Received: 28 April 2021 | Revised: 1 April 2022 | Accepted: 13 August 2022

DOI: 10.1002/jip.1597

RESEARCH ARTICLE

WILEY

Looking guilty: Handcuffing suspects influences judgements of deception

Mircea Zloteanu[1] | Nadine L. Salman[2] | Eva G. Krumhuber[3] | Daniel C. Richardson[3]

[1]Department of Criminology and Sociology, Kingston University London, London, UK

[2]Department of Security and Crime Science, University College London, London, UK

[3]Department of Experimental Psychology, University College London, London, UK

Correspondence
Mircea Zloteanu, Department of Criminology and Sociology, Kingston University London KT1 2EE, UK.
Email: m.zloteanu@kingston.ac.uk

Abstract

Veracity judgements are important in legal and investigative contexts. However, people are poor judges of deception, often relying on incorrect behavioural cues when these may reflect the situation more than the sender's internal state. We investigated one such situational factor relevant to forensic contexts: handcuffing suspects. Judges—police officers (n = 23) and laypersons (n = 83)—assessed recordings of suspects, providing truthful and deceptive responses in an interrogation setting where half were handcuffed. Handcuffing was predicted to undermine efforts to judge veracity by constraining suspects' gesticulation and by priming stereotypes of criminality. It was found that both laypersons and police officers were worse at detecting deception when judging handcuffed suspects compared to non-handcuffed suspects, while not affecting their judgement bias; police officers were also overconfident in their judgements. The findings suggest that handcuffing can negatively impact veracity judgements, highlighting the need for research on situational factors to better inform forensic practice.

KEYWORDS
accuracy, bias, deception detection, interviewing, police officers, veracity judgements

1 | INTRODUCTION

Deception detection can be crucial in investigative, forensic, and legal contexts where the outcome of a charging decision or criminal trial can hang on the credibility of the victim, witness, or suspect testimony (Horvath et al., 1994). However, veracity judgements are challenging, especially for human judges (Zloteanu, 2020; Zloteanu, Bull, et al., 2021). People tend to be poor detectors of deception (Bond & DePaulo, 2006), biased towards overestimating others' honesty (Levine et al., 1999), and overconfident in their judgements (DePaulo et al., 1997). Given the relevance of veracity judgements in forensic and legal contexts, it is important to examine the role of situational factors in this process. Using an experimental scenario that simulates a common real-life application—handcuffing suspects—we examine how situational factors can affect veracity judgements in both laypersons and police officers.

1.1 | Behavioural cues to deception

Human judges typically make poor veracity judgements (Bond & DePaulo, 2006, 2008), especially when judging lies (Levine et al., 1999, 2014). This lacklustre performance has been, in part, attributed to the lack of reliable behavioural cues (i.e. verbal, paraverbal, and nonverbal) that differentiate liars and truth-tellers (DePaulo et al., 2003; Hartwig & Bond, 2014). Indeed, the relationship between behavioural cues and deception detection is controversial and fraught with contradicting evidence as to the type and direction of effects (Burgoon, 2018; Luke, 2019; Zloteanu, Bull, et al., 2021).

Relevant to our current manipulation, liars are reported to make fewer hand and finger movements, use fewer illustrators, pointing gestures, and self-touching behaviours compared to truth-tellers (Caso et al., 2006; Sporer & Schwandt, 2007; Vrij et al., 2004). This is also demonstrated in motion-tracking research which finds liars and truth-tellers to differ in the use of their arms, torso, and head (Duran et al., 2013). Hence, the restriction of hand gestures may negatively impact the discriminability of liars and truth-tellers, thereby reducing deception detection accuracy. Moreover, regardless of their diagnostic value, people hold strong beliefs about specific behaviours revealing deceit (The Global Deception Research Team, 2006) and often make veracity judgements based on these beliefs (Bogaard et al., 2016; Zloteanu, Bull, et al., 2021). People believe that liars display more self-touching, fidgeting, gaze aversion, and generally move more than truth-tellers because they are anxious and/or nervous, however, these beliefs rarely match reality (Akehurst et al., 1996; Anderson et al., 1999; Masip & Herrero, 2015). Such misinformation is also propagated by police training manuals (Vrij et al., 2010), suggesting that experience with nonverbal cues can make officers better lie detectors (Gudjonsson, 2005).

Furthermore, in the forensic and legal literature, demeanour evidence is considered an important cue for witness credibility (Mack, 2001; Varinsky, 1992). A focus on such behavioural 'cues' can consequently be a source of misleading information (Denault et al., 2020; Denault & Patterson, 2021; Zloteanu, 2020). Thus, if behavioural cues are diagnostic, as so many believe them to be, then we would expect that accuracy would decrease if judges were no longer able to rely on them. But, our claim is not that such cues are useful in (accurately) detecting deception, only that they are influential, and limiting the ability of senders to move and gesticulate by handcuffing will reduce any real or expected nonverbal differences between liars and truth-tellers, affecting judges' veracity judgements[1]

1.2 | Situational factors

To illustrate the complexity of veracity judgements, we bring attention to an often-underrepresented component of the judgement process: the role of situational or contextual elements. The deception detection literature has largely overlooked the impact of situational factors on senders (i.e., liars and truth-tellers) and judges (i.e., individuals making veracity judgements). The situation in which senders find themselves may affect their behaviour and ability to accurately portray themselves, influencing how they act and are perceived.

Situational factors and their effects are here considered separate from the natural variability between senders' demeanour (e.g., physical appearance; Funk & Todorov, 2013) or differences between judges' ability (e.g., training; Meissner & Kassin, 2002), which are innate to any task but can also affect veracity judgements (see also, Zloteanu, Krumhuber, et al., 2021). Instead, we define situational factors as *external elements which influence the judgement process*. For example, manipulating the attire (nurse, military, casual) of individuals can affect judges' ratings of perceived dominance and empathy (Küster et al., 2019), while adding glasses to a portrait photo can increase ratings of intelligence, honesty, dependability, and industriousness (Thornton, 1943). Such effects are explained by the stereotypical associations of situational elements being used to ascribe personal characteristics to individuals.

To outline the effects of such situational[2] factors on veracity judgements, we adapt the decision-making framework by Wieser and Brosch (2012). The framework proposes that a judgement is a product of the encoding stage (i.e., information from the sender), the transmission stage (i.e., presentation of information), and the decoding stage (i.e., receiver perception) and that situational effects play a role at each stage of this process to influence the final judgement (for more detail, see Wieser & Brosch, 2012). Here, we retain two types of influence described in their framework: *within-judge elements* (i.e., differences emerging from variations between receivers' thought processes) and *external features from the environment*; the focus of our paper is on both elements while providing a novel manipulation for the latter type; we note that a clear demarcation of situational effects based on these types is not simple due to potential interactive effects. A directed acyclic graph (DAG) of the theoretical causal model of the elements and factors of interest and the final veracity judgement is presented in the Supplemental Information (SI).

By employing a manipulation that is relevant to forensic settings, we investigate the effect of handcuffing to illustrate how a routine practice can affect the veracity judgement process. Handcuffs are used in many forensic and legal contexts. For example, in the United States arrestees are routinely handcuffed before an interview (Virginia Commonwealth University Police Department, 2014). In the United Kingdom, police officers can handcuff suspects if it is deemed justified (Association of Chief Police Officers of England Wales & Northern Ireland, 2009). Their use in UK courts is also permitted in exceptional circumstances (Courts and Tribunals Judiciary, 2010). Our focus is on how this routine element can affect person perception and veracity judgements.

In one of few studies on situational factors in veracity judgement settings, ten Brinke et al. (2015) coded the behaviour of liars and truth-tellers interviewed in a scarce environment (a plain, undecorated room) and an enriched environment (a room filled with colourful artwork and a desk full of office-related objects). They found more behavioural differences between liars and truth-tellers in the scarce environment, arguing that the setting increased liars' production of nonverbal cues to deception, concluding that this manipulation could be used as a tactic to improve deception detection. However, Verschuere et al. (2016) contested these claims, showing that such environmental manipulations have the same effect on both liar and truth-teller behaviour, resulting in overall poorer deception detection accuracy. Moreover, depriving suspects of resources, as advised by certain police training manuals (Inbau et al., 2011), can increase interviewer suspiciousness, resulting in a stronger tendency to assume that senders are lying (i.e., lie-bias) whilst also increasing the likelihood of eliciting a false confession (Meissner & Kassin, 2002; Toris & DePaulo, 1984; Vrij et al., 2006).

