

**Factors affecting Observers' Accuracy when Assessing Credibility:
The Effect of the Interaction between Media, Senders' Competence and
Veracity**

Letizia Caso¹, Fridanna Maricchiolo², Stefano Livi³, Aldert Vrij⁴, and Nicola Palena¹

¹*Università degli Studi di Bergamo (Italy)*

²*Università degli Studi Roma Tre (Italy)*

³*Sapienza Università di Roma (Italy)*

⁴*University of Portsmouth (UK)*

Correspondence concerning this article should be addressed to Letizia Caso. Università degli Studi di Bergamo. Scienze Umane e Sociali. 24129 Bergamo Lombardia (Italy). E-mail: letizia.caso@unibg.it

Factors affecting observers' accuracy when assessing credibility: the effect of the interaction between media, senders' competence and veracity

Abstract

The present experiment examined how the interaction between senders' communicative competence, veracity and the medium through which judgments were made affected observers' accuracy. Stimuli were obtained from a previous study. Observers ($N = 220$) judged the truthfulness of statements provided by a good truth teller, a good liar, a bad truth teller, and a bad liar presented either via an audio-only, video-only, audio-video, or transcript format. Log-linear analyses showed that the data were best explained via the saturated model, therefore indicating that all the four variables interacted ($G^2(0) = 0, p = 1, Q^2 = 1$). Follow-up analyses showed that the good liar and bad liar were best evaluated via the transcript ($Z = 2.5$) and the audio-only medium ($Z = 3.9$), respectively. Both the good truth teller and the bad truth teller were best assessed through the audio-video medium ($Z = 2.1$, good truth teller, $Z = 3.4$, bad truth teller). Results indicated that all the factors interacted and played a joint role on observers' accuracy. Difficulties and suggestions for choosing the right medium are presented.

Keywords: Detecting Deception, accuracy, medium, communication

Introduction

Two important aspects of deception research are the study of objective (DePaulo, et al., 2003; Vrij, 2008a) and subjective (Hartwig & Bond 2011; Hartwig & Granhag, 2015; Vrij, 2008a) cues to deception. Objective cues refer to which verbal and nonverbal cues are actually related to the act of lying (DePaulo et al., 2003; Hartwig & Bond, 2011; Vrij, 2008a), whereas subjective cues refer to which cues observers direct their attention to when making veracity decisions¹ (Hartwig & Bond, 2011).

Research has shown that people's accuracy in distinguishing truth tellers from liars is close to chance-level (Bond & DePaulo, 2006, 2008) and that such accuracy is generally consistent across laypeople and professionals (Aamodt & Custer, 2006; Bond & DePaulo, 2006; Vrij & Mann, 2001), with some exceptions (Ekman, O' Sullivan, & Frank, 1999; Mann, Vrij, & Bull, 2004). Some have argued that the low accuracy rate is due to wrong beliefs held by observers (Strömwall, Granhag, & Hartwig, 2004; The Global Deception Research Team, 2006). However, in their meta-analysis, Hartwig and Bond (2011) have found that people actually look for the correct cues, but that the differences between truth tellers and liars regarding these cues are barely noticeable. Moreover, Levine et al. (2011) assumed that the transparency of senders also affects judges' accuracy. With transparency, they refer to how well senders portray their honesty or dishonesty. Transparency is also influenced by amount of "matching" between apparent and actual honesty. In particular, if a sender is honest and appears as honest (or if s/he is lying and appears to be lying) then it is possible to claim that there is a "matching demeanor", as the observable behaviour of the subject is indicative of the actual veracity condition. When a sender appears honest while lying (or when s/he give the impression to be a liar when actually telling the truth), then the subject demeanor mismatches the actual veracity condition ("mismatching condition"). It has

¹ The label "Subjective cues" can be interpreted in two different ways. First, as we reported in the text, it can refer to the cues to which the observer *actually* directs his/her attention when making veracity judgments (Hartwig & Bond, 2011). Second, it can refer to cues to which the observer *believes/reports* s/he is directing his/her attention (Hartwig & Granhag, 2015; Taylor & Hick, 2007; The Global Deception Research Team, 2006).

been shown that higher transparency causes higher accuracy rates in observers (Levine, Shaw, & Shulman, 2010; Levine et al., 2011). A recent study (Levine, 2016) also found that most of the variance for deception detection accuracy is observed in senders' than in observers' characteristics, as well as that there is a lack of evidence for deception-general ability.

The effect of senders' characteristics also emerged in Bond and DePaulo (2008) meta-analysis. The authors looked at differences between people in ability to detect lies, tendency to regard someone as honest (which they labeled "credulity"), how much a sender makes obvious whether s/he is lying or telling the truth ("detectability"), and how much a sender tends to appear as credible ("credibility"), regardless of whether s/he is lying or telling the truth. Meta-analytic results showed that the smallest difference was found in judge's ability, whereas the highest was senders' credibility, which is consistent with Levine and colleagues (2011). Indeed, Bond and DePaulo found that judges' ability difference only reached .80%. On the other hand, the difference in judges' biases reached 5.13%, that of senders' detectability 5.49% and that of senders' credibility 11.58%. Bond and DePaulo's results stress again the importance of senders' characteristics on judges' ability, as found in Levine and colleagues (2011). However, Levine et al. only examined one type of lie (lying about an event), whereas it is known that deceptive messages can differ based on the type of lie being told (Warren, Schertler, & Bull, 2009). For example, people can lie about emotions, opinions, events, or intentions. Therefore, it is important to explore if the match/mismatch manipulation has a different impact when different types of lies are told. An individual may be able to portray an emotion s/he is not really feeling and appear as if s/he is actually experiencing that particular emotion but, at the same time, s/he may be unable to be convincing when lying about an intention. In this example, the person may be labelled as a good liar when lying about emotions but as a bad liar when lying about an intention. Furthermore, sender's competence, type of lie and communication medium may interact,

especially when taking into account that the medium the observer is presented with has an impact on judgment accuracy (Bond & DePaulo, 2006; Bond, Levine, & Hartwig, 2015).

