which the thesis' author explained the experiment and asked the participants to adjust the volume of their speakers/headphones so that the audio is heard clearly. This was done to ensure that the subjects can properly hear the following videos and would not need to change the volume during the experiment.

The experiment was set up so that no video could appear twice for the same subject to avoid repetition of stimuli. Furthermore, subjects could never watch both the true and false story on the same topic told by the same person. This was important to prevent any chance of comparison between the truths and lies. Videos were randomised for all subjects. Experiment trial setup is shown in Figure 2.

Figure 2

Illustration of the experiment setup.



Note. This figure illustrates the general setup of the experiment. The frame in grey represents the feedback that subjects received during the training session of the experiment. Only frames in white were included in the pre- and post-training parts of the test during which participants did not receive any feedback.

Pre- and post-training tests.

Pre- and post-training tests were identical in structure. Each video in the pre- and post-training test was preceded by a fixation cross for 3000 milliseconds, after which a randomised stimulus video automatically started. After watching the video, participants had to make their veracity judgment by choosing between two buttons: 'The person in the video told the truth' and 'The person in the video lied'. Participants were asked to make this judgment as quickly as possible. Following this, subjects were also asked to give a confidence judgement for their answer on a scale of 0-100 (0 = 'I am certain my answer is incorrect'; 50 = 'I am uncertain of my answer'; 100 = 'I am certain my answer is correct'). Both the pre- and post-training tests included five videos. After finishing the post-training test, participants were asked whether they knew anyone from the videos. Then, participants were thanked for their contribution and relevant information was displayed if a participant wished to contact the thesis' author.

Training

The training session of the experiment included 15 videos and was identical with the pre- and post-training tests with one addition. In addition to giving veracity judgements and confidence ratings, subjects received feedback on the correctness of their answer for each video. Feedback was presented directly after the true/lie judgment and confidence rating.

Ethical statement

This study was approved by the Research Ethics Committee of University of Tartu (application number 350/T-19). Important ethical aspects specific to this study are discussed below.

One important ethical aspect in the current study concerns the fact that actors shared personal stories on camera. To ensure that this was approached ethically, the topics for the stimulus videos were kept as broad as possible to give the actors control to only share stories that they were comfortable with. The second important ethical consideration is that the actors were asked to tell lies. As lying could be seen as an immoral act, it was important that the actors volunteered and were not pressured to participate in any way. This was achieved by informing volunteers about lying and sharing personal stories before their participation. Furthermore, actors did not receive any compensation (monetary or otherwise) for their participation, meaning that they did not have to act against their moral compass for any gains.

Subjects participating in the experiment and watching the stimulus material could also be affected by the actors lying to them. Receiving feedback on their truth and lie detection accuracies during the training part of the experiment can be upsetting if the person is more

inaccurate than they thought. This experience could possibly lead to uncomfortable thoughts about their possible inaccuracies in lie/truth detection in their daily lives and personal relationships. To lessen the possible stress from the study, all participants could discontinue the experiment at any time and were given the contacts of the study's author to discuss any questions or problems if they arise.

Data Analysis

Statistical analyses were performed on RStudio version 2021.09.2+382 (RStudio Team, 2022) and JASP version 0.16 (JASP Team, 2021). RStudio was used for data transformation and clean-up as well as for generalized linear mixed model (GLMM) analyses using the lme4 package (Bates et al., 2015) and signal detection analyses using the psycho package (Makowski, 2018). JASP was used for descriptive statistics and paired-samples t-tests. Wilcoxon signed-rank tests were used when data did not meet the assumptions for paired-samples t-tests. The variables used in the analyses were age, gender, detection accuracy, confidence rating and decision-making time. A statistical significance level of .05 was applied in all statistical analyses.

Age was measured in years and gender was presented in three categories: male, female, other. Age and gender were only used for analyses important to describe the sample of this study. Detection accuracy was first coded as ‘0’ for an incorrect veracity judgment and ‘1’ for a correct veracity judgment. The 0/1 formatting was used in GLMM analyses. For other analyses, the average accuracy per subject was calculated and transformed into percentage form. Confidence ratings were given on a 0 – 100 scale, which were explained in more detail above. Decision-making time was the time measured from the start of the veracity judgment frame until the subject made their choice. This was measured automatically during the experiment and is presented in seconds.