Further examples of situational effects can be found in the related investigative interviewing literature on 'context manipulation', that is, techniques to modify the interview space to maximise the probability of obtaining accurate and reliable information (Kelly et al., 2013). Context manipulation research finds that elements relating to the physical environment in which an interview is conducted, such as lighting, decoration, and room size can have an effect on interpersonal communication, disclosure, and interview outcomes (Chaikin et al., 1976; Dawson et al., 2017, 2017, 2017; Gifford, 1988; Kelly et al., 2013; Knapp et al., 2013). Some even demonstrate how such elements—including physical restraints—can be used strategically to exert or reduce coercive pressure (Goodman-Delahunty & Sivasubramaniam, 2013) and/or build rapport with interviewees (Goodman-Delahunty et al., 2014; Hoogesteyn et al., 2019).

Similar effects of situational factors have also been reported in the counselling literature, where alterations to the environment are found to influence how forthcoming clients are towards their therapist (e.g., Chaikin et al., 1976;

Okken et al., 2012). Importantly, people seem to have strong a priori expectations and preferences for interview environments, with the latter seemingly being more important for perceptions and outcomes (Hoogesteyn et al., 2020). While such research focuses more on disclosure rather than deception detection, an overarching finding is that situational factors can be influential despite often being overlooked when judging others (i.e., the fundamental attribution error; Nisbett & Ross, 1991). This is pertinent given how increased suspiciousness and distrust can be counterproductive to communication and accurate judgements (Burgoon et al., 1996; Toris & DePaulo, 1984). These studies demonstrate that situational factors, such as the interrogation environment, can affect sender behaviour and judge accuracy although not always in the desired fashion.

1.3 | Veracity judgement differences: Police officers versus laypersons

Research using police officers is rare in the deception field. The available data suggest that police officers display the same underwhelming performance as laypersons (Aamodt & Custer, 2006; Vrij & Semin, 1996). This may be because police officers also rely on the same (incorrect) cues to determine deceit, including cues that relate to hand and arm movements (Akehurst et al., 1996; Bogaard et al., 2016; Colwell et al., 2006; Strömwall & Granhag, 2003). However, police officers may differ in the direction and strength of their judgemental biases (Hauch et al., 2014; Vrij, 2008). While laypersons are typically truth-biased by believing that senders are likely to be telling the truth (Levine et al., 1999), studies using police officers find a weaker truth-bias or complete reversal thereof (i.e., lie-bias; Garrido et al., 2004; Masip et al., 2008; Meissner & Kassin, 2002, but see Hurst & Oswald, 2012). This can be attributed to police officers being more suspicious of others (Kim & Levine, 2011) whilst also typically being (over)confident in their abilities (DePaulo & Pfeifer, 1986).

Given police officers' experience with interview scenarios and handcuffs, the handcuffing manipulation may affect their judgements and those of laypersons differently. The findings of the current investigation can be relevant at different stages in the judicial process, namely for police officers involved in the initial investigation and charging decision, and for laypersons who serve as jurors in sentencing decisions. Hence, we considered both police officers and laypersons as target groups in our research.

2 | THE PRESENT STUDY

The current study explores the potentially detrimental effects of situational factors on the veracity judgement process. From an external context perspective, handcuffs are associated with criminality which in turn can influence judges' suspiciousness, resulting in a stronger lie bias in their judgements (Bond et al., 1992; Levine et al., 2000; Levine & McCornack, 1991). Perceived credibility is the strongest factor impacting veracity decisions (George et al., 2014). Thus, if a sender appears less credible due to the presence of an added cue to criminality, they are more likely to be rated as dishonest regardless of their veracity (Burgoon et al., 2008; Levine et al., 2011; Porter & ten Brinke, 2009). Thus, we hypothesised that the presence of handcuffs will bias judgements of suspects, leading to a stronger tendency to assume handcuffed suspects are more dishonest than non-handcuffed suspects (i.e., a reduction or reversal of the truth-bias; H_1).

From a within-judge context perspective, handcuffing may alter the behaviours that judges expect from liars and truth-tellers (Akehurst et al., 1996), affecting their ability to classify veracity. Handcuffing suspects may limit the movements of both sender groups (liars and truth-tellers) making the two appear more similar. Thus, we hypothesised that restricting the movement of suspects by handcuffing will lead to poorer veracity discrimination for handcuffed suspects compared to non-handcuffed suspects (i.e., lower deception detection accuracy; H_2).

From an interactive perspective, the effects of handcuffing could manifest differently in the two groups of judges based on their prior expectations of sender appearance and behaviour. Hence, we predict that situational factors

interact with judge type due to differences in experience with forensic settings. Specifically, laypersons' judgement response bias (e.g., truth-bias) should be more strongly affected by the presence of handcuffs (H_3), while police officers are expected to be more confident (H_4) and lie-biased (H_5) in all their veracity judgements than laypersons (e.g., Garrido et al., 2004).

3 | METHOD

3.1 | Design

A three-way mixed design was employed. The between-subjects factors were handcuffing (handcuffed vs. non-handcuffed) and judge (laypersons vs. police officers). The within-subjects factor was statement veracity (truth and lie). Judges were measured on deception detection accuracy, judgement bias, and confidence. Given the difficulties in recruiting police officers, an a priori power analysis was not conducted. To estimate the minimum detectable effect (MDE), a sensitivity analysis considering the final sample size revealed that effect sizes of Cohen's $f = 0.28$ or partial eta$^2 = 0.07$ (small-to-moderate) could be detected with 80% power and an alpha criterion of 0.05.

4 | PARTICIPANTS

Ninety laypersons were recruited as judges through online advertisement and the university's online subject pool. After screening for incomplete responses ($n = 6$) and for deducing the aim of the study ($n = 1$) the final sample comprised 83 judges (36 males, 47 females, $M_{age} = 24.13$, $SD = 6.93$). Participating students were given course credits; no other incentive was offered. Twenty-three police officers were recruited (17 males, six females, $M_{age} = 31.00$, $SD = 6.69$). Police officers were contacted directly by the experimenters after obtaining approval from the London (UK) Metropolitan Police Service's Research Department. Table 1 provides details on their levels of experience. All aspects of the study had been approved by the ethics committee of the university. Informed consent was obtained from all participants prior to data collection.

4.1 | Suspect interrogation

A semi-scripted interview approach was employed in which a handcuffed "suspect" was interviewed by a police interrogator. Nineteen suspects (9 males, 10 females, $M_{age} = 21.30$ years, $SD = 3.00$) were recruited from the student population for this part of the study in return for course credit. One police officer (male, 35 years old) with previous training in interviewing and deception detection served as the interrogator.

TABLE 1 Police officers' self-reported level of experience

Experience level	N
Special constable (part-time volunteer)	11
Probationer (trainee officer)	3
2-5 years	3
5-10 years	4
10+ years	2
Total	23

Note: The table contains the self-reported experience level of the police officers.

Suspects were randomly allocated to either the handcuffed (n = 10) or non-handcuffed (n = 9) condition. Before the interrogation, they completed four items from the Mach-IV questionnaire, a psychometric test measuring individual differences in Machiavellianism (Christie & Geis, 1970), with responses made on a 10-point Likert-type scale ("*strongly disagree*"—"*strongly agree*"). The questions were: Q1. "The best way to handle people is to tell them what they want to hear", Q2. "It is safest to assume that all people have a vicious streak and it will come out when they are given the chance", Q3. "There is no excuse for lying to someone else", and Q4. "All in all, it is better to be humble and honest than to be important and dishonest". Suspects' true responses were then transferred to a new sheet, and two of the four responses were modified by moving the rating by 5 points (e.g., a response of 2 was changed to 7), while the other two remained unchanged. The two changed responses and the direction of change (±5) were counterbalanced between suspects. Thus, each suspect had two honest and two deceptive answers (for more details on this procedure, see Levine et al., 1999).

