

Police officers' beliefs about, and use of, cues to deception

Clea Wright Whelan and Jacqueline M. Wheatcroft

Address correspondence to Clea Wright Whelan, Department of Psychology, University of Chester, Chester, CH1 4BJ clea.wright@chester.ac.uk (email)

Jacqueline M. Wheatcroft, Department of Psychological Sciences, Witness Group, Eleanor Rathbone Building, University of Liverpool, Liverpool, L69 7ZA, UK jacmw@liverpool.ac.uk (email)

Keywords: deception detection; cues to deception; police officers

Abstract

Meta-analytic findings indicate that people, including police officers, are generally poor at detecting low-stakes deception. Related to this, investigations of behaviours that people reportedly use to make truth/lie judgements tend to conclude that people rely on incorrect stereotypes. However, consistent findings suggest that police officers are able to detect high-stakes deception; this implies that, at least in some contexts, police officers utilise reliable cues to deception. The research presented here was an investigation of cues to deception used by police officers ($N = 69$), when making veracity decisions about real world, high stakes communications. Data were collected on both free report cues, and also prescribed cues that were known (from previous research), to discriminate between liars and truth-tellers in the communications that the police officers observed. Officers free reported using cues related to verbal content, emotion, body language, eyes, vocal cues, and external cues. Most prescribed cues were self-reportedly used correctly by large majorities of the officers, suggesting that they may not rely on inaccurate stereotypes. Self-report use of categories of free report cues, and prescribed cues, was not related to accuracy in detecting deception. As people may not always be aware of the behaviours on which their judgements are based, the relationships between some of the behaviours actually displayed in the communications, and group accuracy in detecting deception in those communications, were also investigated. Group accuracy was related to the presence of subjective, emotion-related cues in the communications.

Being deceived is a normal part of human life; indeed, some studies suggest that most people are probably lied to several times a day (DePaulo, Kashy, Kirkendol, Wyer & Epstein, 1996; Serota, Levine & Boster, 2010). However, findings consistently indicate that people are generally poor lie detectors; for example, Bond and DePaulo's meta-analysis (2006) found that average accuracy in detecting deception was only 54%, which is barely above chance. There are two main overarching reasons why this might occur. For deception to be successfully detected, a) there must exist observable behaviours that differ between liars and truth-tellers; and b) people must know what those behaviours are, so that they can correctly utilise them. Failure or weakness in either or both of these will limit the ability to detect deception. The detection of deception is, of course, not limited to the use of behavioural signals, and may occur through investigative activities (for example, gathering evidence). The focus here, however, is on observable behaviours.

With regard to knowledge of deception cues, investigations of behaviours that people reportedly use to make veracity judgements have tended to conclude that people generally rely on incorrect stereotypes of non-verbal deceptive behaviour. For instance, the Global Deception Research Team (2006) found that, across 58 countries, the most commonly reported cue that people thought was related to deception was gaze aversion (64% of participants), followed by nervousness (28% of participants), and then body movements, facial expression, and facial colour. Similarly, in their review of the topic, Stromwall, Granhag and Hartwig (2004) report that the most commonly used non-verbal cues to deception are, again, gaze aversion, and signs of nervousness (shifting position, more illustrators, more self-manipulations, more arm/hand movements, more leg/feet movements, more blinking, and a higher vocal pitch). Importantly, as both sets of researchers note, these findings contrast with those reported in DePaulo et al.'s (2003) comprehensive meta-analysis on actual cues to deception; for example, the evidence suggests that liars do not avert their gaze more than truth tellers, and liars make fewer arm/hand/finger movements, fewer leg/feet movements, and fewer illustrators. Hence, people's beliefs about non-verbal cues to deception appear to bear very little resemblance to behaviours that have actually been found to discriminate between liars and truth tellers.