An example of the interaction between medium, senders' competence and type of lie is the work by Frank and Ekman (2004). They explored whether a truthful demeanor, that is appearing truthful regardless of whether one is actually truthful or deceptive, generalizes across different media and whether this competence is related to the type of lie. Such truthful demeanor refers to how credible the sender appears (and therefore how convincing s/he is) regardless of the actual veracity condition. The more the sender shows a truthful demeanor, the more likely s/he will be labelled as a truth teller. Therefore, assessing a person who shows a truthful demeanor when telling the truth will result in a hit, but assessing a person who, although showing such a truthful demeanor, is lying will result in a miss. Frank and Ekman's experiment concerned a mock crime and an opinion statement. The authors evaluated the proportion of observers who judged senders as truthful, which they called perceived senders' believability. Their results showed that observers' ratings of senders' honesty for one scenario correlated with ratings for the other scenario for facial expressions, but not for body language and vocal aspects. That is, what made the observers think that the senders were honest (regardless of whether they actually were) was mostly portrayed via facial expressiveness. Such results indicate that when facial expressiveness is taken into account, the type of lie does not play a main role, as people who appear honest for one type of lie appear honest also for the other type of lie. On the other hand, since the correlation for the remaining media across the two scenarios was not significant, the type of lie may play a role in such cases. For example, a sender who is being judged via the vocal channel may appear as truthful when lying about a past event, but s/he may appear as deceptive when lying about an opinion.

Bond et al. (2015) also explored how different media affected how credible a sender appeared, regardless of their actual veracity condition, across four different channels. The

question therefore was “Does a sender who appear honest in his/her nonverbal demeanor also sound honest in his/her speech?”. The authors accounted for the following four media: i) audio only, ii) video only, iii) audio with video, and iv) transcripts. They found that each channel correlated with each other. That is, when a sender appeared honest via the video-only condition, s/he also sounded honest, or when s/he appeared honest via the audio-video condition his/her written statement also appeared credible, and so on. However, it appeared that the audio only condition correlated with transcripts and audio-video conditions more than the video only condition did. This may be due to the fact that people appear less credible when assessed only by their body language due to lying stereotypes (Vrij, 2008b) and it also relates to Bond and DePaulo's (2006) results, who found that people are more accurate when attending to audio cues than when attending to video cues. All of this can also be rooted to the evidence that verbal cues are more diagnostic than nonverbal cues when assessing credibility (Vrij, 2008a, 2008b).

An important research question is if senders appear as consistently credible across different media as found by Frank and Ekman (2004) and by Bond et al. (2015) when controlling for communication skills. That is, senders' believability may be influenced by both the sender's competence and the medium. In short, Levine et al. (2011) shed light on the fact that observers' accuracy is driven by the transparency of the sender (matching vs. mismatching condition, or sender's believability, which we can relate to the communicative competence) but they did not account for the communication medium. On the contrary, Frank and Ekman found that believability mainly remains constant between different scenarios (type of lie) and Bond et al. found that apparent honesty between different media correlated, but neither of them accounted for the interaction between the medium, observers' accuracy and senders' transparency. It is then logical to explore if the medium of communication, the competence of the sender and his/her veracity status affected the observers' accuracy. The main goal of this study was to explore if such a relationship exists.

Based on the role of transparency previously found by Levine et al. (2011), the importance of both judges' (accuracy and credulity) and senders' (detectability and credibility) reported in Bond and DePaulo (2008) and considering the role of the medium (Bond & DePaulo, 2006; Bond et al., 2015; Frank & Ekman, 2004) our goal was to explore whether the three factors (competence, veracity and medium) interacted and played a joint role on observers' accuracy. If this is the case, then, using a log-linear analysis, the statistical model which best fits the data should be the one that accounts for all the four variables taken together (for a detailed explanation of log-linear modeling in psychology, see Knoke & Burke, 1980; Madsen, 1976).

Method

Participants

A total of 220 university students (31 males, 189 females, 55 for each experimental condition) took part in the experiment. Their average age was $M = 22.9$ years ($SD = 5.9$ years).

Material

The video stimuli used in this experiment were obtained from a previous study (Caso, Maricchiolo, Bonaiuto, Vrij, & Mann, 2006) where 128 participants had to lie or to tell the truth about the possession of a backpack and about a meeting with another person (a confederate who gave them the backpack). In a follow-up study focusing on participants' skills in decoding truthfulness and lies, 16 of the 128 videos were randomly selected (eight truthful and eight deceptive senders). Thirty observers were asked to indicate whether each of the 16 senders portrayed in the stimuli was telling the truth or lying.