Prior to the interrogation, participants were allowed to read through their changed responses and were instructed to justify them to the interrogator during a videotaped interview; they also had the opportunity to ask the researcher questions before being escorted to the interview room. The interrogator was blind to the veracity of these responses. Suspects in the handcuffed condition were placed in standard UK police-style rigid handcuffs, with their hands placed in front of them before entering the interrogation room. After being seated, the interrogator would begin the recording, confirming this to the suspect. The camera was placed on the interview table on a tripod in full view and aimed at the suspect; the videos do not show the interrogator, but their voice can be heard when questioning the suspect.

The interrogator read out the questions alongside the suspect's response. He then asked, "Why did you answer this way?". After suspects provided an initial response, the interrogator probed suspects on their statements. During probing, the suspect is required to produce additional information to answer a question (Granhag & Strömwall, 2001). Probing is recommended by many police manuals due to the assumption that it increases the likelihood of a confession (Inbau et al., 2011), although it has not been found to enhance deception detection (Buller et al., 1991; Stiff & Miller, 1986). Here, probing was used to provide flexibility in the interrogator's behaviour (i.e., by reducing a scripted and anticipatory talking pattern), and to force suspects to provide lengthier statements, thereby closely mirroring common police procedures.

The interrogator chose from one of three probes: (a) positive: "That sounds reasonable but tell me a little more about why you answered it that way", (b) neutral: "Tell me a little more about why you answered it that way", or (c) negative: "I don't believe you really think that. Tell me a little more about why you answered it that way". The interrogator then judged and marked the statement made by the suspects as either a "lie" or "truth" and rated his confidence in these judgements on a 5-point Likert-type scale (1 "*not at all confident*" to 5 "*very confident*"). After the interrogation, the handcuffs were removed, and suspects were asked four post-interrogation questions (see SI for more details on the interrogator and suspects' data and analyses).

Only responses to the Mach-IV's Q2 were considered in the current paper to ensure consistency in the question asked and answered across all videos (devoid of technical problems) and to prevent judges in the deception detection task from seeing the same suspects twice, which may introduce carryover and/or practice effects (namely, if judges see a suspect multiple times, their responses may be influenced by the past judgement they provided irrespective of the content of the new statement). This resulted in sixteen videos being used for the deception detection task.

5 | STIMULI

The Q2 videos used in the deception detection task contain the full statements made by each suspect beginning with the question being asked by the interrogator, the suspect's initial response, the interrogator's probe, and the suspect's elaboration on their answer; while the voice of the interrogator can be heard during the video, only the suspect is visible in the video. To ensure that an equal number of videos were presented to the judges in each condition, two handcuffed videos (one lie and one truth), and one non-handcuffed video (one lie) were excluded (randomly selected).

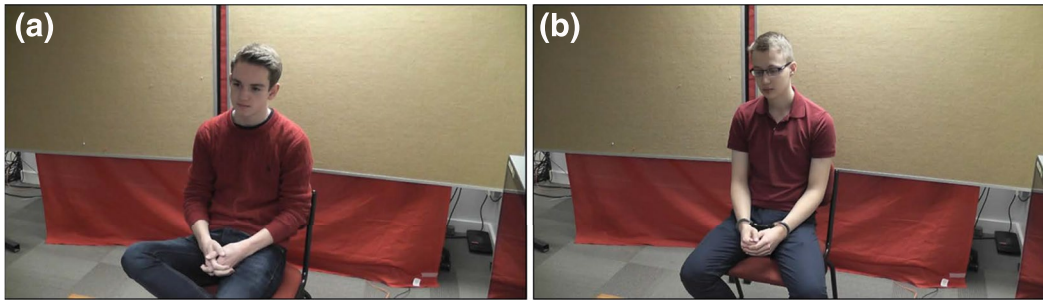

FIGURE 1 Examples of Stimuli Depicting Suspects in the (a) Non-Handcuffed and (b) Handcuffed Condition. *Note*: Both suspects provided consent for the publication of their images

This resulted in eight videos in the handcuffed condition (4 lies, 4 truths; 4 males, 4 females) and eight videos in the non-handcuffed condition (4 lies, 4 truths; 2 males, 6 females). All videos are 1920 × 1080 pixels at 30 frames-per-second, and around 2 min in length. See Figure 1 for example, stimuli.[3]

5.1 | Procedure

The task was distributed online using the Qualtrics survey platform. Participants were randomly allocated to view either the handcuffed or non-handcuffed videos. Participants were asked for their age, gender, previous deception detection training, and whether they were a police officer. Police officers were further asked about their length of service.

Before starting the task, participants viewed a condition-specific test video to ensure they understood the instructions and that the videos were displaying correctly. The instructions stated that individuals in the videos were either answering truthfully or deceptively about their personal opinions on a topic during an interrogation setting. The videos were then presented in random order. For each video, judges provided a veracity judgement using a 7-point scale (1 "*very dishonest*"—4 "*don't know*"—7 "*very honest*") and a confidence rating using a 7-point scale (1 "*not at all confident*"—7 "*very confident*"). Employing an honesty scale permits for more variance in judges' response uncertainty to be captured while not affecting accuracy (e.g., selecting "very dishonest" is not more 'accurate' than selecting "dishonest" but it is informative), and avoids issues with social norms and moral condemnation not permitting judges to express strict opinions (Levine, 2001). They were also asked to specify which cues they used to reach their judgement, phrased as "What was the most important factor that made you determine if the person was lying or telling the truth?", using a list provided at the end of the task. Finally, participants were debriefed.

6 | RESULTS

Judges' deception detection accuracy, judgement bias, and confidence scores were analysed based on the handcuffing manipulation and judge type. For the analysis and breakdown of the self-reported cues used by judges during the deception task, see SI.

6.1 | Deception detection accuracy

Judges' honesty ratings were used to compute accuracy scores; values 1-3 were coded as lie and values 5–7 as truth, while 4 ("don't know") was coded as incorrect (for a discussion on continuous vs. dichotomous judgements, see Levine, Shaw, & Shulman, 2010). These responses were compared to the veracity of the videos to create accuracy

scores. Thus, if the judge responded with "truth" to a truthful video it was considered correct (score = 1), but if they answered "lie" it was considered incorrect (score = 0) and vice versa for deceptive videos. These values were summed and converted to percentage accuracy scores (see SI for a complementary analysis preserving the original structure of the data).[4] To investigate veracity judgements, lie and truth accuracies were treated separately to account for potential differences in veracity-specific accuracy rates of the judges (see Levine et al., 1999).

First, an analysis of the accuracy scores was conducted with handcuffing (non-handcuffed vs. handcuffed), judge (police officers vs. laypersons), and veracity (lies and truths) as predictors. To account for the unbalanced data structure, a mixed-model ANOVA (Type III, full factorial) using the Parametric Bootstrap (PB) method (with 10,000 simulations) was used (see Krishnamoorthy et al., 2007). Significant omnibus effects were followed up with robust post-hoc tests (trimmed means and Winsorized variance), estimating 95% bootstrapped confidence intervals around robust effects sizes.[5] A robust heteroscedastic analysis of the explanatory measure of effect size (ξ) is reported for between-subjects effects and post-hoc comparisons (interpretation: small = 0.10, medium = 0.30, large >0.50; range: 0–1; Wilcox & Tian, 2011), and the robust Cohen's d (d_R) is reported for within-subjects effects and post-hoc comparisons (interpretation: small = 0.20, medium = 0.50, large >0.80; range: 0–1; Algina et al., 2005). The assumption of homoscedasticity held for all analyses, $F(3, 102) = 1.87$, $p = 0.139$ (truths), $F(3, 102) = 1.41$, $p = 0.243$ (lies).

There was a statistically significant main effect of veracity on accuracy, $\chi^2(1, N = 106) = 7.11$, $p = 0.012$, $d_R = 0.21$, 95% CI_{PB} [0.04, 0.46], Vovk-Sellke Maximum p-Ratio (VS-MPR; Sellke et al., 2001) = 7.01, with lies ($M = 43.63\%$, $SD = 24.65\%$) being harder to detect than truths ($M = 52.60\%$, $SD = 22.08\%$). There was also a significant main effect of handcuffing, $\chi^2(1, N = 106) = 4.82$, $p = 0.029$, $\xi = 0.33$, 95% CI_{PB} [0.11, 0.56], VS-MPR = 3.56, with lower accuracy for handcuffed suspects ($M = 44.81\%$, $SD = 14.39\%$) than non-handcuffed suspects ($M = 51.41\%$, $SD = 13.35\%$). There was no statistically significant difference in accuracy between police officers ($M = 45.13\%$, $SD = 15.43\%$) and laypersons ($M = 48.95\%$, $SD = 13.84\%$), $\chi^2(1, N = 106) = 0.88$, $p = 0.349$, nor interactions, χ^2s ≤ 2.26, ps > 0.144.