In contrast, there appears to be some evidence that verbal and para-verbal behaviours reportedly used by people in their veracity decisions, may have more overlap with those actually found to be related to deception. For example, compared to truth-

tellers, liars are believed to be less coherent, to produce more 'uhs' and pauses, to be less consistent, less plausible and direct, provide fewer details and shorter stories, make fewer self-references, and provide more negative statements and irrelevant information (The Global Deception Research Team, 2006; Stromwall et al. , 2004). Most of these behaviours have been reported by researchers as related to deception (see, for example, DePaulo et al., 2003). Nevertheless, there are also a number of verbal behaviours that have been shown to relate to deception, that people do not commonly report using when making veracity judgements; for instance, lies tend to contain less temporal, spatial, and perceptual information, and liars produce more chronologically correct stories (DePaulo et al., 2003). Consequently, despite evidence of an overlap for some verbal behaviours, researchers have generally concluded that overall, for both non-verbal and verbal behaviours, 'People's beliefs about deception cues simply are not very realistic' (Stromwall et al., 2004, p.233). This conclusion seems to sit comfortably with Bond and DePaulo's (2006) finding that people are generally poor at detecting deception.

For some groups of professionals, such as police officers and other legal professionals, this issue is potentially extremely important as, for them, making decisions about veracity is an everyday occurrence which may have serious and far reaching consequences. In view of this, some researchers have specifically investigated the beliefs of these 'professional lie catchers'. A number of studies have found that there appear to be few differences between the cues that professional lie catchers report, and those reported by lay people (see, for example, Akehurst, Kohnken, Vrij & Bull, 1996; Vrij & Semin, 1996). In their review of professional lie catchers' expressed beliefs about cues to deception, Stromwall et al. (2004) conclude that such individuals still incorrectly rely predominantly on gaze aversion and nervousness in making veracity judgements. If this is the case, it is not surprising to find, as some investigators have, that professional lie catchers are no more accurate at detecting deception than lay people, (see, for example, DePaulo & Pfeifer, 1986; Ekman & O'Sullivan, 1991; Meissner & Kassin, 2002; Vrij, 1993).

Nevertheless, these findings seem to contrast with a developing body of more recent research indicating that, in some situations, police officers are consistently accurate at detecting deception at rates significantly above chance. Importantly, in this body of research, police officers made veracity judgements of real life, high stakes contexts (people in police suspect interviews for serious crimes, and people making public pleas for help with

missing or murdered relatives). Using stimulus materials of this kind, accuracy rates between 64% and 72% have been reported (Mann & Vrij, 2006; Mann, Vrij & Bull, 2004; Mann, Vrij & Bull, 2006; Vrij & Mann, 2001; Vrij, Mann, Robbins & Robinson, 2006; Wright Whelan, Wagstaff & Wheatcroft, 2015a). In using high stakes situations, the studies differ from those of most previous studies, which have tended to focus on low stakes lies produced in the laboratory; this suggests that the nature of the stimulus materials may have an important effect on the accuracy of observers. Indeed, some researchers have long argued that the majority of deception studies have been limited by poor ecological validity, and that there are very likely to be differences between cues to deception exhibited in low stakes situations, and those exhibited in high stakes situations (for example, DePaulo & Morris, 2004; Frank & Svetieva, 2012; Granhag & Stromwall, 2004; Porter & ten Brinke, 2008; Shuy, 1998; Vrij, 2004). An important implication of these findings is that the police officers were utilising reliable cues to deception to make their decisions about veracity in those high stakes contexts. Therefore, the findings in the literature that police officers may have incorrect stereotypical beliefs about cues to deception, may be relevant only to cues in low stakes situations. In fact, in De Paulo et al.'s meta-analysis of cues to deception, which has been used by researchers to test the validity of beliefs about cues to deception, of the 120 studies included, only four reportedly used a 'naturalistic' paradigm. These considerations also raise the possibility that the same stereotypical beliefs might be more accurate when applied to highly motivated liars, as suggested by Stromwall et al. (2004).