Results

Preliminary analysis

Preliminary analysis was carried out to check whether video difficulty could play an important role in the upcoming analyses. As mentioned above, stimulus validation procedure detected five videos for which the correct/incorrect veracity judgment ratio was over 25%:75% (in either direction). To investigate whether removing these five videos significantly affects the effect of training, pre-test-post-test comparisons on overall accuracy were conducted for both datasets (dataset with five videos removed vs dataset with all videos). Paired samples t-test did not find a significant difference between pre- ($M = 51.84$,

$SD = 27.42$) and post-training ($M = 48.38$, $SD = 27.24$) accuracy $t(65) = 0.83$, $p = 0.41$, $d = 0.1$. The pre-test-post-test comparison on overall accuracy was also insignificant for the dataset with all videos (see below). This suggested, that removing the data for five videos did not have a significant impact on the training's effectiveness and thus data from all videos was kept and used for subsequent analyses.

Descriptive statistics

In total, participants made 1615 veracity judgments. On 818 occasions, participants made the judgment that the person in the video was truthful. On 797 occasions, they believed the content of the stimulus video was deceitful. Table 1 represents average overall, truth and lie accuracies as well as confidence ratings and decision-making times. Table 1 was compiled by first calculating average results for every subject across all variables and for every stage of the experiment. The values in the tables represent average results calculated from the subject averages. Subject-based averages were used for t-tests and their non-parametric versions.

For GLMM analysis trial-level data was used. This means, accuracies, confidence ratings and decision-making times were calculated by calculating experiment stage averages straight away, without having calculated subject averages first. This differentiation was important, as by removing some datapoints due to acquaintances in videos for some participants, the pre- and post-training datasets became unequal in number of observations (pre-training $N = 324$, post-training $N = 321$). This is not problematic when describing pre- and post-test results separately (like in GLMMs), however for paired-samples t-tests (and their parametric versions) an equal amount of datapoints is required in the pairs, which was only achievable by calculating subject averages.

Inferential statistics

Accuracy

Pre- and post-training test comparisons were conducted for total accuracy rates as well as for lie detection and truth detection accuracy rates separately. Skewness and kurtosis suggested a non-normal distribution of the differences between pre- and post-training tests for both overall accuracy and truth detection accuracy. Shapiro-Wilk test confirmed violation of normality for overall accuracy ($W = 0.95$, $p = 0.008$) and truth detection accuracy ($W = 0.96$, $p = 0.02$) differences between pre- and post-training tests. Thus, Wilcoxon-signed rank tests were used to analyse these data. Participants overall accuracy did not differ significantly between pre-training ($Mdn = 60$, $SD = 25.98$) and post-training ($Mdn = 40$; $SD = 24.25$) measurements, $W = 899$, $p = 0.098$, $r = 0.26$. Truth detection accuracy was found to be

significantly different in pre- ($Mdn = 66.67$, $SD = 33.75$) and post-training ($Mdn = 33.33$, $SD = 38.24$) parts of the experiment, $W = 1175.5$, $p = 0.005$, $r = 0.42$.

The differences between pre- and post-training tests for lie detection accuracy were normally distributed and thus paired samples t-test was used. No significant difference was found between pre-training ($M = 47.6$, $SD = 33.99$) and post-training ($M = 49.87$, $SD = 34.98$) lie accuracy scores, $t(65) = -0.42$; $p = 0.67$, $d = -0.05$.

Table 1

Descriptive statistics for detection accuracies, confidence rating and decision-making time calculated by the subjects' averages

Stage	Accuracy (%)						Confidence (0-100)						Decision-time (s)		
	Overall			Lie detection			Truth detection								
	M	Mdn	SD	M	Mdn	SD	M	Mdn	SD	M	Mdn	SD	M	Mdn	SD
Pre-training	53.91	60	25.99	47.6	50	33.99	63.26	66.67	33.75	68.36	69.3	12.54	5.54	4.32	3.95
Training	51.96	53.33	14.99	50.48	50	21.22	50.89	45.45	24.11	63.05	62.47	13.27	4.5	3.79	3.25
Post-training	48.81	40	24.25	49.87	50	34.98	44.57	33.33	38.24	63.86	62	12	4.84	3.22	4.59
Total	51.56	52	10.96	50.18	50	15.9	51.55	50	18.33	64.37	63.78	11.88	4.78	4	2.65

Note. Values in this table were calculated by first calculating subject-based averages, which were then averaged once more.

Signal detection analyses.

Signal detection analyses were also conducted on accuracy scores for a more detailed and informative approach to deception detection research. As the focus of the current study was on deception detection, hit, miss, false alarm and correct rejection rates were assigned like in Reinhard et al. (2011) and Porter et al. (2010) (see Table 2).

Table 2

The differentiation of values for signal detection analysis based on the true value of the stimulus and the subject's veracity judgment.