Scores (total and for each veracity type) were also compared to chance accuracy. Laypersons were statistically no different from chance either overall, $t(82) = -0.69$, $p = 0.489$, lies, $t(82) = -1.90$, $p = 0.062$, or truths, $t(82) = -1.25$, $p = 0.213$. Police officers were statistically no different from chance either overall, $t(22) = -1.52$, $p = 0.142$, or for truths, $t(22) = 0.23$, $p = 0.824$; however, lies were detected with below chance accuracy, $t(22) = -2.10$, $p = 0.047$, $d_z = -0.44$, 95% CI [−0.86, −0.01], VS-MPR = 2.56.

The accuracy data were also analysed using Signal Detection Theory (SDT), separating judgement accuracy from response bias (Masip et al., 2009; Meissner & Kassin, 2002). A two-way independent-measures parametric bootstrapped ANOVA (with 10,000 simulations) compared discriminability (A'; Rae, 1976) between judges and handcuffing conditions. For A′, a value of 0.50 indicates chance level performance. A main effect of handcuffing was found, $F(1, 102) = 6.04$, $p = 0.016$, $\xi = 0.37$, 95% CI_{PB} [0.10, 0.58], VS-MPR = 5.64, where the veracity of handcuffed suspects ($M = 0.45$, $SD = 0.23$) was overall harder to discriminate than that of non-handcuffed suspects ($M = 0.55$, $SD = 0.19$). No effect of judge was found, $F(1, 102) = 2.52$, $p = 0.114$, nor a judge by handcuffing interaction, $F(1, 102) = 0.49$, $p = 0.484$.

6.2 | Judgement bias

To understand the impact of handcuffing on judgement bias, each veracity judgement was coded as −1 for a lie and +1 for a truth and then summed; a score of 0 reflects no bias, >0 a truth-bias, and <0 a lie-bias. A two-way independent-measures parametric bootstrapped ANOVA (with 10,000 simulations) on judgement bias based on judge and handcuffing did not reveal any statistically significant differences, Fs ≤ 2.08, ps > 0.152. The bias data were also analysed using SDT (B"; Donaldson, 1992), but this did not reveal any statistically significant effects, Fs ≤ 2.18, ps > 0.143.

Considering participants' overall responses to the base-rate of being unbiased (0), a one-sample t-test revealed a statistically significant difference, $t(105) = 2.48$, $p = 0.015$, $d_z = 0.24$, 95% CI [0.05, 0.44], VS-MPR = 5.93, suggesting that judges were overall truth-biased in their judgements ($M = 0.72$, $SD = 2.98$).

6.3 | Confidence

The confidence ratings for each veracity decision were summed across videos. A two-way independent-measures parametric bootstrapped ANOVA (with 10,000 simulations) compared confidence ratings based on judge and handcuffing condition. This revealed a main effect of judge, with police officers (M = 42.04, SD = 4.63) being more confident in their judgements than laypersons (M = 38.00, SD = 6.36), $F(1, 102) = 8.33$, $p = 0.004$, $\xi = 0.50$, 95% CI_{PB} [0.10, 0.82], VS-MPR = 16.66. It also revealed a main effect of handcuffing, with handcuffed videos (M = 37.74, SD = 6.80) being less confidently judged that non-handcuffed videos (M = 40.02, SD = 5.44), $F(1, 102) = 4.12$, $p = 0.042$, $\xi = 0.34$, 95% CI_{PB} [0.06, 0.61], VS-MPR = 2.78. The interaction term was not statistically significant, $F = 0.90$, $p = 0.346$.

Pearson's correlations for confidence and accuracy revealed no statistically significant relationship, $r(106) = 0.055$, $p = 0.574$, 95% CI [-0.14, 0.24]. Considering each veracity type separately, a significant positive correlation was observed for truth detection, $r(106) = 0.242$, $p = 0.012$, 95% CI [0.05, 0.41], VS-MPR = 6.93, and a negative correlation for lie detection, $r(106) = -0.211$, $p = 0.030$, 95% CI [-0.39, −0.02], VS-MPR = 3.42; the difference between the two veracity correlations was statistically significant, Steiger's $Z = -2.94$, $p = 0.003$, $r_{diff} = -0.45$, 95% CI [-0.73, −0.16], VS-MPR = 21.11.

7 | DISCUSSION

The present research explored whether a situational factor related to interrogation procedures (i.e., the use of handcuffs on suspects) can negatively impact veracity judgements. Confirming our hypothesis, the handcuffing manipulation affected both laypersons' and police officers' ability to detect deception (i.e., H_2 was supported; moderate effect size). Statements made by handcuffed suspects were harder to classify for both police officers and laypersons. Converting the handcuffing effect size ($\xi = 0.37$) to more intuitive estimates (as recommended by Fritz et al., 2012), we obtain a Number Needed to Treat (NNT) of 5.01. Meaning for every fifth person that is interviewed wearing handcuffs we would expect one more misclassification of veracity. Or, based on the Common Language (CL) effect size, the probability that a suspect selected at random from the handcuffed condition is misclassified in terms of statement veracity compared to a suspect from the non-handcuffed condition is 64.3%. This decrease in accuracy was attributable to the study's manipulation affecting veracity discriminability rather than a shift in judgement response tendencies (H_1 was not supported), as all judges remained truth-biased overall (H_3 was not supported; NNT = 10.54, CL = 56.7%). For both judge groups, truths were easier to detect than lies (NNT = 12.02, CL = 55.9%; replicating the veracity effect; Levine et al., 1999).

Unsurprisingly, police officers did not perform better at judging veracity than laypersons (see Aamodt & Custer, 2006), and judging handcuffed suspects made this process even harder. However, the manipulation did not affect officers' response bias (H_5 was not supported). This contrasts research arguing for a veracity detection reversal in professionals (i.e., police officers showing higher lie detection, but lower truth detection compared to laypersons; Meissner & Kassin, 2002). The similarity in response patterns with laypersons indicates that police officers were not overall more suspicious of suspects. This could, however, be due to the relatively junior sample of officers recruited (see Table 1), or, potentially, due to the "suspects" being naïve students which may have mitigated lie bias towards them; however, we note that the instructions never mention the status of suspects.

A more worrying result, and per our prediction, police officers displayed higher confidence while being no more accurate than laypersons (i.e., H_4 was supported; moderate-to-large effect size; NNT = 3.66, CL = 70.2%), even showing a trend towards lower accuracy (e.g., below chance lie detection; NNT = 5.88, CL = 62.2%). This parallels findings of professionals tending to be overconfident in their veracity judgements (Aamodt & Custer, 2006; DePaulo & Pfeifer, 1986; Masip et al., 2016). While the police officers' level of experience may have not been sufficient to bias their judgements in the direction of a lie, it was able to increase their confidence in catching liars (e.g., Masip et al., 2016).

Overall, judges performed worse at discriminating veracity when viewing handcuffed suspects, supporting our assertions that situational factors can negatively impact the discriminability between deceptive and honest suspects (for a more detailed breakdown of the honesty scale data, see SI). Such effects may have serious ramifications for the forensic domain (Verschuere et al., 2016), especially when considering the already poor deception detection rates in the absence of the handcuffing manipulation. Interestingly, both laypersons and police officers were less confident in their judgements when they watched the handcuffed (vs. non-handcuffed) videos (NNT = 5.32, CL = 63.6%). Judges may have found deception detection more difficult when suspects were handcuffed, tempering their confidence.

These results illustrate that situational elements can impact the perception and judgement of both laypersons and police officers. Reducing the impact of such artificial factors could improve forensic practices and deception detection procedures, whilst reducing the risk of potential miscarriages of justice. Such effects are especially pertinent in situations of judgement under uncertainty where external and contextual information often influence the perception of ambiguous or ambivalent information (Masip et al., 2009; Mobbs et al., 2006). In line with research on investigative interviewing, it would seem recommendable that the space and circumstances under which an interrogation takes place are comfortable and do not restrict the individual (Goodman-Delahunty et al., 2014; Kelly et al., 2013).