With regard to the latter point, it can be noted that studies focusing on the accuracy of the beliefs of police officers about deceptive behaviours have tended to rely on comparisons with behaviours alleged to differentiate deceptive from non-deceptive behaviours in the experimental literature; i.e. studies that have used mainly low stakes materials (Akehurst et al., 1996; Vrij & Semin, 1996). Indeed, a number of findings now suggest that the cues available and useful in the kinds of real-life contexts in which police officers work may differ from those available and useful in low stakes, laboratory-based situations. For example, gaze aversion consistently does not relate to deception in low stakes situations (DePaulo et al., 2003), but some studies report that gaze aversion is related to deception in certain high stakes contexts (Vrij & Mann, 2001; Wright Whelan, Wagstaff & Wheatcroft, 2014). Similarly, how smiling is best used as an indicator of deception may vary depending on the context: in studies predominantly of low stakes lies, DePaulo et al. (2003)

found that smiles were more likely to occur in honest communications, but in their investigation of pleas for help with missing or murdered relatives, ten Brinke and Porter (2012) found that deceptive pleaders were more likely to smile than truthful pleaders. In addition, a number of other behaviours, not previously identified in the literature and which may be context specific, have been related to deceptive communications in specific high stakes contexts (Harpster, Adams, & Jarvis, 2009; Hunt & Bull, 2012; ten Brinke & Porter, 2012; Wright Whelan et al., 2014; Wright Whelan, Wagstaff & Wheatcroft, 2015b). For example, in deceptive 911 homicide calls (in which the caller has actually murdered the victim), the caller is more likely to insult/blame the victim, and to use inappropriate politeness, than in truthful 911 calls (Harpster et al., 2009).

It may be the case, therefore, that the cues used by police officers to detect deception may be more applicable to high stakes situations of the kinds that they encounter in their everyday work, which might account for their improved performance in these situations. However, an important additional possibility is that, in high stakes situations, police officers may use cues other than those about which they have been asked. Typically, they are provided with a list of cues, and asked to indicate whether or not they use the cues, but being asked only to indicate beliefs about prescribed behaviours does not allow respondents to express beliefs about other possible cues. Mann et al. (2004) addressed this in a study that investigated free report cues that police officers reported using when watching excerpts from real-life suspect interviews. In line with other research, the cue most frequently reported by the police officers as being useful for lie detection was gaze aversion, followed by body movements. However, a wide range of cues were mentioned, including some related specifically to the account of events (for example, contradictions, minimising offence, and evidence), and some not commonly investigated (for example, posture, covering face, confidence, defensive). After the researchers had grouped the cues into categories, it was found that good lie detectors reported relying more on verbal story cues, and less on non-verbal stereotypical cues (gaze, fidget, and self-manipulation), than did poor lie detectors.

Furthermore, Mann et al. (2004) reasoned that people may not be aware of the cues that they are using to make their veracity decisions, and so coded some of the behaviours of the suspects in the videos, and investigated whether these coded behaviours were related to the veracity decisions made by the observers. This analysis showed that good lie

detectors were more likely to judge a suspect as deceptive if he or she used fewer illustrators (less use of hands), even though relatively few officers reported using this cue. As a decrease in illustrators has been shown to be related to deception, in both low stakes (DePaulo et al., 2003) and high stakes contexts (Koper & Sahlman, 1991), it seems that the more accurate police officers were using a reliable cue, without necessarily realising that they were doing so.

In a recent meta-analysis, Hartwig and Bond (2011) investigated whether deception detection is generally poor due to the reliance on invalid cues, or whether the general weakness of valid cues limits accuracy. DePaulo et al. (2003) report that the median effect size for the cues that differed between truthful and deceptive communications in their meta-analysis was just $d = 0.10$, suggesting that differences in behaviour between liars and truth-tellers are small and so may be difficult to discern. Hence, Hartwig and Bond's analysis (2011) looked not at self-report use of cues to deception, but the behaviours actually produced by senders, and how these related to judgements of deception by observers. The two behaviours most strongly related to judgements of deception were appearing incompetent, and statements in which events are not placed in context, though other large effects were found for fidgeting with objects, sounding uncertain, and appearing ambivalent or indifferent. It should be noted that some of the cues included in the analysis have only been investigated in two studies (e.g. context, and fidgeting with objects), and that although the effect size was large, it was not significant. Judgements of honesty were associated with sounding immediate, a pleasant face, being cooperative and involved, and statements that seem plausible, realistic and spontaneous. Not only do these behaviours differ to those generally cited in most self-report studies, but they also align with those found to be valid in DePaulo et al.'s meta-analysis (2003). The researchers concluded, therefore, that the generally accepted conclusion that people rely on invalid cues when making veracity judgements may have been over-stated, and it is primarily the weakness and lack of availability of cues that hampers accuracy. One implication of this is that lie detection accuracy will most likely be improved in situations that magnify potential differences between liars and truth tellers. If so, this might account, at least in part, for the higher lie detection accuracy rates reported in studies using high stakes materials, as one might expect the factors that underlie the production of cues to deception to be exaggerated in high stakes situations. Hartwig and Bond's (2011) findings could also be construed as further

supporting the idea that intuitive judgments about deceptive behaviour (of which people may not be aware) can be more accurate than those based on explicit beliefs.