For the present study, four videos were selected from the original 16 of the follow-up study, according to two criteria. First, a Veracity factor was introduced, so that two videos showed truth tellers (a good truth teller and a bad truth teller) and two videos showed liars (a good liar and a bad liar). Second, a quality of communication (sender competence) factor was

introduced, resulting in four video stimuli: good truth teller, bad truth teller, good liar and bad liar. To assign the senders to each of the four conditions, we proceeded as follows: the good truth teller is a sender who was (correctly) judged as truth teller by at least 70% of observers in the previous study (high believability, matching condition). The bad truth teller is a sender who was correctly identified by no more than 30% of observers (low believability, mismatching condition). Similarly, a good liar is someone who was correctly identified by no more than 30% of observers (high believability, mismatching condition), whereas the bad liar was judged as a liar by at least 70% of observers (low believability, matching condition). All four senders were about twenty years old and white-Caucasian; three of them were women, the remaining was a man. The video depicted the full body of the senders. Our material therefore differed from Levine et al. (2011) in two ways. First, instead of having only two conditions, either matching or mismatching, we made a distinction on the basis of the interaction between the competence of the sender, either good or bad, and his/her veracity conditions (honest or liar), and therefore had four conditions. Second, we classified the senders based on the proportion of correct answers made by the observers in a previous study, whereas Levine et al. selected the five most often believed and five most often disbelieved senders. The mean duration of each video segment was of $M = 150$ s.

Design and Dependent Variables

All the above conditions were presented in four different media: a transcript, an audio recording, a video without sound and a video with sound. The Veracity and Sender's Competence were within-subjects factors, and Medium was a between-subject factor. Observers' accuracy was the dependent variable and was measured on a dichotomous scale: either accurate or inaccurate. The order of Veracity was counterbalanced, and observers always saw the 'good' example first. This resulted in a 2 (Veracity: truth teller vs. liar) X 2 (Sender's competence: bad vs. good) X 4 (Medium: transcript vs. audio vs. video vs. audio-video) X 2 (Participants' accuracy: accurate vs. inaccurate) contingency table which we used

for log-linear model testing. The log-linear analysis modeling aims at showing which model best fits the data (Bakeman & Robinson, 1994; Gnisci & Bonaiuto, 2003). Contrarily to what is usually expected, the focus here is on the first model that does not show significance, as here the lack of significance indicates that the model can explain the data (Knoke & Burke, 1980). The procedure starts from the saturated model, which accounts for all the variables and their interactions as a whole, and then eliminates the effect of singular variables and interactions in the following steps, in a similar way to what is done in a multiple regression. The first model that shows non-significance is the one that explains the data the best. In our case, the four factors entered in the model were Accuracy [A], Medium [M], Sender's Competence [S] and Veracity [V]. Therefore, the first model that is tested by the log-linear analysis is the saturated model [AMSV]. Then, the effect of each interaction and individual variable is deleted from the analysis. For example, a second step may analyze the effect of the [V] taken alone, plus the interaction between [AMS], etc. When the individual factors are deleted from the model, the analysis tests if it played a significant effect in the model. That is, if [AMSV] fits the data (thus, it is not significant) and after the removal of [A], the [MSV] model no longer fits the data (thus it is significant) and we find that the variable [A] is, on its own, significant, then this means that [A] played a central role in the full [AMSV] model. For model testing, the G^2 is chosen over the χ^2 , as the former is unaffected by sample size and unequal marginal distributions. The effect of each removed variable (or interaction of variables) is tested via the analysis of the ΔG^2 and the ΔQ^2 variables. The latter indicates the amount of variability explained by the selected variable with respect to the base model (Gnisci, 2005).

Procedure

Data were collected at University "La Sapienza" (Rome) and at University of Bergamo. Participants watched the stimulus-material in groups ranging from 27 (La Sapienza) to 28 (Bergamo). Eight different experimental sessions were generated: for each

medium of presentation modality (between-subjects factor), four groups saw the two truth tellers first, whereas the remaining four groups saw the two liars first. Participants were firstly welcomed and explained what they had to do. Then, they were shown the stimuli and asked to give their answer to the question “*Do you think that the interviewee lied?*”. They had to choose between two (dichotomous) options, “yes” or “no”. The participants sat far enough to not be able to influence each other's answer. Additionally, participants were also asked to answer the following question “*How much convincing do you think the interviewee was, on a scale from 1 (not at all) to 9 (completely)?*”. Scores to this answer were used to explore if observers could evaluate senders' competence. Indeed, appearing convincing, therefore truthful, can be seen as being a good (competent) communicator, for both cases: truth telling and lie telling. In case of truth telling, being highly convincing can be seen as being a competent sender as one is able to portray his/her actual veracity condition. In case of lying, being highly convincing can be seen as being a competent sender as one is able to hide the fact that s/he is lying whilst appearing as a truth teller, which is what liars aim to do. Once the participant watched all stimuli, s/he was thanked and debriefed.