7.1 | Future directions

The current work sought to highlight the effects of situational factors on veracity judgements, particularly in forensic contexts. Future research could elaborate on the different ways in which handcuffing affects senders and judges by separating their influence on suspect perceptions (e.g., handcuffs as a visual cue of criminality; Stiff et al., 1992) from the effect on suspects' ability to gesticulate (within-sender features). For this, handcuffed and non-handcuffed suspects' movements could be restricted by asking them, for example, to place their hands flat on a table throughout the interrogation. This would equate the nonverbal differences whilst having the presence/absence of handcuffs as the only factor that differs between conditions. Alternatively, the videos could be edited to show the same suspect with or without handcuffs, revealing whether any impressions brought about by being handcuffed are due to the presence of external visual cues.

Considerations should also be given to the content of the stimuli themselves. An analysis of the videos may reveal verbal, paraverbal, and/or nonverbal cues which may aid in understanding the current findings. Such an investigation could uncover if behavioural differences between the liars and truth-tellers are indeed reduced by handcuffing and if differences in impression management are brought about by the manipulation (e.g., handcuffed suspects may "compensate" for their restricted gesticulation by modifying their speech and, by extension, their verbal cues may differ; see Verschuere et al., 2021).

Additionally, given the within-sender variability typically seen in deception research (Levine, 2010; Zloteanu, Bull, et al., 2021), the current stimulus set may be expanded to show a larger number of senders which would provide more precise effect size estimates and reduced uncertainty (Levine et al., 2022). Future research should also employ a more in-depth statistical approach (i.e., multi-level modelling) that accounts for both sender and decoder variability. This may be especially relevant in understanding if handcuffing interacts with senders' demeanour and judges' expectations. The possibility exists that the manipulation may not affect all individuals to the same degree or in the same manner (see DAG in SI for the potential influence of within/between subject and stimuli variance on the judgement process).

Subsequent work may also explore the effect of handcuffing on the relationship quality between suspect and interrogator (also, see SI). Due to the interactive nature of the interrogation task, handcuffs may have affected the rapport between the interrogator and suspect, which in turn could shape the behaviour of suspects (Kassin et al., 2003; Paton et al., 2018). The present manipulation demonstrates that deception detection does not happen in isolation. Future studies investigating veracity judgements should expand the range of factors being considered, both within the lab and in the real world.

7.2 | Limitations

The issue of generalisability in the deception field is rarely addressed; nonetheless, a few elements of the current research must be considered. First, the type of lie told by suspects related to personal information that liars misrepresented. It can be argued that differences in performance and judgement may emerge if other types of lies (e.g., lies about transgressions) are employed (Levine, Kim, & Blair, 2010; cf. Hartwig & Bond, 2014; Hauch et al., 2014). Second, although some have argued that using students instead of real suspects may impact the detection rate (see O'Sullivan et al., 2009), both empirical investigations and meta-analyses report that deception detection is unaffected by whether the sender is a student or not (Hartwig & Bond, 2014; Zhang et al., 2013), nor do police officers show better accuracy rates even in naturalistic high-stakes settings (Hartwig, 2004; Meissner & Kassin, 2002). However, using different type of senders may influence perceptions and judgements.

Presently, it is difficult to separate the effect of handcuffing on judges' perception (i.e., pure external features) from that on sender performance (i.e., within-sender features) as our manipulation may have been affecting either or both. For example, handcuffing could attenuate behavioural differences between liars and truth-tellers resulting in poorer overall veracity discrimination. However, considering the dynamics between the interrogator and the suspects, being handcuffed could have also prompted senders as to the added scrutiny and behavioural restrictions, and compensated through increased impression management to produce a more convincing performance (Buller & Burgoon, 1996; Burgoon et al., 1996). The interplay between the interviewee and the interviewer is an important unknown, as some response variability may be due to the interrogator himself, given that rapport strongly influences interviewing outcomes (Abbe & Brandon, 2013).

The interrogation style used should also be weighed. Currently, while we did not find any effect of probing, this element could not be explored in depth due to a lack of variability in the use of the three probes by the interrogator (see SI). The literature on probing is equivocal on its use impacting veracity judgements (Buller et al., 1991). Nonetheless, it may impact rapport building and disclosure (Paton et al., 2018). Different probes may result in changes in the interdynamics of the interrogator and suspect, as well as subsequent judges (e.g., biasing impressions based on the valence of the probe used during the questioning). Future research could consider manipulating (e.g., standardising) the probing element to investigate how it interacts with the handcuffing element (e.g., Granhag & Strömwall, 2001); specific probes may bolster (e.g., negative) or attenuate (e.g., positive) the effects of handcuffing.

Finally, a more pronounced limitation is the relatively small and unbalanced sample. Underpowered studies are less likely to find true effects (i.e., Type II error), have a higher chance of found effects being statistical artefacts (i.e., Type I error), inflate estimates of true effects (i.e., Type M error), and have lower replicability (Fraley & Vazire, 2014; Gelman & Carlin, 2014). For instance, the CIs around the handcuffing effect indicate that the data is compatible with a wide range of effect sizes, from large and of potential interest ($\xi = 0.58$) to small and potentially unimportant ($\xi = 0.10$). Thus, we advise readers to interpret the results with care. Still, considering the forensic-relevant sample alongside the implications of our findings (especially for miscarriages of justice), on balance, we consider that the value of the research outweighs its drawbacks (Eckermann et al., 2010; Sterling et al., 1995).

To increase usability, we report all necessary measurements of uncertainty and variability (Calin-Jageman & Cumming, 2019), permitting future hypothesis generation and integration into meta-analyses (Cumming, 2014; Fritz et al., 2012). For example, replications can consider the effect sizes reported and their confidence intervals to estimate future results (e.g., prediction intervals; Cumming, 2008), and calculate the statistical power needed to reproduce the effect (e.g., considering $\xi_{33\%}$; see, Simonsohn, 2015).

8 | CONCLUSION

Handcuffing was found to impact several aspects of the veracity judgement process. Judging veracity was more difficult when suspects were handcuffed, resulting in lower accuracy for both police officers and laypersons. Although

police officers were no more accurate at detecting lies and truths than laypersons, they were more confident in their judgements. Both groups, however, showed a decrease in confidence for handcuffed suspects, indicating that this scenario was more difficult. Handcuffing is a common approach to handling suspects, yet the addition of this external element influences how the deception detection process unfolds, specifically in ways that can be detrimental to the forensic process. Deception researchers and practitioners should address situational factors that can interfere with judgemental accuracy and attempt to reduce their potentially negative influence.

AUTHOR CONTRIBUTIONS

Mircea Zloteanu: Conceptualization (lead); writing – original draft (lead); formal analysis (lead); methodology (lead); writing – review and editing (equal); investigation (lead); resources (equal). **Nadine L. Salman:** Writing – original draft (supporting); investigation (supporting); writing – review and editing (equal); resources (equal). **Eva G. Krumhuber:** Writing – review and editing (equal). **Daniel C. Richardson:** Writing – review and editing (supporting); resources (equal).

CONFLICT OF INTEREST

The authors declared no conflicts of interest with respect to the research, authorship, and/or publication of this article.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study and videos are available from the corresponding author on request.

ORCID


Mircea Zloteanu https://orcid.org/0000-0002-2753-637X

ENDNOTES

[1] The terms "deception detection" and "veracity judgment" are often used interchangeably in the literature, yet we consider they reflect an important theoretical and operational distinction. Notably, the former assumes the existence of a correct answer, with the judge's task being to "detect" the relevant information and make an accurate inference (i.e., it is treated as an *intellective task*), while the latter investigates the formation of a decision when accurate inference may be impossible (i.e., a *judgmental task*; see Carey & Laughlin, 2011).

[2] The term "situational" and "contextual" are used interchangeably; here, we prefer the former as in the deception literature "context" is occasionally used to refer to the lie scenario (e.g., the lie or judgment being made in a forensic, clinical, or social context) rather than the elements peripheral to the deceptive event. However, we caution readers that this operational definition should not be construed as an exhaustive or exact term, and should be interpreted within the context of the current work (see Lilienfeld et al., 2015).

[3] Examples of the suspect videos and additional information on the methodology can be obtained from the corresponding author on request.

[4] The pattern of results of the analysis using the raw honesty scale support our primary hypothesis; see SI.