In view of these considerations, the present study sought to address a number of questions concerning police officers' beliefs about cues to deception, and the actual behaviours that they use to make veracity judgements, in high stakes contexts. The first of these was, which cues do police officers self-reportedly use when making veracity judgements? Second, is there a relationship between these self-reported cues and accuracy in detecting high stakes lies? Mann et al. (2004) found that self-report use of story cues was related to accuracy, but it is useful to extend this research in to a different forensic context; if story cues are not available, are police officers able to be flexible and use other, more relevant, cues (for example, in appeals for missing and murdered relatives, commonly the speaker does not provide an account of events)? Third, some cues have been found to discriminate between liars and truth-tellers in the specific context of pleas for help with missing and murdered relatives (Wright Whelan et al., 2014; Wright Whelan et al., 2015b); do police officers report using these cues when viewing this type of material? Fourth, is use of these cues related to accuracy? Finally, if it is the case that people may be unaware of the cues that they are using to make correct veracity judgements, is there a relationship between independently coded behavioural measures in the pleas, and the accuracy with which these pleas are categorised as truthful or deceptive by police officers? The aim of the present study was to investigate these issues.

Method

Participants

Sixty nine police officers from a large UK police service took part in the study. Originally 70 were recruited, but one officer was required to leave before completing all aspects of the study, and so his partial data were removed. The participants were recruited using opportunity sampling, and participated voluntarily. There were 50 male police officers and 19 female police officers. Ages ranged from 26 to 53 years ($M = 33.90$, $SD = 6.86$), and years of experience as a police officer ranged from two to 29 years ($M = 11.75$, $SD = 6.25$).

Materials and Procedure

Participants attended the study in a psychology laboratory at a University in 10 groups, ranging in size from three participants to 14 participants. This variation in group size was a result of the number of police officers available to participate at any particular session, and did not affect the running of the study.

The study used the same real life, high stakes stimulus materials detailed by the authors elsewhere (Wright Whelan et al., 2015a). In brief, these were video footage of real life pleas for help with missing or murdered relatives. Videos of 32 pleaders were used, 16 of whom were truthful (the pleader was not involved in the death or disappearance of the relative), and 16 of whom were deceptive (the pleader was involved in the death or disappearance of the relative). Cases classified as deceptive were those in which the pleader had been convicted of being involved in the death or disappearance of their relative, based on very strong evidence (for example, CCTV footage of the accused disposing of the body of the victim). Cases classified as being truthful were those in which somebody else had been convicted of the murder of the relative, or the relative had not been the victim of foul play (for example, the relative returned after having voluntarily disappeared), again based on very strong evidence. Issues relating to the utilisation of these videos, including establishing ground truth, have been extensively addressed elsewhere (Wright Whelan et al., 2014). Cases were selected from USA, Canada, Australia, New Zealand, and the UK; cases from the UK were not recent and high profile.

Each participant was given an instruction and response sheet. Participants were informed that they would be viewing short video clips of people pleading for help with missing or murdered relatives, and would be asked to decide whether they thought each person was lying (i.e. was actually involved in the death or disappearance of their relative), or telling the truth (i.e. was not involved in the death or disappearance of their relative). Participants were asked to check a box on the response sheet if they were familiar with the person featured in the clip, or the outcome of the case, and to not complete the section of the response sheet for that appeal. Participant familiarity with cases ranged from zero to six cases ($M=2.91$, $SD= 1.58$). For each group, the appeals were shown in a different, randomised order. This was followed by a free report section, in which participants were asked the open-ended question 'What verbal or nonverbal cues did you use most in this study to decide whether the people in the video clips were lying or telling the truth?'