True value of stimulus	Veracity judgment	
	Lie	Truth
Lie	Hit	Miss
Truth	False alarm	Correct rejection

Overall discrimination accuracy (d') and response bias (β) were calculated. As explained by Porter and colleagues (2010), a positive d' indicates that the subject is good at distinguishing between truths and lies, whereas a negative d' shows that the subject misjudges deceptive stimulus as truthful and vice versa. A d' value of zero indicates chance level performance. A positive β signifies the participants tendency to judge the stimulus as deceptive, while a negative β indicates a response bias of judging the stimulus as not deceptive.

In the current study, d' was close to 0 ($M = 0.053$; $SD = 0.590$), suggesting poor discrimination accuracy. A paired samples t-test did not show a significant difference between pre-training d' ($M = 0.183$, $SD = 1.069$) and post-training d' ($M = -0.081$, $SD = 0.988$), $t(65) = 1.611$, $p = 0.112$. Total β demonstrated a positive average of 1.086 ($SD = 0.475$). No significant difference between pre-training β ($M = 1.064$, $SD = 0.339$) and post-training β ($M = 1.117$, $SD = 0.382$) was found, $t(65) = -0.817$, $p = 0.417$.

Confidence

Due to violation of normality for the differences in confidence ratings between pre- and post-training tests, Wilcoxon signed-rank tests were used to analyse them. Overall confidence ratings were shown to have a significant difference between pre- ($Mdn = 69.3$ $SD = 12.54$) and post-training ($Mdn = 63.86$, $SD = 12$) parts of the experiment. $W = 1527.5$, $p = 0.001$, $r = 0.47$. Average confidence rating for correct veracity judgments was 64.94, ($SD =$

18.44) and 63.64 ($SD = 18.24$) for incorrect veracity judgments. A paired-samples t-test was conducted to investigate whether confidence ratings changed significantly for correct and incorrect veracity judgments between pre- and post-training tests. A paired-samples t-test found a significant difference in confidence ratings for incorrect veracity judgments between pre- ($M = 69.78$, $SD = 14.43$) and post-training ($M = 62.34$, $SD = 13.44$), $t(53) = 2.72$, $p = 0.009$. Confidence ratings for correct veracity judgments did not differ significantly between pre- ($M = 69.90$, $SD = 14.91$) and post-training ($M = 66.83$, $SD = 12.60$), $t(59) = 1.74$, $p = 0.087$.

Generalized linear mixed model with binomial family and logit function was performed on pre-training and post-training data separately. In these models, the dependent variable was accuracy coded as 0 = incorrect veracity judgment and 1 = correct veracity judgment. The fixed effect variable was the confidence rating in each trial and the random effects grouping factor was the subject ID. Pre-training confidence did not predict overall accuracy in pre-training ($\beta = -0.07$, $z = -1.09$, $p = 0.277$). However, post-training confidence did predict overall accuracy in post-training ($\beta = 0.02$, $z = 2.01$, $p = 0.044$).

Decision-Time

Wilcoxon signed-rank test showed a significant difference in decision-time between pre- ($Mdn = 4.32$, $SD = 3.95$) and post-training ($Mdn = 3.22$, $SD = 4.59$) parts of the experiment $W = 1675$, $p < 0.001$, $r = 0.52$. Average decision-time for correct veracity judgments was 4.6 seconds ($SD = 7.25$) and 5.03 seconds ($SD = 12.83$) for incorrect veracity judgments. To investigate changes in decision-time for correct and incorrect veracity judgments before and after training a Wilcoxon signed-rank test was conducted as violation of normality was shown for pre-and post-training decision-time differences. Wilcoxon signed-rank test found a significant difference in pre- ($Mdn = 3.74$; $SD = 4.36$) and post-training ($Mdn = 2.75$, $SD = 7.91$) decision-making time for correct veracity judgments, $W = 1278$, $p = 0.008$. Decision-making time did not change significantly for incorrect veracity judgments from pre- ($Mdn = 4.29$; $SD = 4.61$) to post-training ($Mdn = 3.39$, $SD = 5.50$), $W = 951$, $p = 0.073$.

Similarly to confidence ratings, generalized linear mixed model analyses were run on pre- and post-training data where the fixed effect variable was decision-making time in seconds. In these models, the dependent variable was accuracy coded as 0 = incorrect veracity judgment and 1 = correct veracity judgment and the random effects grouping factor was the subject ID. Random intercepts and slopes were allowed to vary for participants. Pre-training decision-making time did not predict overall accuracy in pre-training ($\beta = -0.004$, $z =$

0.23, $p = 0.819$). Post-training decision-making time also did not predict overall accuracy in post-training ($\beta = 0.003$, $z = -0.24$, $p = 0.814$).