Results

Model testing

A 2 (Veracity: truth teller vs. liar) X 2 (Sender's Competence: bad vs. good) X 4 (Medium: transcript vs. audio vs. video vs. audio-video) X 2 (Participants' accuracy: accurate vs. inaccurate) hierarchical log-linear model was carried out. The first three variables were entered as the independent variables, whereas accuracy was entered as the dependent variable. The analysis showed that the model which best explained the data was the saturated model [AMSV], $G^2(0) = 0, p = 1, Q^2 = 1$. Using the backward elimination, the removal of the [AMSV] effect, resulting in the [AMS] [AMV] [ASV] [M] model, yielded a significant result. This means that the [AMSV] model played a main role for the modeling and that the new [AMS] [AMV] [ASV] [M] model no longer fitted the data. As a result, it is possible to

conclude that the four variables interacted with each other and that participants' accuracy was influenced by the three factors (Veracity, Sender's Competence and Medium) taken as a whole (Table 1). A finest exploration of ΔQ^2 values indicates that most of the variability of the base model was explained by the [AMV] ($\Delta Q^2 = .20$) and [A] effects ($\Delta Q^2 = .40$), indicating a main role of accuracy. All relevant parameter estimates are reported in Table 2.

Follow-up exploratory analyses

To examine the relations between all the variables we explored the model in more depth. We inspected the adjusted standardized residuals and kept the value of $z \pm 1.96$ as the cut-off score. Therefore, we had a contingency table for each combination of sender's competence and veracity status: good liar, bad liar, good truth teller, bad truth teller. The rows represent the accuracy outputs, either accurate or inaccurate, whereas columns represent the medium. This resulted in four contingency tables with eight different cells. Any cell with a z -score beyond the ± 1.96 limit indicated an association between presence (or absence) of accuracy and the specific medium.

Insert Table 1 and Table 2 about here

For the good liar (Table 3), the analysis showed that accuracy was associated with the transcript modality ($Z = 2.5$), whereas the lack of accuracy was associated with the audio-only modality ($Z = 2.6$). Twenty-four observers (43%) correctly identified the good liar when provided with the transcript, whereas forty-six (83%) made an inaccurate judgment of good liar when provided with the audio-only stimulus.

Enter Table 3 about here

For the bad liar (Table 4), adjusted standardized residuals showed that the audio-only modality was associated with accuracy ($Z = 3.9$, 40 accurate participants, 72%), whereas the video-only modality was associated with lack of accuracy ($Z = 2.0$, 34 inaccurate participants, 61%).

Enter Table 4 about here

For the good truth teller (Table 5), the adjusted standardized residuals showed a relationship between accuracy and the audio-video modality ($Z = 2.1$, 45 accurate participants, 81%) and an association between lack of accuracy and the video-only modality ($Z = 3.4$, 26 inaccurate participants, 47%).

Enter Table 5 about here

For the bad truth teller (Table 6), the adjusted standardized residuals showed that accuracy was associated with the audio-video modality ($Z = 3.4$, 44 participants, 80%), whereas the lack of accuracy was associated with the video-only modality ($Z = 5.9$, 40 participants, 72%).

Enter Table 6 about here

Additional analyses

Since the results underlined that observers' accuracy was related to both Senders' Competence and Veracity conditions, it is essential to explore if observers can actually detect the communication competence of the Sender. This was done via repeated measures ANOVAs with the four different conditions (good liar, bad liar, good truth teller, bad truth teller) as the factor and with convincing scores as the dependent variable.

The test was significant for all four conditions, audio-video condition, $F(3, 52) = 7.665, p < .001, partial \eta^2 = .30$, transcript, $F(3, 52) = 4.130, p = .01, partial \eta^2 = .19$, audio, $F(3, 52) = 32.239, p < .001, partial \eta^2 = .65$, and video, $F(3, 51) = 4.592, p = .006, partial \eta^2 = .21$. Follow-up t -tests revealed several significant results.

Enter Table 7 about here

For the audio-video condition, the bad liar obtained lower scores than each of the other three senders.

For the transcript condition, the bad truth teller and the good liar obtained higher scores than the good truth teller and the bad liar.

For the audio condition, both truth tellers obtained the highest scores, while the bad

liar obtained the lowest score.

Finally, for the video condition, the bad liar appeared as the most convincing, whereas the bad truth teller appeared as the least convincing. All the differences are reported in detail in Table 7.

Discussion

In this experiment, we tested if senders' competence, their veracity status and the medium of communication affected observers' accuracy. First, we found that the full model, which accounted for all these three variables together, influenced participants' accuracy. In addition, the fact that the highest levels of variability were those of accuracy [A] and accuracy, medium and veracity [AMV] gives further support to the idea that observers' accuracy is related to and depends on all of the three other variables. This implies that observers' accuracy builds on several factors, including senders' competence, veracity status and medium.