Participants were asked to indicate whether they believed the reported cues related to honesty or deception.

Participants then completed the final part of the response sheet, on which they were asked to indicate whether or not they used prescribed cues to come to their veracity decisions. If they used a prescribed cue, they were asked to indicate whether they thought it indicated honesty or deception. Twenty two cues were included in this section, all had been identified in previous research as discriminating between truthful and deceptive pleaders in the stimulus sample used in the present study. For example, on the basis of frequency count data, gaze aversion, and use of equivocal language were related to deception, whereas related to honesty were verbal expressions of positive emotion expressed towards the relative, expressions of hope that the relative may still be alive (in cases in the relative was missing and not already known to be dead), and references to norms of emotion or behaviour (Wright Whelan et al., 2014). In a further study, observers' subjective impressions and responses to pleaders in this sample were investigated by observers rating their agreement with a number of statements about each pleader (Wright Whelan et al., 2015b). It was found that, compared to truthful pleaders, deceptive pleaders were judged to display more fake emotion and fake facial expressions, to put on a performance, appear unnatural, and seem creepy. Moreover, observers had less sympathy for them and disliked them more than truthful pleaders. In contrast, truthful pleaders were regarded as being more sad, having a sadder expression in their eyes, being more genuine and heartfelt, urgent, plausible, and normal, using a more personal and expressive voice, having a voice quivering with emotion, and observers said they felt the pleaders' pain, and felt sorry for them.

Results

Following the same process used in previous studies (Mann et al., 2004; Masip & Herrero, 2015), the free-report behaviours provided in response to the open-ended question were sorted into 28 cues, see Table 1. The aim was to achieve a list of cues that reflected the diversity of answers, but was still manageable (there were at least two responses in each group). Each response of every participant was coded in terms of one of the 28 cues (no response was assigned to more than one category). A second, independent coder then coded a random sample of 25% of the response sheets into the 28 cues. Inter-rater

reliabilities were reasonably high and ranged from $r = 1.00$ to $r = 0.69$ (though inter-rater reliability could not be computed for five cues, head position/movement, focus on self, plausibility, appearance, and circumstances, because at least one variable was constant due to floor effects). The data produced by the coder who had coded all responses were, therefore, used in all analyses.

The free-report cues were then further grouped into six categories based on general, observable themes. Although using by-category analyses risks losing some information that may have been retained in using by-cue analyses, it was favoured as a parsimonious option. The categories were eyes (as 86% of police officers mentioned eyes at least once, it was decided to categorise cues related to eyes separately from other body language cues), body language, emotion, vocal characteristics, verbal content, and external cues (see Table 1). The total number of times that each cue was mentioned by participants was calculated and expressed as a percentage of the total number of participants (see Table 1). The total number of cues each participant reported using in each category was also calculated.

(insert table 1 around here)

The 22 prescribed cues that had been identified in previous research were grouped into five categories; eyes (gaze aversion), emotion (fake emotion, fake facial expressions, putting on a performance, sad, sad eyes, genuine and heartfelt, urgency), vocal characteristics (personal and expressive voice, voice quivering with emotion), verbal content (equivocal language, positive emotion expressed towards the relative, expressions of hope, and references to norms of emotion or behaviour), and subjective global responses (unnatural, creepy, lack of sympathy, dislike, plausible, normal, feel the appellant's pain, feeling sorry for the appellant); see Table 2. The number of participants who reported using the prescribed cue in the 'correct' manner according to previous studies (for example, as being related to truthfulness if the cue had actually been found to relate to truthfulness) in each category was then calculated and expressed as a percentage of the total number of participants (see Table 2). The total number of cues each participant reported correctly using in each category was also calculated.

(insert table 2 around here)

A detailed examination of the accuracy findings for this sample has been presented elsewhere (Wright Whelan et al., 2015a); accuracy in detecting deception ranged between 56% and 91%, with a mean of 72.20% ($SD=8.18$). On the basis of these previous findings,

police officers were divided into two ability groups; good lie detectors were those who had scored above the mean ($n=37$, $M=78.49\%$, $SD=4.31$), and poor lie detectors were those that scored below the mean ($n=32$, $M=64.94\%$, $SD=4.86$).