Discussion

The thesis at hand investigated the effects of intuitive deception detection training, which aimed to improve detection accuracy by lessening any previous biases about lying and by providing participants feedback on their performance. More specifically, the study investigated how training deception detection affects both lie and truth detection accuracy as well as confidence ratings and decision-making time. Each of the results are discussed below.

Effect of training on detection accuracy

The results showed that subject's overall mean base rate accuracy (in this context, base rate accuracy is accuracy in pre-training) was approximately 54%, which means subjects in the current study were detecting truths and lies is almost at chance level. This result corresponds with previous findings in the literature (Bond & DePaulo, 2006), suggesting that the current sample and results are comparable to other studies in the deception detection field.

As the current study investigated training deception detection, it was hypothesised (H1) that participants' lie detection accuracy will improve with training, meaning it was expected that subjects' will be more accurate in their veracity judgments when watching a deceitful stimulus. The results showed that lie detection accuracy did see a slight improvement from pre-training to post-training, however the increase was not statistically significant. Thus, hypothesis H1 was not supported by the data. This result does not support the previously mentioned theory (Porter et al., 2000), that receiving feedback is essential for improving deception detection as in the current study, feedback-based training did not significantly improve subjects' accuracy.

The thesis at hand also posed a research question (RQ1) which aimed to understand what effect training has on truth detection. It was found that truth detection accuracy decreased over the course of the training and this change was significant. This result is especially interesting in the context of our finding that on average, participants detected truths more accurately ($M = 63.26\%$) than lies ($M = 47.65\%$) at base rate, however, the opposite was found after training. This suggests that participants possibly possessed a truth-bias at base-rate, causing a veracity effect, which was then lost with training. This offers some support the previously mentioned Truth-Default Theory (Levine, 2014) which states that people generally hold a default assumption of others being honest. The decrease seen in truth deception accuracy may have been caused by receiving feedback. Although no official

feedback was asked from participants, some subjects did contact the author of the current study to mention that receiving feedback during the training part of the experiment was sometimes unsettling as in it made participants more doubtful of the stories told in the videos and their answers. This could explain why truth accuracy declined as participants started to lose their default assumption of truthfulness and became more critical of the stimulus material. Overall, these results demonstrate that the current deception detection training was not effective in improving participants lie or truth detection rates. The failure of the current experiment in improving detection accuracy can be caused by many factors from the length of the training to subjects' motivation.

Compared to previous 'brief' training programs in deception detection research, the current experiment was the shortest with only 15 videos in the training stage, which took approximately 20-25 minutes to complete. For example, Porter et al., (2010) included a three-hour workshop as part of their training. Although the authors do not specify how much of that time was spent doing feedback exercises (similar to the training part of the current study) and how much time was dedicated to information provision and myth dissolution, it is likely that subjects spent longer on watching videos than in the current study. Participants' overall accuracy increased significantly between pre- and post-training tests (Porter et al., 2010). Hence, it is possible, that the training part of the current experiment was too short to give participants the necessary amount of experience to improve truth/lie detection. However, as seen above, lie detection accuracy saw a 2.27% increase from pre- to post-training. This outcome in conjunction with Porter et al. (2010) results hints on the possibility that longer training in the current study could have increased accuracy more. Investigation into longer feedback training programs could help us identify whether the lack of improvement in accuracy is more to do with the content or the length of the training. Long experiments, however, can be very tiring for the participants, which can seriously affect their motivation.

As mentioned above, participants' motivation may play a role in their detection accuracy. Motivation is a multi-level problem in the current study, as arguably both the sender's motivation to deceive and the receiver's motivation to detect can affect detection accuracy (Hauch et al., 2016). The current study did not motivate senders in any way and receivers were motivated with a lottery for monetary gain and participation credits, which might have not been enough to induce prolonged motivation throughout the experiment. However, stronger enforcement of motivation by different means was considered unethical by the thesis' author and was thus not used. Low motivation can cause giving random veracity judgments or not watching the stimulus material at all, which could make results

meaningless. However, research has shown that too high motivation can cause motivational impairment effect, leading to participants paying careful attention to stereotypical deception cues that may lead to a misjudgement of deception (Forrest & Feldman, 2000). As it can be difficult to motivate all participants ethically and moderately (not too little and not too much) when it comes to the field of deception, it would be more beneficial to ask participants about their motivation in different parts of the experiment (e.g. on a Likert scale) in future research. This would allow us to see, if and how different levels of motivation affect deception detection accuracy.