We also found that, when evaluating the good liar, participants were more accurate when their judgments were based on transcripts, whereas when evaluating the bad liar, they were more accurate when their judgments were based on audio cues. Being a good liar may be related to being able to control both one's own demeanor (Buller & Burgoon, 1996; Hartwig, Granhag, Strömwall, & Doering, 2010) and verbal content (Hartwig et al., 2010). Yet, research found that senders sometime find it more difficult to control their nonverbal behaviour than their verbal content (Caso, Vrij, Mann, & De Leo, 2006), which goes against our results for the good liar. Therefore, it may be difficult to untangle the reasons why we found the transcript to be the best medium to assess the good liar, yet a good explanation may lie on the fact that verbal cues are more effective than nonverbal cues when it comes to evaluating senders' honesty (Bond & DePaulo, 2006; Vrij, 2008a). In any case, our results have applied value as transcripts is the medium experts typically use during legal proceedings, showing potential ecological validity for such a result. Nonetheless, it is

essential to underline that the good liar was correctly identified by only 43% of observers when they attended a transcript, and that their accuracy decreased even more when they paid attention to the audio medium. In this case, only 17% of participants gave the correct answer. Consequently, whilst it still has limitations, the transcript remains the best option. It is interesting to note that, contrary to what has been found in Bond and DePaulo (2006), the audio medium brought observers to a low accuracy rate when assessing the good liar. This stresses again the importance of the interaction between all the variables (medium, sender's competence and veracity). One explanation why this happened is that the good liar could manage his demeanor and emotions to a better extent than bad liar, resulting in fewer cues to deception in their prosody² (Buller & Burgoon, 1996; Caso, Gnisci, Vrij, & Mann, 2005). In fact, we did find that the bad liar was best assessed through the audio cue, which supports this explanation and can be expected being this cue among the few valid cues to deception found in DePaulo's et al. (2003) meta-analysis. One may wonder why the other channels did not work better. It is possible that a bad liar gets stressed during questioning and because of the fear of being caught (Vrij, 2008a) let higher pitch betray his/her lie. It is indeed true that pitch is also available in the audio-video medium, as well as that other cues to deception (fewer details, for example) may be available in other channels such as the transcript. Yet, it has not to be excluded that the stress was so strong for the interviewee that the pitch became the clearer cue to deception.

Our results also showed that participants' accuracy for both the good truth teller and the bad truth teller was positively associated with the audio-video modality. Although research has shown that paying attention to nonverbal behaviour weakens accuracy, it has also been shown that consistency between the verbal and nonverbal medium is important for correct classifications (DePaulo et al., 2003; Heinrich & Borkenau, 1998). Such consistency can be observed only when both audio and visual channels are available, and this may be the

² Prosody refers to all those speech features that are not related to the content of the statement, such as pitch, intonation, volume, etc.

reason why this was the most effective medium for the good truth teller and the bad truth teller. We did not make any prediction about the effect of sender's competence on observers' accuracy as it is likely that both the good truth teller and the bad truth teller told the story "as it is", which should result in verbal-nonverbal consistency regardless of sender's competence. Similarly, we did not predict that inconsistency between the channels may help lie detection, as visual cues may trick people into believing that someone is telling the truth, also due to the well-known truth bias (Vrij, 2008b). That is, consistency may be more evident than inconsistency, particularly when considering that "perceiving true statements may be accompanied by an intuitive feeling of certainty" (Bond, Levine, & Hartwig, 2015, p.52). Furthermore, our results are also in accordance with the idea that audio-visual cues are related to perceived truthfulness (Bond et al., 2015; Frank & Ekman, 2004). Similarly, it may be that our truth tellers' (both the good truth teller's and the bad truth teller's) efforts to appear credible (IDT, Buller & Burgoon, 1996) were more evident through this medium, likely because they either were automatically, or tried to be as much as possible, consistent in their nonverbal and verbal messages.

The most relevant result here is that it is important to shed light on the importance of clarifying the interaction between both senders (*encoding*) and observers (*decoding*) characteristics. Senders' characteristics include their communicative competence and idiosyncrasies, which may result in specific cues to deception (Vrij, 2016). Observers' characteristics may include their biases, which are the cues they rely on to make their decision and their attitudes toward the sender. In fact, although there is a generalization about which are the most reliable cues to deception (DePaulo et al., 2003) it would be important to account for such interpersonal differences (Vrij, 2016) as this may improve accuracy rates. Different people may show different cues to deception and they may do this via different media according to senders' competence. However, it is not possible, at least to date, to decide which media should be selected for veracity assessment. Further studies are needed in

this direction. This becomes important when considering that cues to deception (DePaulo et al., 2003; Vrij, 2008a) and senders' competence (Levine et al., 2011) taken alone may not be able to give a full picture of the variability in observers' accuracy. One possibility to deal with the issue may be to provide observers with one medium first, then to ask to them to think about the perceived sender's honesty. Subsequently, observers may be provided with additional media and then asked if this reinforces or weakens their certainty about the interviewee's honesty. Asking observers to judge senders repeatedly has already proved effective (Granhag & Strömwall, 2001) and it reduces the truth bias (Street & Masip, 2015). Repeated judgments may become even more effective when using several media. The approach may also stimulate observers to think more in depth about what let them make a specific decision. This would also give them the opportunity to change their decisions, with possible bias reduction. Indeed, it has been found that critical thinking is effective even when a bogus training is provided (for a review of training effects, see Hauch, Sporer, Michael, & Meissner, 2016; Vrij, 2008a). Further research is needed to explore if such an approach works as well as which order of (media) presentation may be the best one. Finally, it is interesting to spend some words on the *convincingness* rates made by the observers. For example, for the audio-video condition, the good liar did not appear more or less convincing than both the good truth teller and the bad truth teller. It is plausible that this happened because the good liar was proficient at appearing truthful, which again stresses the importance of accounting for the impact of senders' transparency on observers' accuracy. On the other hand, some unexpected results appeared. To give an example, in the transcript condition the good liar was rated as more convincing than the bad liar and the good truth teller. Yet, the good liar was best assessed through this medium than via the others.