[5] All analyses were conducted in *R* (R Core Team, 2020). *p*-values were determined using the *afex* (Singmann et al., 2021) and the *pbkrtest* (Halekoh & Højsgaard, 2014) packages. Robust effect size estimates and bootstrapped confidence intervals were calculated using the *WRS2* package (Mair & Wilcox, 2020).

REFERENCES

Aamodt, M. G., & Custer, H. (2006). Who can best catch a liar? *Forensic Examiner*, *15*(1), 6–11.

Abbe, A., & Brandon, S. E. (2013). The role of rapport in investigative interviewing: A review: The role of rapport in investigative interviewing. *Journal of Investigative Psychology and Offender Profiling*, *10*(3), 237–249. https://doi.org/10.1002/jip.1386

Akehurst, L., Köhnken, G., Vrij, A., & Bull, R. (1996). Lay persons' and police officers' beliefs regarding deceptive behaviour. *Applied Cognitive Psychology*, *10*(6), 461–471. https://doi.org/10.1002/(sici)1099-0720(199612)

Algina, J., Keselman, H. J., & Penfield, R. D. (2005). An alternative to Cohen's standardized mean difference effect size: A robust parameter and confidence interval in the two independent groups case. *Psychological Methods*, *10*(3), 317–328. https://doi.org/10.1037/1082-989x.10.3.317

Anderson, D. E., DePaulo, B. M., Ansfield, M. E., Tickle, J. J., & Green, E. (1999). Beliefs about cues to deception: Mindless stereotypes or untapped wisdom? *Journal of Nonverbal Behavior*, *23*(1), 67–89.

Association of Chief Police Officer of England Wales & Northern Ireland. (2009). ACPO: Guidance on the use of handcuffs.

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. https://doi.org/10.1371/journal.pone.0156615

Bond, C. F., & DePaulo, B. M. (2006). Accuracy of deception judgments. *Personality and Social Psychology Review*, *10*(3), 214–234. https://doi.org/10.1207/s15327957pspr1003_2b

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

Bond, C. F., Omar, A., Pitre, U., Lashley, B. R., Skaggs, L. M., & Kirk, C. T. (1992). Fishy-looking liars: Deception judgment from expectancy violation. *Journal of Personality and Social Psychology*, *63*(6), 969–977. https://doi.org/10.1037/0022-3514.63.6.969

Buller, D. B., & Burgoon, J. K. (1996). Interpersonal deception theory. *Communication Theory*, *6*(3), 203–242. https://doi.org/10.4135/9781412959384.n206

Buller, D. B., Strzyzewski, K. D., & Comstock, J. (1991). Interpersonal deception: I. Deceivers' reactions to receivers' suspicions and probing. *Communication Monographs*, *58*(1), 1–24. https://doi.org/10.1080/03637759109376211

Burgoon, J. K. (2018). Microexpressions are not the best way to catch a liar. *Frontiers in Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.01672

Burgoon, J. K., Blair, J. P., & Strom, R. E. (2008). Cognitive biases and nonverbal cue availability in detecting deception. *Human Communication Research*, *34*(4), 572–599. https://doi.org/10.1111/j.1468-2958.2008.00333.x

Burgoon, J. K., Buller, D. B., Ebesu, A. S., White, C. H., & Rockwell, P. A. (1996). Testing interpersonal deception theory: Effects of suspicion on communication behaviors and perceptions. *Communication Theory*, *6*(3), 243–267. https://doi.org/10.1111/j.1468-2885.1996.tb00128.x

Calin-Jageman, R. J., & Cumming, G. (2019). The new statistics for better science: Ask how much, how uncertain, and what else is known. *The American Statistician*, *73*(sup1), 271–280. https://doi.org/10.1080/00031305.2018.1518266

Carey, H. R., & Laughlin, P. R. (2011). Groups perform better than the best individuals on letters-to-numbers problems: Effects of induced strategies. *Group Processes & Intergroup Relations*, *15*(2), 231–242. https://doi.org/10.1177/1368430211419174

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. https://doi.org/10.1007/s10919-005-0001-z

Chaikin, A. L., Derlega, V. J., & Miller, S. J. (1976). Effects of room environment on self-disclosure in a counseling analogue. *Journal of Counseling Psychology*, *23*(5), 479–481. https://doi.org/10.1037/0022-0167.23.5.479

Christie, R. G., & Geis, F. (1970). *Studies in machiavellianism*. Academic Press.

Colwell, L. H., Miller, H. A., Lyons, P. M., & Miller, R. S. (2006). The training of law enforcement officers in detecting deception: A survey of current practices and suggestions for improving accuracy. *Police Quarterly*, *9*(3), 275–290. https://doi.org/10.1177/1098611104273293

Courts and Tribunals Judiciary. (2010). Protocol: Management of prisoners that present a risk of escape or violence when attending Criminal Courts.

Cumming, G. (2008). Replication and *p* intervals: *P* values predict the future only vaguely, but confidence intervals do much better. *Perspectives on Psychological Science*, *3*(4), 286–300. https://doi.org/10.1111/j.1745-6924.2008.00079.x

Cumming, G. (2014). The new statistics: Why and how. *Psychological Science*, *25*(1), 7–29. https://doi.org/10.1177/0956797613504966

Dawson, E., Hartwig, M., Brimbal, L., & Denisenkov, P. (2017). A room with a view: Setting influences information disclosure in investigative interviews. *Law and Human Behavior*, *41*(4), 333–343. https://doi.org/10.1037/lhb0000244

Denault, V., & Patterson, M. L. (2021). Justice and nonverbal communication in a post-pandemic world: An evidence-based commentary and cautionary statement for lawyers and judges. *Journal of Nonverbal Behavior*, *45*(1), 1–10. https://doi.org/10.1007/s10919-020-00339-x

Denault, V., Plusquellec, P., Jupe, L. M., St-Yves, M., Dunbar, N. E., Hartwig, M., Sporer, S. L., Rioux-Turcotte, J., Jarry, J., Walsh, D., Otgaar, H., Viziteu, A., Talwar, V., Keatley, D. A., Blandón-Gitlin, I., Townson, C., Deslauriers-Varin, N., Lilienfeld, S. O., Patterson, M. L., & Koppen, P. J. van (2020). The analysis of nonverbal communication: The dangers of pseudoscience in security and justice contexts. *Anuario de Psicología Jurídica*, *30*(1), 1–12. https://doi.org/10.5093/apj2019a9

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. https://doi.org/10.1207/s15327957pspr0104_5

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. https://doi.org/10.1037/0033-2909.129.1.74

DePaulo, B. M., & Pfeifer, R. L. (1986). On-the-job experience and skill at detecting deception. *Journal of Applied Social Psychology*, *16*(3), 249–267. https://doi.org/10.1111/j.1559-1816.1986.tb01138.x

Donaldson, W. (1992). Measuring recognition memory. *Journal of Experimental Psychology: General*, *121*(3), 275–277. https://doi.org/10.1037/0096-3445.121.3.275

Duran, N. D., Dale, R., Kello, C. T., Street, C. N. H., & Richardson, D. C. (2013). Exploring the movement dynamics of deception. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00140

Eckermann, S., Karnon, J., & Willan, A. R. (2010). The value of value of information: Best informing research design and prioritization using current methods. *PharmacoEconomics*, *28*(9), 699–709. https://doi.org/10.2165/11537370-000000000-00000

Fraley, R. C., & Vazire, S. (2014). The N-pact factor: Evaluating the quality of empirical journals with respect to sample size and statistical power. *PLoS One*, *9*(10), e109019. https://doi.org/10.1371/journal.pone.0109019

Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: Current use, calculations, and interpretation. *Journal of Experimental Psychology: General*, *141*(1), 2–18. https://doi.org/10.1037/a0024338

Funk, F., & Todorov, A. (2013). Criminal stereotypes in the courtroom: Facial tattoos affect guilt and punishment differently. *Psychology, Public Policy, and Law*, *19*(4), 466–478. https://doi.org/10.1037/a0034736

Garrido, E., Masip, J., & Herrero, C. (2004). Police officers' credibility judgments: Accuracy and estimated ability. *International Journal of Psychology*, *39*(4), 254–275. https://doi.org/10.1080/00207590344000411

Gelman, A., & Carlin, J. (2014). Beyond power calculations: Assessing type S (sign) and type M (magnitude) errors. *Perspectives on Psychological Science*, *9*(6), 641–651. https://doi.org/10.1177/1745691614551642

George, J. F., Tilley, P., & Giordano, G. (2014). Sender credibility and deception detection. *Computers in Human Behavior*, *35*, 1–11. https://doi.org/10.1016/j.chb.2014.02.027

Gifford, R. (1988). Light, decor, arousal, comfort and communication. *Journal of Environmental Psychology*, *8*(3), 177–189. https://doi.org/10.1016/S0272-4944(88)80008-2

Goodman-Delahunty, J., Martschuk, N., & Dhami, M. K. (2014). Interviewing high value detainees: Securing cooperation and disclosures. *Applied Cognitive Psychology*, *28*(6), 883–897. https://doi.org/10.1002/acp.3087

Goodman-Delahunty, J., & Sivasubramaniam, D. (2013). *Investigative and intelligence interviewing in Asia-Pacific jurisdictions*. Report to the US Department of Justice, Federal Bureau of Investigation and Centre for Law and Human Behaviour University of Texas. University of Texas.