Effect of training on discrimination accuracy and response bias

To investigate detection accuracies in more detail, supplementary analyses were conducted to calculate the participants' discrimination accuracy and response bias. Participants demonstrated poor discrimination accuracy throughout the experiment, meaning they were not good at distinguishing between truths and lies. Discrimination accuracy did not change significantly from pre-test to post-test. Although participants detected truths more accurately during post-training, response bias analyses showed that overall participants had the tendency to judge the stimulus as deceptive throughout the experiment. This response bias did not change significantly from pre- to post-training. These results further show that the currently used feedback-based training on intuitive detection was not successful in improving participants' detection accuracy.

Effect of training on confidence and its relation to accuracy

Based on previous literature, it was hypothesised (H2) that confidence ratings will not be predictive of detection accuracy (DePaulo et al., 1997; Leach et al., 2004). Data analysis showed that confidence did not predict accuracy in the pre-training stage of the experiment. However, confidence did predict overall accuracy in the post-training test. This means hypothesis H2 found partial support as confidence was not predictive at base rate but was after the training. Participants overall confidence ratings decreased significantly between pre- and post-training. A closer investigation revealed that confidence decreased significantly for incorrect veracity judgments. Confidence saw a decrease for correct veracity judgments as well, but this change was not significant. Although no official feedback was asked from participants, some subjects did contact the author of the current study to mention that receiving feedback during the training part of the experiment was sometimes unsettling as in it made participants more doubtful of their answers. This explains why confidence decreased by the end of the experiment.

When considering these findings together, they suggest that training did not increase detection accuracy but made participants more accurate in their estimates of their confidence. In simple terms, during the post-training test, participants were more accurate when choosing an estimate on the “I am sure my answer is incorrect” and “I am sure my answer is correct” scale. This means, that the training made participants more careful in their confidence judgments and modified their confidence from liberal to more conservative.

Effect of training on decision-making time and its relation to accuracy

The study at hand investigated decision-making time exploratively, asking whether decision-making time predicts detection accuracy. Statistical analysis showed that although average decision-making time saw a decrease from pre- to post-training, it was not predictive of deception accuracy in neither part of the experiment. Statistical analysis showed that decision-time between pre- and post-training changed significantly. On further investigation, it was found that decision-making time decreased significantly for correct veracity judgments from pre- to post-training tests. The change was not significant for incorrect veracity judgments.

Shorter decision-making times do not allow for deep and conscious processing of incoming information, thus forcing the person to make the judgment based on their first impression, or in other words, intuitively (Zhou et al., 2013). If we recognise shorter decision-times as a reflection of more intuitive processing, it is evident that intuition does not benefit veracity judgments in the current experiment, as participants’ accuracy did not improve, but, in fact, worsened. Accuracy was poor throughout the experiment and especially in the post-training part, where decision-making times were shorter. This suggests that either intuitive decision-making in the truth/lie detection context is not an optimal approach or that decision-making time is not the main factor to reflect and affect intuitive processing. More research is needed on the topic to unravel the relationship between decision-making time, intuitive processing, and deception detection.

The decrease in decision-making time can also be explained separately from accuracy or intuitive processing with the idea that subjects become more tired and unmotivated by the end of the experiment. As the experiment lasted for 30-40 minutes and was repetitive in its nature, it is entirely possible, that participants gave their veracity judgments impatiently and with little motivation to answer correctly. In addition to the length of the experiment, negative feedback on accuracy may have further decreased their motivation and confidence, leading to more impatient judgments.

Limitations and future research

Research into deception and deception training is facing a problem when it comes to the ecological validity of these studies. This concerns the fact experiments often include passive truth/lie detection, which has been a point of critique in many deception detection studies (DePaulo et al., 2003). In this context, passive detection means making a veracity judgment without any communication with the sender. In the current study, receivers only viewed pre-filmed audio-visual stimulus material without any possibility to interact with the senders. Although such a setting may be common in everyday life, where most of consumed media is in similar video form, passive lie detection is not ordinary in the criminal context, where interrogations are conducted in real time and are interactive in nature. However, creating an experiment with interactions is a challenge, as it is difficult to find a valid way to control the interactions to see what effect communication has on the accuracy. Nonetheless, this limitation does not render the current study useless, as with many discussions around fake news and false information sharing, finding ways to improve non-interactive video-based truth/lie detection accuracy could become very beneficial.