Limitations

Although this study gives some insights into the impact of competence, veracity, and medium on observers' accuracy, there are some limitations that can be dealt with in future studies.

First, there were only four stimuli, which may reduce the understanding of the processes under investigation. Therefore, future research should try to replicate our result with more and more varied stimuli. Second, we tried to control for confounding variables having all participants coming from the same background (all white-Caucasian students) and all being of the same age. Nonetheless, future studies should also actively control for personal characteristics (gender, age, attractiveness, education, ecc.) to reach a more precise picture. Indeed, although we can be quite confident about our senders' competence manipulation due to our selection criteria, it may be possible that idiosyncratic characteristics played an important role as well, as we had only one sender in each condition. We proceeded as explained also to reduce the complexity of the design and analyses, but it is paramount that future studies explore such a relation.

References

- Aamodt, M. G., & Mitchell, H. (2006). Who can best catch a liar? A meta-analysis of individual differences in detecting deception. *Forensic Examiner, 15*, 6–11.
- Bakeman, R., & Robinson, B. F. (1994). Understanding Log-linear analysis with Ilog: An interactive approach. Hove, UK: LEA.
- Bond Jr, C. F., & DePaulo, B. M. (2006). Accuracy of deception judgments. *Personality and Social Psychology Review, 10*(3), 214-234. doi:10.1207/s15327957pspr1003_2.
- Bond Jr, C. F., & DePaulo, B. M. (2008). Individual differences in judging deception: accuracy and bias. *Psychological Bulletin, 134*(4), 477-492. doi:10.1037/0033-2909.134.4.477.
- Bond, C. F., Levine, T. R., & Hartwig, M. (2015). New Findings in Non-Verbal Lie Detection. In P. A. Granhag, A. Vrij, & B. Verschuere (Eds.), *Deception detection: Current challenges and cognitive approaches* (pp. 37 – 58). Chichester, England: Wiley.
- Buller, D. B., & Burgoon, J. K. (1996). Interpersonal deception theory. *Communication Theory, 6*(3), 203-242. doi:10.1111/j.1468-2885.1996.tb00127.x.
- Caso, L., Gnisci, A., Vrij, A., & Mann, S. (2005). Processes underlying deception: An empirical analysis of truth and lies when manipulating the stakes. *Journal of Investigative Psychology and Offender Profiling, 2*(3), 195-202. doi:10.1002/jip.32.
- Caso, L., Vrij, A., Mann, S., & Leo, G. (2006). Deceptive responses: The impact of verbal and non-verbal countermeasures. *Legal and Criminological Psychology, 11*(1), 99-111. doi:10.1348/135532505X49936
- Caso, L., Maricchiolo, F., Bonaiuto, M., Vrij, A., & Mann, S. (2006). The impact of deception and suspicion on different hand movements. *Journal of Nonverbal Behavior, 30*(1), 1-19. doi:10.1007/s10919-005-0001-z.
- DePaulo, B. M., Lindsay, J. L., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H.

- (2003). Cues to deception. *Psychological Bulletin*, *129*, 74–118. doi:10.1037/0033-2909.129.1.74.
- Ekman, P., O'Sullivan, M., & Frank, M. G. (1999). A few can catch a liar. *Psychological Science*, *10*(3), 263-266. doi:10.1111%2F1467-9280.00147.
- Frank, M. G., & Ekman, P. (2004). Appearing truthful generalizes across different deception situations. *Journal of Personality and Social Psychology*, *86*(3), 486-495. doi:10.1037/0022-3514.86.3.486.
- Gnisci, A. (2005). Sequential strategies of accommodation: A new method in courtroom. *British Journal of Social Psychology*, *44*(4), 621-643. doi:10.1348/014466604X16363.
- Gnisci, A., & Bonaiuto, M. (2003). Grilling Politicians: Politicians' Answers to Questions in Television Interviews and Courtroom Examinations. *Journal of language and social psychology*, *22*(4), 385-413. doi: 10.1177/0261927X03258088.
- Gnisci, A., Caso, L., & Vrij, A. (2010). Have you made up your story? The effect of suspicion and liars' strategies on reality monitoring. *Applied Cognitive Psychology*, *24*(6), 762-773. doi:10.1002/acp.1584.
- Granhag, P. A., & Strömwall, L. A. (2001). Deception detection based on repeated interrogations. *Legal and Criminological Psychology*, *6*(1), 85-101. doi: [10.1348/135532501168217](https://doi.org/10.1348/135532501168217)
- Hartwig, M., & Bond, C. F., Jr. (2011). Why do lie-catchers fail? A lens model meta-analysis of human lie judgments. *Psychological Bulletin*, *137*, 643-659. doi:10.1037/a0023589.
- Hartwig, M., & Granhag, P. A. (2015). Exploring the nature and origin of beliefs about deception: Implicit and explicit knowledge among lay people and presumed experts. In P. A. Granhag, A. Vrij, & B. Verschuere (Eds.), *Deception detection: Current challenges and cognitive approaches* (pp. 125 – 154). Chichester, England: Wiley.