Granhag, P. A., & Strömwall, L. A. (2001). Deception detection: Interrogators' and observers' decoding of consecutive statements. *Journal of Psychology*, *135*(6), 603–620. https://doi.org/10.1080/00223980109603723

Gudjonsson, G. H. (2005). Fitness to be interviewed. In J. Payne-James, R. W. Byard, T. S. Corey, & C. Henderson (Eds.), *Encyclopedia of forensic and legal medicine* (Vol. 2, pp. 169–174). Elsevier.

Halekoh, U., & Højsgaard, S. (2014). A Kenward-Roger approximation and parametric bootstrap methods for tests in linear mixed models – the R package pbkrtest. *Journal of Statistical Software*, *59*(9), 1–30. https://doi.org/10.18637/jss.v059.i09

Hartwig, M., & Bond, C. F. (2014). Lie detection from multiple cues: A meta-analysis. *Applied Cognitive Psychology*, *28*(5), 661–676. https://doi.org/10.1002/acp.3052

Hartwig, M., Granhag, P. A., Stromwall, L. A., & Vrij, A. (2004). Police officers' lie detection accuracy: Interrogating freely versus observing video. *Police Quarterly*, *7*(4), 429–456. https://doi.org/10.1177/1098611104264748

Hauch, V., Sporer, S. L., Michael, S. W., & Meissner, C. A. (2014). Does training improve the detection of deception? A meta-analysis. *Communication Research*, *43*(3), 283–343. https://doi.org/10.1177/0093650214534974

Hoogesteyn, K., Meijer, E., & Vrij, A. (2019). The influence of room spaciousness on investigative interviews. *Legal and Criminological Psychology*, *24*(2), 215–228. https://doi.org/10.1111/lcrp.12156

Hoogesteyn, K., Meijer, E. H., & Vrij, A. (2020). Detainee and layperson's expectations and preferences regarding police interview rooms. *PLoS One*, *15*(11), e0241683. https://doi.org/10.1371/journal.pone.0241683

Horvath, F., Jayne, B., & Buckley, J. (1994). Differentiation of truthful and deceptive criminal suspects in behavior analysis interviews. *Journal of Forensic Sciences*, *39*(3), 793–807. https://doi.org/10.1520/JFS13657J

Hurst, M., & Oswald, M. (2012). Mechanisms underlying response bias in deception detection. *Psychology, Crime and Law*, *18*(8), 759–778. https://doi.org/10.1080/1068316X.2010.550615

Inbau, F. E., Reid, J. E., Buckley, J. P., & Jayne, B. C. (2011). *Criminal interrogation and confessions* (4th ed.). Jones & Bartlett Publishers.

Kassin, S. M., Goldstein, C. C., & Savitsky, K. (2003). Behavioral confirmation in the interrogation room: On the dangers of presuming guilt. *Law and Human Behavior*, *27*(2), 187–203. https://doi.org/10.1023/A:1022599230598

Kelly, C. E., Miller, J. C., Redlich, A. D., & Kleinman, S. M. (2013). A taxonomy of interrogation methods. *Psychology, Public Policy, and Law*, *19*(2), 165–178. https://doi.org/10.1037/a0030310

Kim, R. K., & Levine, T. R. (2011). The effect of suspicion on deception detection accuracy: Optimal level or opposing effects? *Communication Reports*, *24*(2), 51–62. https://doi.org/10.1080/08934215.2011.615272

Knapp, M. L., Hall, J. A., & Horgan, T. G. (2013). *Nonverbal communication in human interaction*. Wadsworth Cengage Learning.

Krishnamoorthy, K., Lu, F., & Mathew, T. (2007). A parametric bootstrap approach for ANOVA with unequal variances: Fixed and random models. *Computational Statistics & Data Analysis*, *51*(12), 5731–5742. https://doi.org/10.1016/j.csda.2006.09.039

Küster, D., Krumhuber, E. G., & Hess, U. (2019). You are what you wear: Unless you moved—Effects of attire and posture on person perception. *Journal of Nonverbal Behavior*, *43*(1), 23–38. https://doi.org/10.1007/s10919-018-0286-3

Levine, T. R. (2010). A few transparent liars: Explaining 54% accuracy in deception detection experiments. In C. T. Salmon (Ed.), *Communication yearbook* (Vol. 34, pp. 40–61). SAGE Publications, Inc. https://doi.org/10.1080/23808985.2010.11679095

Levine, T. R. (2001). Dichotomous and continuous views of deception: A reexamination of deception ratings in information manipulation theory. *Communication Research Reports*, *18*(3), 230–240. https://doi.org/10.1080/08824090109384803

Levine, T. R., Anders, L. N., Banas, J., Baum, K. L., Endo, K., Hu, A. D. S., & Wong, N. C. H. (2000). Norms, expectations, and deception: A norm violation model of veracity judgments. *Communication Monographs*, *67*(2), 123–137. https://doi.org/10.1080/03637750009376500

Levine, T. R., Clare, D. D., Green, T., Serota, K. B., & Park, H. S. (2014). The effects of truth–lie base rate on interactive deception detection accuracy. *Human Communication Research*, *40*(3), 350–372. https://doi.org/10.1111/hcre.12027

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., Kim, R. K., & Blair, J. P. (2010). (In) accuracy at detecting true and false confessions and denials: An initial test of a projected motive model of veracity judgments. *Human Communication Research*, *36*(1), 82–102. https://doi.org/10.1111/j.1468-2958.2009.01369.x

Levine, T. R., & McCornack, S. A. (1991). The dark side of trust: Conceptualizing and measuring types of communicative suspicion. *Communication Quarterly*, *39*(4), 325–340. https://doi.org/10.1080/01463379109369809

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

Levine, T. R., Serota, K. B., Shulman, H., Clare, D. D., Park, H. S., Shaw, A. S., Shim, J. C., & 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. https://doi.org/10.1111/j.1468-2958.2011.01407.x

Levine, T. R., Shaw, A. S., & Shulman, H. (2010). Assessing deception detection accuracy with dichotomous truth–lie judgments and continuous scaling: Are people really more accurate when honesty is scaled? *Communication Research Reports*, *27*(2), 112–122. https://doi.org/10.1080/08824090903526638

Lilienfeld, S. O., Sauvigné, K. C., Lynn, S. J., Cautin, R. L., Latzman, R. D., & Waldman, I. D. (2015). Fifty psychological and psychiatric terms to avoid: A list of inaccurate, misleading, misused, ambiguous, and logically confused words and phrases. *Frontiers in Psychology*, *6*, 1100. https://doi.org/10.3389/fpsyg.2015.01100

Luke, T. J. (2019). Lessons from Pinocchio: Cues to deception may be highly exaggerated. *Perspectives on Psychological Science*, *14*(4), 646–671. https://doi.org/10.1177/1745691619838258

Mack, K. (2001). Challenging assumptions about credibility. *Law Society Journal*, *39*(10), 60–63.

Mair, P., & Wilcox, R. (2020). Robust statistical methods in R using the WRS2 package. *Behavior Research Methods*, *52*(2), 464–488. https://doi.org/10.3758/s13428-019-01246-w

Masip, J., Alonso, H., Garrido, E., & Barba, A. (2008). Police officers' suspicion and veracity judgments. In *Comunicación presentada en La 18th conference of the European association of psychology and law*. Maastricht, Holanda.