Research application and conclusions

The current sample size ($N=66$) is considered small-to-moderate in social sciences, creating a question whether the results of the current study are generalisable to a wider population. However, Levine et al. (2021) argue that the number of total judgments made during a deception detection study is more important than the sample size. Levine and colleagues (2022) advise for at least 500 total judgments, which the current study exceeds by 1115. Thus, it can be argued that the current results are generalisable in deception detection research in terms of numbers, but it should not be forgotten that it only consists of Estonian participants. Furthermore, the current study is strongly imbalanced in terms of gender, so that analysis on gender differences was deemed unreasonable and thus lessening the results' generalisability. Yet, the current study is valuable for showing that deception detection in an Estonian sample with good age variance is similar to findings in other parts of the world, suggesting this accuracy rate may be universal. Furthermore, the current training set-up is one of its kind and thus an important addition to the current literature by showing that intuitive processing may not be superior to other types of detection techniques. A future suggestion for similar research in Estonia is to collect information about participants vocations. Specifically, it would be interesting and possibly beneficial to investigate detection accuracies in Estonia's law enforcement compared to other professions. However, in relation to law enforcement, it is important to note that studies on intuitive deception detection are unpractical as "intuition"

should never be used as an argument in the criminal justice context. Nonetheless, the current study is an important addition to the new area of intuitive deception detection and to the field of experimental psychology.

In conclusion, the current study investigated the effects of a brief feedback-based training on intuitive deception detection. Results demonstrated that training did not improve deception detection significantly, but truth deception accuracy decreased significantly with training. This means training had an overall negative effect, making people less accurate in detection. Also, the current training was successful in adjusting participants' confidence ratings more accurately to their overall detection accuracy. The study shows what effect intuitive processing can have on deception detection and related variables, and makes recommendations for future research in the same field.

References

- Albrechtsen, J. S., Meissner, C. A., & Susa, K. J. (2009). Can intuition improve deception detection performance?. *Journal of Experimental Social Psychology*, 45(4), 1052-1055. and *Social Psychology Review*, 10(3), 214-234.
doi:10.1207/s15327957pspr1003_2
- American Psychological Association. (n.d.). Intuition. In *APA dictionary of psychology*. Retrieved April 30, 2022, from <https://dictionary.apa.org/intuition>
- Bates D, Mächler M, Bolker B, Walker S (2015). “Fitting Linear Mixed-Effects Models Using lme4.” *Journal of Statistical Software*, 67(1), 1–48. doi: 10.18637/jss.v067.i01.
- Biziou-van-Pol, L., Haenen, J., Novaro, A., Occhipinti Liberman, A., & Capraro, V. (2015). Does telling white lies signal pro-social preferences?. *Judgment and Decision Making*, 10, 538-548. doi:10.2139/ssrn.2617668
- Bogaard, G., Meijer, E. H., Vrij, A., & Merckelbach, H. (2016). Strong, but wrong: Lay people’s and police officers’ beliefs about verbal and nonverbal cues to deception. *PloS one*, 11(6), e0156615.
- Bond, J., Charles F, & DePaulo, B. M. (2006). Accuracy of deception judgments. *Personality and Social Psychology Review*, 10(3), 214-234. doi:10.1207/s15327957pspr1003_2
- Cantarero, K., Van Tilburg, W. A., & Szarota, P. (2018). Differentiating everyday lies: A typology of lies based on beneficiary and motivation. *Personality and Individual Differences*, 134, 252-260. doi:10.1016/j.paid.2018.05.013
- Capraro, V., Schulz, J., & Rand, D. G. (2019). Time pressure and honesty in a deception game. *Journal of Behavioral and Experimental Economics*, 79, 93-99.
doi: 10.1016/j.soc.2019.01.007
- Code of Criminal Procedure. (2022).
<https://www.riigiteataja.ee/en/eli/527122021006/consolide>. Accessed 16.05.2022
- De Turck, M. A., & Miller, G. R. (1990). Training observers to detect deception: Effects of self-monitoring and rehearsal. *Human Communication Research*, 16(4), 603-620.
doi:10.1111/j.1468-2958.1990.tb00224.x
- DePaulo, B. M., Charlton, K., Cooper, H., Lindsay, J. J., & Muhlenbruck, L. (1997). The Accuracy-Confidence Correlation in the Detection of Deception. *Personality and Social Psychology Review*, 1(4), 346–357. doi:10.1207/s15327957pspr0104_5