- Hartwig, M., Granhag, P. A., Strömwall, L. A., & Doering, N. (2010). Impression and information management: On the strategic self-regulation of innocent and guilty suspects. *The Open Criminology Journal*, 3(1), 10-16.
doi:10.2174/1874917801003010010.
- Hauch, V., Sporer, S. L., Michael, S. W., & Meissner, C. A. (2016). Does training improve the detection of deception? A meta-analysis. *Communication Research*, 43(3), 283-343. doi:10.1177/0093650214534974.
- Heinrich, C. U., & Borkenau, P. (1998). Deception and deception detection: The role of cross-modal Inconsistency. *Journal of Personality*, 66(5), 687-712. doi:10.1111/1467-6494.00029.
- Knoke, D., & Burke, P. J. (1980). *Log-linear models. Series: Quantitative applications in the social sciences*. Beverly Hills: Sage Publications Inc.
- Levine, T. R. (2010). A few transparent liars: Explaining 54% accuracy in deception detection experiments. *Annals of the International Communication Association*, 34(1), 41-61. doi:10.1080/23808985.2010.11679095.
- Levine, T. R. (2016). Examining sender and judge variability in honesty assessments and deception detection accuracy: Evidence for a transparent liar but no evidence of deception-general ability. *Communication Research Reports*, 33(3), 188-194. doi: [10.1080/08824096.2016.1186629](https://doi.org/10.1080/08824096.2016.1186629)
- Levine, T. R., Serota, K. B., Shulman, H., Clare, D. D., Park, H. S., Shaw, A. S., ... & Lee, J. H. (2011). Sender demeanor: Individual differences in sender believability have a powerful impact on deception detection judgments. *Human Communication Research*, 37(3), 377-403. doi:10.1111/j.1468-2958.2011.01407.x.
- Levine, T. R., Shaw, A., & Shulman, H. C. (2010). Increasing deception detection accuracy with strategic questioning. *Human Communication Research*, 36(2), 216-231. doi:10.1111/j.1468-2958.2010.01374.x.

- Madsen, M. (1976). Statistical analysis of multiple contingency tables. Two examples. *Scandinavian Journal of Statistics*, 3(3), 97-106.
- Mann, S., Vrij, A., & Bull, R. (2004). Detecting true lies: police officers' ability to detect suspects' lies. *Journal of Applied Psychology*, 89(1), 137. doi:10.1037/0021-9010.89.1.137.
- Mann, S. A., Vrij, A., Fisher, R. P., & Robinson, M. (2008). See no lies, hear no lies: Differences in discrimination accuracy and response bias when watching or listening to police suspect interviews. *Applied Cognitive Psychology*, 22(8), 1062-1071. doi:10.1002/acp.1406.
- Street, C. N., & Masip, J. (2015). The source of the truth bias: Heuristic processing?. *Scandinavian Journal of Psychology*, 56(3), 254-263. doi: 10.1111/sjop.12204.
- Strömwall, L., Granhag, P. A., & Hartwig, M. (2004). Practitioners' beliefs about deception. In P. A. Granhag & L. A Strömwall (Eds.), *The detection of deception in forensic contexts* (pp. 229-250). New York, NY: Cambridge University Press.
- Taylor, R., & Hick, R. F. (2007). Believed cues to deception: Judgments in self-generated trivial and serious situations. *Legal and Criminological Psychology*, 12(2), 321-331. doi:10.1348/135532506X116101.
- The Global Deception Research Team (2006). A world of lies. *Journal of Cross Cultural Psychology*, 37(1), pp. 60-74. doi:10.1177/0022022105282295.
- Vrij, A. (2008a). *Detecting lies and deceit: Pitfalls and opportunities*. Chichester: John Wiley and Sons.
- Vrij, A. (2008b). Nonverbal dominance versus verbal accuracy in lie detection a plea to change police practice. *Criminal Justice and Behavior*, 35(10), 1323-1336. doi:10.1177/0093854808321530.
- Vrij, A. (2016). Baseline as a Lie Detection Method. *Applied Cognitive Psychology*, 30(6),

1112-1119. doi:10.1002/acp.3288.

Vrij, A., & Mann, S. (2001). Who killed my relative? Police officers' ability to detect real-life high-stake lies. *Psychology, Crime and Law*, 7(1-4), 119-132.

doi:10.1080/10683160108401791.

Warren, G., Schertler, E., & Bull, P. (2009). Detecting deception from emotional and unemotional cues. *Journal of Nonverbal Behavior*, 33(1), 59-69. doi:10.1007/s10919-008-0057-7.

Table 1. Log-linear Analysis $A \times M \times S \times V$ ($N = 220$).

Model	G^2	df	p	Deleted effect	ΔG^2	df	p	Q^2	ΔQ^2
[AMSV]	.0	0	1.00	-				1	-
[AMS] [AMV] [ASV] [M]	14.1	3	< .01	[AMSV]	14.1	3	< .01	.9	.1
[AMV] [ASV] [MSV]	31.8	6	< .01	[AMS]	17.7	3	< .01	.8	.1
[ASV] [MSV] [AM]	66.3	9	< .01	[AMV]	34.5	3	< .01	.6	.2
[MSV] [AM] [AS] [AV]	67.0	10	< .01	[ASV]	.7	1	.41	.6	.0
[AM] [AS] [AV] [MS] [MV]	67.0	13	< .01	[MSV]	.0	3	1.00	.6	.0
[AS] [AV] [MS] [MV] [SV]	83.4	16	< .01	[AM]	16.4	3	< .01	.5	.1
[AV] [MS] [MV] [SV]	99.5	17	< .01	[AS]	16.1	1	< .01	.4	.1
[MS] [MV] [SV] [A]	101.3	18	< .01	[AV]	1.8	1	.18	.4	.0
[MV] [SV] [A]	101.3	21	< .01	[MS]	.0	3	1.00	.4	.0