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*(9), 1282–1296. https://doi.org/10.1002/acp.1535

Masip, J., Alonso, H., Herrero, C., & Garrido, E. (2016). Experienced and novice officers' generalized communication suspicion and veracity judgments. *Law and Human Behavior*, *40*(2), 169–181. https://doi.org/10.1037/lhb0000169

Masip, J., & Herrero, C. (2015). Police detection of deception: Beliefs about behavioral cues to deception are strong even though contextual evidence is more useful. *Journal of Communication*, *65*(1), 125–145. https://doi.org/10.1111/jcom.12135

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. https://doi.org/10.1023/A:1020278620751

Mobbs, D., Weiskopf, N., Lau, H. C., Featherstone, E., Dolan, R. J., & Frith, C. D. (2006). The Kuleshov Effect: The influence of contextual framing on emotional attributions. *Social Cognitive and Affective Neuroscience*, *1*(2), 95–106. https://doi.org/10.1093/scan/nsl014

Nisbett, R. E., & Ross, L. (1991). *The person and the situation: Perspectives of social psychology*. Temple University Press.

Okken, V., van Rompay, T., & Pruyn, A. (2012). Room to move: On spatial constraints and self-disclosure during intimate conversations. *Environment and Behavior*, *45*(6), 737–760. https://doi.org/10.1177/0013916512444780

O'Sullivan, M., Frank, M. G., Hurley, C. M., & Tiwana, J. (2009). Police lie detection accuracy: The effect of lie scenario. *Law and Human Behavior*, *33*(6), 530–538. https://doi.org/10.1007/s10979-008-9166-4

Paton, W., Bain, S. A., Gozna, L., Gilchrist, E., Heim, D., Gardner, E., Cairns, D., McGranaghan, P., & Fischer, R. (2018). The combined effects of questioning technique and interviewer manner on false confessions. *Journal of Investigative Psychology and Offender Profiling*, *15*(3), 335–349. https://doi.org/10.1002/jip.1513

Porter, S., & ten Brinke, L. M. (2009). Dangerous decisions: A theoretical framework for understanding how judges assess credibility in the courtroom. *Legal and Criminological Psychology*, *14*(1), 119–134. https://doi.org/10.1348/135532508X281520

Rae, G. (1976). Table of A'. *Perceptual and Motor Skills*. *42*(1), 98. https://doi.org/10.2466/pms.1976.42.1.98

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for statistical computing. Retrieved from https://www.R-project.org/

Sellke, T., Bayarri, M. J., & Berger, J. O. (2001). Calibration of p values for testing precise null hypotheses. *The American Statistician*. *55*, 62–71. https://doi.org/10.1198/000313001300339950

Simonsohn, U. (2015). Small telescopes: Detectability and the evaluation of replication results. *Psychological Science*, *26*(5), 559–569. https://doi.org/10.1177/0956797614567341

Singmann, H., Bolker, B., Westfall, J., Aust, F., & Ben-Shachar, M. S. (2021). afex: Analysis of factorial experiments. Retrieved from https://CRAN.R-project.org/package=afex

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

Sterling, T. D., Rosenbaum, W. L., & Weinkam, J. J. (1995). Publication decisions revisited: The effect of the outcome of statistical tests on the decision to publish and vice versa. *The American Statistician*, *49*(1), 108. https://doi.org/10.2307/2684823

Stiff, J. B., Kim, H. J., & Ramesh, C. N. (1992). Truth biases and aroused suspicion in relational deception. *Communication Research*, *19*(3), 326–345. https://doi.org/10.1177/009365092019003002

Stiff, J. B., & Miller, G. R. (1986). Come to think of it…. *Human Communication Research*, *12*(3), 339–357. https://doi.org/10.1111/j.1468-2958.1986.tb00081.x

Strömwall, L., & Granhag, P. A. (2003). How to detect deception? Arresting the beliefs of police officers, prosecutors and judges. *Psychology, Crime and Law*, *9*(1), 19–36. https://doi.org/10.1080/10683160308138

ten Brinke, L. M., Khambatta, P., & Carney, D. R. (2015). Physically scarce (vs. Enriched) environments decrease the ability to tell lies successfully. *Journal of Experimental Psychology: General*, *144*(5), 982–992. https://doi.org/10.1037/xge0000103

The Global Deception Research Team. (2006). A world of lies. *Journal of Cross-Cultural Psychology*, *37*(1), 60–74. https://doi.org/10.1177/0022022105282295

Thornton, G. R. (1943). The effect upon judgements of personality traits of varying a single factor in a photograph. *The Journal of Social Psychology*, *18*(1), 127–148. https://doi.org/10.1080/00224545.1943.9921704

Toris, C., & DePaulo, B. M. (1984). Effects of actual deception and suspiciousness of deception on interpersonal perceptions. *Journal of Personality and Social Psychology*, *47*(5), 1063–1073. https://doi.org/10.1037/0022-3514.47.5.1063

Varinsky, H. (1992). Non-verbal communication enhances witness credibility. https://web.archive.org/web/20041019021447/http://www.varinsky.com/publications/nonverbal.html

Verschuere, B., Bogaard, G., & Meijer, E. (2021). Discriminating deceptive from truthful statements using the verifiability approach: A meta-analysis. *Applied Cognitive Psychology*, *35*(2), 374–384. https://doi.org/10.1002/acp.3775

Verschuere, B., Meijer, E., & Vrij, A. (2016). Increasing pressure does not benefit lie detection: A reply to ten Brinke et al. (2015). *Psychology, Crime and Law*, *22*(9), 1–920. https://doi.org/10.1080/1068316X.2016.1202249

Virginia Commonwealth University Police Department. (2014). VCUPD Written directives manual.

Vrij, A. (2008). *Detecting lies and deceit: Pitfalls and opportunities* (2nd ed.). Wiley.

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

Vrij, A., Granhag, P. A., & Porter, S. (2010). Pitfalls and opportunities in nonverbal and verbal lie detection. *Psychological Science in the Public Interest*, *11*(3), 89–121.https://doi.org/10.1177/1529100610390861

Vrij, A., Mann, S., & Fisher, R. P. (2006). Information-gathering vs accusatory interview style: Individual differences in respondents' experiences. *Personality and Individual Differences*, *41*(4), 589–599. https://doi.org/10.1016/j.paid.2006.02.014

Vrij, A., & Semin, G. R. (1996). Lie experts' beliefs about nonverbal indicators of deception. *Journal of Nonverbal Behavior*, *20*(1), 65–80. https://doi.org/10.1007/BF02248715

Wieser, M. J., & Brosch, T. (2012). Faces in context: A review and systematization of contextual influences on affective face processing. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00471

Wilcox, R. R., & Tian, T. S. (2011). Measuring effect size: A robust heteroscedastic approach for two or more groups. *Journal of Applied Statistics*, *38*(7), 1359–1368. https://doi.org/10.1080/02664763.2010.498507

Zhang, K., Frumkin, L. A., Stedmon, A., & Lawson, G. (2013). Deception in context: Coding nonverbal cues, situational variables and risk of detection. *Journal of Police and Criminal Psychology*, *28*(2), 150–161. https://doi.org/10.1007/s11896-013-9127-9

Zloteanu, M. (2020). Reconsidering facial expressions and deception detection. In A. Freitas-Magalhaes & J. Borod (Eds.), *Handbook of facial expression of emotion* (Vol. 3, pp. 238–284). FEELab Science Books & Leya.

Zloteanu, M., Bull, P., Krumhuber, E. G., & Richardson, D. C. (2021a). Veracity judgement, not accuracy: Reconsidering the role of facial expressions, empathy, and emotion recognition training on deception detection. *Quarterly Journal of Experimental Psychology*, *74*(5), 910–927. https://doi.org/10.1177/1747021820978851

Zloteanu, M., Krumhuber, E. G., & Richardson, D. C. (2021b). Sitting in judgment: How body posture influences deception detection and gazing behavior. *Behavioral Sciences*, *11*(6), 85. https://doi.org/10.3390/bs11060085

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Zloteanu, M., Salman, N. L., Krumhuber, E. G., & Richardson, D. C. (2022). Looking guilty: Handcuffing suspects influences judgements of deception. *Journal of Investigative Psychology and Offender Profiling*, *19*(3), 231–247. https://doi.org/10.1002/jip.1597