- DePaulo, B. M., 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. doi: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. (1997). Lying and Deception. In N.L. Stein, P.A. Ornstein, B. Tversky & C. Brainerd (Eds.), *Memory for Everyday and Emotional Events* (pp. 333-349). Lawrence Erlbaum Associates, Inc.
- Ekman, P., O'Sullivan, M., & Frank, M. G. (1999). A Few Can Catch a Liar. *Psychological Science*, 10(3), 263–266. <https://doi.org/10.1111/1467-9280.00147>
- Elaad, E. (2009). Lie-detection biases among male police interrogators, prisoners, and laypersons. *Psychological Reports*, 105(3_suppl), 1047-1056.
- Erat, S., & Gneezy, U. (2012). *White Lies*. *Management Science*, 58(4), 723–733. doi:10.1287/mnsc.1110.1449
- Finger, H., Goeke, C., Diekamp, D., Standvoß, K., & König, P. (2017). LabVanced: a unified JavaScript framework for online studies. In International Conference on Computational Social Science (Cologne)
- Forrest, J. A., & Feldman, R. S. (2000). Detecting Deception and Judge's Involvement: Lower Task Involvement Leads to Better Lie Detection. *Personality and Social Psychology Bulletin*, 26(1), 118–125. <https://doi.org/10.1177/0146167200261011>
- Franz, V. H., & von Luxburg, U. (2015). No evidence for unconscious lie detection: A significant difference does not imply accurate classification. *Psychological science*, 26(10), 1646-1648.
- George, J.F., Biros, D.P., Burgoon, J.K., Nunamaker, J.F. (2003). Training Professionals to Detect Deception. In: Chen, H., Miranda, R., Zeng, D.D., Demchak, C., Schroeder, J., Madhusudan, T. (eds) *Intelligence and Security Informatics*. ISI 2003. Lecture Notes in Computer Science, vol 2665. Springer, Berlin, Heidelberg. https://doi.org/10.1007/3-540-44853-5_31
- Gillett, R. and Lee, S., 2019. *You can tell someone's lying to you by watching their face — here are 12 dead giveaways*. [online] Business Insider. Available at: <<https://www.businessinsider.com/how-to-tell-someones-lying-by-watching-their-face-2016-1>> [Accessed 28 April 2022].

- Hartwig, M., & Bond, C. F., Jr. (2011). Why do lie-catchers fail? A lens model meta-analysis of human lie judgments. *Psychological Bulletin*, 137(4), 643–659. <https://doi.org/10.1037/a0023589>
- Hartwig, M., Granhag, P. A., Strömwall, L. A., & Andersson, L. O. (2004). SUSPICIOUS MINDS: CRIMINALS' ABILITY TO DETECT DECEPTION. *Psychology, Crime and Law*, 10(1), 83-95.
- Hauch, V., Sporer, S., Michael, S., & Meissner, C. (2016). Does Training Improve the Detection of Deception? A Meta-Analysis. *Communication Research*, 43(3), 283-343. doi: 10.1177/0093650214534974
<https://books.google.ee/books?id=m28eAgAAQBAJ&lpg=PA333&dq=lying%2C%20deception&lr&pg=PP1#v=twopage&q&f=false>
- JASP Team (2021). JASP (Version 0.16) [Computer software]
- Leach, A.-M., Talwar, V., Lee, K., Bala, N., & Lindsay, R. C. L. (2004). “Intuitive” Lie Detection of Children’s Deception by Law Enforcement Officials and University Students. *Law and Human Behavior*, 28(6), 661–685. doi:10.1007/s10979-004-0793-0
- Levine, T. R. (2014). Truth-default theory (TDT) a theory of human deception and deception detection. *Journal of Language and Social Psychology*, 33(4), 378-392.
- 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., Daiku, Y., & Masip, J. (2022). The Number of Senders and Total Judgments Matter More Than Sample Size in Deception-Detection Experiments. *Perspectives on Psychological Science*, 17(1), 191–204. <https://doi.org/10.1177/1745691621990369>
- Levine, T. R., Feeley, T. H., McCornack, S. A., Hughes, M., & Harms, C. M. (2005). Testing the effects of nonverbal behavior training on accuracy in deception detection with the inclusion of a bogus training control group. *Western Journal of Communication*, 69(3), 203-217. <https://doi.org/10.1080/10570310500202355>
- 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.
- Lohse, T., & Qari, S. (2021). Gender differences in face-to-face deceptive behavior. *Journal of Economic Behavior & Organization*, 187, 1-15. <https://doi.org/10.1016/j.jebo.2021.03.026>