[SV] [A] [M]	101.3	24	< .01	[MV]	.0	3	1.00	.4	.0
[A] [M] [S] [V]	101.3	25	< .01	[SV]	.0	1	.95	.4	.0
[M] [S] [V]	168.7	26	< .01	[A]	67.4	1	<.001	.0	.4
[S] [V]	168.7	29	< .01	[M]	.0	3	1.00	.0	.0
[V]	168.7	30	< .01	[S]	.0	1	.95	.0	.0
0	168.7	31	< .01	[V]	.0	1	.95	.0	.0

Table 2. *Parameter estimates for Log-linear analysis*

Effect		Log-Linear	
		Parameter (Lambda)	Estimate Lambda/SE
Fourth-order effect:			
Veracity	by		
Competence	by	.110	1.766
Medium by Accuracy			
Third-order effects:			
Veracity	by		
Competence	by	-.025	-.400
Medium			
Veracity	by		
Competence	by	-.169	-4.477
Accuracy			
Veracity by Medium by			
Accuracy		.059	.941
Competence	by		
Medium by Accuracy		-.058	-.926
Second-order effects:			
Veracity	by		
Competence		.015	.409
Veracity by Medium		-.010	-.159
Competence	by		
Medium		.010	.158
Veracity by Accuracy		-.286	-7.586
Competence	by		
Accuracy		.054	1.423
Medium by Accuracy		-.084	-1.346

Table 3. *Contingency table for good liar.*

		Medium				
		Transcript	Audio-Only	Video-Only	Audio-Video	
Was the observer accurate?	No	Count	31	46	40	36
		Expected count	38.3	38.3	38.3	38.3
		Percentage	57%	83%	72%	65%
		Adjusted Residuals	-2.45	2.62	.59	-.76
	Yes	Count	24	9	15	19
		Expected count	16.8	16.8	16.8	16.8
		Percentage	43%	17%	28%	35%
		Adjusted Residuals	2.45	-2.62	-.59	.76

Table 4. *Contingency table for bad liar.*

		Medium				
		Transcript	Audio-Only	Video-Only	Audio-Video	
Was the observer accurate?	No	Count	28	15	34	33
		Expected count	27.5	27.5	27.5	27.5
		Percentage	50%	28%	61%	59%
		Adjusted Residuals	.16	-3.89	2.02	1.71
	Yes	Count	27	40	21	22
		Expected count	27.5	27.5	27.5	27.5
		Percentage	50%	72%	39%	41%
		Adjusted Residuals	-.16	3.89	-2.02	-1.71

Table 5. Contingency table for good truth teller.

		Medium				
		Transcript	Audio-Only	Video-Only	Audio-Video	
Was the observer accurate?	No	Count	16	12	26	10
		Expected count	16	16	16	16
		Percentage	29%	22%	47%	19%
		Adjusted Residuals	.00	-1.37	3.43	-2.06
	Yes	Count	39	43	29	45
		Expected count	39	39	39	39
		Percentage	71%	78%	53%	81%
		Adjusted Residuals	0.00	1.37	-3.43	2.06

Table 6. Contingency table for bad truth teller.

		Medium				
		Transcript	Audio-Only	Video-Only	Audio-Video	
Was the observer accurate?	No	Count	19	16	40	11
		Expected count	21.5	21.5	21.5	21.5
		Percentage	35%	29%	72%	20%
		Adjusted Residuals	-.80	-1.75	5.9	-3.35
	Yes	Count	36	39	15	44
		Expected count	.80	1.75	-5.9	3.35
		Percentage	65%	71%	28%	80%
		Adjusted Residuals	.80	1.75	-5.9	3.35

Table 7. *Observers' evaluation of senders' convincingness.*

Medium condition	Condition Mean (SD)			
	Good truth teller	Bad truth teller	Good Liar	Bad liar
Audio-Video	$M = 5.53^a$ ($SD = 2.11$)	$M = 5.29^a$ ($SD = 2.22$)	$M = 4.95^a$ ($SD = 2.30$)	$M = 3.75^b$ ($SD = 2.07$)
Transcript	$M = 4.49^a$ ($SD = 2.17$)	$M = 5.47^b$ ($SD = 2.13$)	$M = 5.24^b$ ($SD = 1.99$)	$M = 4.36^a$ ($SD = 1.97$)
Audio-Only	$M = 5.93^a$ ($SD = 1.81$)	$M = 5.73^a$ ($SD = 2.16$)	$M = 6.82^b$ ($SD = 1.69$)	$M = 3.62^c$ ($SD = 1.99$)
Video-Only	$M = 4.76^{ab}$ ($SD = 2.22$)	$M = 3.76^c$ ($SD = 1.86$)	$M = 4.09^{bc}$ ($SD = 1.93$)	$M = 4.95^a$ ($SD = 2.08$)

Different superscripts letters indicate significant difference ($p < .05$)