- Makowski D (2018). "The Psycho Package: An Efficient and Publishing-Oriented Workflow for Psychological Science." *Journal of Open Source Software*, 3(22), 470.
doi: 10.21105/joss.00470, R package, <https://github.com/neuropsychology/psycho.R>.
- Masip, J., & Herrero, C. (2017). Examining police officers' response bias in judging veracity. *Psicothema*, 29(4), 490-495.
- Masip, J., Alonso, H., Garrido, E., & Herrero, C. (2009). Training to detect what? The biasing effects of training on veracity judgments. *Applied Cognitive Psychology*, 23, 1282–1296. doi:10.1002/acp.1535
- McCornack, S. A., & Levine, T. R. (1990). When lovers become leery: The relationship between suspicion and accuracy in detecting deception. *Communication Monographs*, 57(3), 219–230.
- Meissner, C. A., & Kassin, S. M. (2002). "He's guilty!": Investigator bias in judgments of truth and deception. *Law and human behavior*, 26(5), 469-480. *Psychology*, 34(2), 138-157. <https://doi.org/10.1023/A:1020278620751>
- Miller, G. R., & Stiff, J. B. (1993). *Deceptive communication*. Sage Publications, Inc.
- O'Sullivan, M. (2003). The Fundamental Attribution Error in Detecting Deception: The Boy-Who-Cried-Wolf Effect. *Personality and Social Psychology Bulletin*, 29(10), 1316–1327. doi:10.1177/0146167203254610
- Porter, S., Juodis, M., ten Brinke, L. M., Klein, R., & Wilson, K. (2010). Evaluation of the effectiveness of a brief deception detection training program. *Journal of Forensic Psychiatry & Psychology*, 21(1), 66–76. doi:10.1080/14789940903174246
- Porter, S., Woodworth, M., & Birt, A. R. (2000). Truth, lies, and videotape: An investigation of the ability of federal parole officers to detect deception. *Law and human behavior*, 24(6), 643-658. <https://doi.org/10.1023/A:1005500219657>
- Reinhard, M.-A., Greifeneder, R., & Scharmach, M. (2013). Unconscious processes improve lie detection. *Journal of Personality and Social Psychology*, 105(5), 721–739. <https://doi.org/10.1037/a0034352>
- RStudio Team (2022). RStudio: Integrated Development Environment for R. RStudio, PBC, Boston, MA, URL <http://www.rstudio.com/>.
- Schindler, S., Wagner, L. K., Reinhard, M. A., Ruhara, N., Pfattheicher, S., & Nitschke, J. (2021). Are criminals better lie detectors? Investigating offenders' abilities in the context of deception detection. *Applied Cognitive Psychology*, 35(1), 203-214.
- Sweeney, C. D., & Ceci, S. J. (2014). *Deception detection, transmission, and modality in age and sex*. *Frontiers in Psychology*, 5. doi:10.3389/fpsyg.2014.00590

- ten Brinke, L., Stimson, D., & Carney, D. R. (2014). Some Evidence for Unconscious Lie Detection. *Psychological Science*, 25(5), 1098–1105.
<https://doi.org/10.1177/0956797614524421>
- Vrij, A. (2008). *Detecting lies and deceit: Pitfalls and opportunities* (pp. 149). John Wiley & Sons.
<https://books.google.ee/books?id=20pg76wmAucC&lpg=PR7&ots=wgqm9iCijg&lr&pg=PA149#v=onepage&q&f=false>
- Vrij, A., & Semin, G. R. (1996). Lie experts' beliefs about nonverbal indicators of deception. *Journal of Nonverbal Behavior*, 20(1), 65–80. doi:10.1007/bf02248715
- Whelan, C. W., Wagstaff, G., & Wheatcroft, J. M. (2015). High stakes lies: Police and non-police accuracy in detecting deception. *Psychology, Crime & Law*, 21(2), 127–138.
<https://doi.org/10.1080/1068316X.2014.935777>
- 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>

Non-exclusive licence to reproduce the thesis and make the thesis public

I, Krõõt Soomets,

1. grant the University of Tartu a free permit (non-exclusive licence) to reproduce, for the purpose of preservation, including for adding to the DSpace digital archives until the expiry of the term of copyright, my thesis

“The effect of feedback-based training on intuitive deception detection accuracy “

supervised by Annegrete Palu.

2. I grant the University of Tartu a permit to make the thesis specified in point 1 available to the public via the web environment of the University of Tartu, including via the DSpace digital archives, under the Creative Commons licence CC BY NC ND 4.0, which allows, by giving appropriate credit to the author, to reproduce, distribute the work and communicate it to the public, and prohibits the creation of derivative works and any commercial use of the work until the expiry of the term of copyright.
3. I am aware of the fact that the author retains the rights specified in points 1 and 2.
4. I confirm that granting the non-exclusive licence does not infringe other persons' intellectual property rights or rights arising from the personal data protection legislation.

Krõõt Soomets

16/05/2022