To investigate potential differences in the use of free report cues between good and poor lie detectors, six independent t-tests (one per cue category), were performed. These analyses were favoured over an ANOVA, as there were differences in the number of cues per category. Levene's test was significant for the cue categories eyes and emotion, and so for these two cue categories, t statistics not assuming homogeneity of variance are reported. No significant differences in use of cue categories were found between good and poor lie detectors: eyes, $t(58.72) = 0.70$, $p = .49$; body language, $t(67) = -0.12$, $p = .90$; emotion, $t(49.93) = 0.86$, $p = .40$; vocal characteristics, $t(67) = 0.25$, $p = .81$; verbal content, $t(67) = -0.87$, $p = .39$; external cues, $t(67) = -0.53$, $p = .60$. The cue categories vocal characteristics, verbal content, and external cues had skewed distributions, therefore non-parametric Mann-Whitney U tests were performed on these cue categories as a check, with equivalent results (all $ps > .40$). See Table 3 for means and SDs of cue use for each category.

(insert table 3 around here)

To investigate whether there were differences in the use of the prescribed cues by good and bad lie detectors, an ANOVA with lie detection ability as the between subjects factor was conducted on the total number of cues each participant reported correctly using each category except the eye cue. Lie detection ability did not have significant effects for the emotion cues, vocal characteristic cues, verbal content cues, or subjective global responses (all $ps > .10$). As there was only one cue in the eye category (gaze aversion), a point-biserial correlation was performed on overall accuracy and correct use of the eye cue, which was not significant, $r_{pb} = .01$, $p = .93$. See Table 4 for means and SDs of cue use for each category.

(insert table 4 around here)

Following suggestions that people may use reliable cues to detect deception without necessarily being aware of the cues that they are using (Hartwig & Bond, 2011; Mann et al., 2004), the final stage of the analyses investigated the relationships between pre-rated cues known to be available in the pleas, and group accuracy in judging the pleas as truthful or deceptive. Thus, the following analysis was an examination of actual use of cues, rather than reported use of cues. The cues investigated in this analysis were those prescribed cues

(Table 2) previously found to discriminate between truthful and deceptive pleas for help with missing or murdered relatives (Wright Whelan et al., 2014; Wright Whelan et al., 2015b). Using a previously established procedure, and previous data (Wright Whelan et al., 2014; Wright Whelan et al., 2015b), for each individual pleader, scores on each cue were compared with the overall cue mean. For deceptive pleas, cue scores that fell above the cue mean on cues to deception were classified as a 'hit', and cues that fell below the cue mean on cues to truthfulness were classified as a 'hit'. For truthful pleas, cue scores that fell below the cue mean on cues to deception were classified as 'hits', and cues that fell above the cue mean on cues to truthfulness were classified as 'hits'. In this way, for each plea, the number of 'hits' was calculated for each category of cue (eyes, emotion, vocal characteristics, verbal content, subjective global responses). The accuracy data were transposed to provide group accuracy for each of the 32 pleaders.

A series of correlations was then conducted to investigate the relationships between group accuracy on each pleader, and the total number of hits in each cue category for each pleader. This was an examination of actual use of cues, rather than reported use of cues. Group accuracy on the pleaders was related to emotion cues ($r = .53, p = .002$), to vocal characteristics ($r = .41, p = .021$), and to subjective global responses ($r = .48, p = .005$); for these cue categories, group accuracy increased with number of 'hits' on the cues. No relationships were found between gaze aversion and group accuracy ($r = .13, p = .60$), or between verbal content and group accuracy ($r = .19, p = .30$).

Discussion

Overall, the findings from the present study suggest that police officers' beliefs about cues to deception may not be inaccurate stereotypes. The high mean accuracy of the officers in classifying pleaders as truthful or deceptive, and the high use of cues known to discriminate between truthful and deceptive pleaders in the stimulus materials, suggest that officers use some reliable cues to make veracity decisions. In line with previous research (The Global Deception Research Team, 2006; Mann et al., 2004; Stromwall et al., 2004), the most often used self-report cue was eye contact. This finding is usually reported as demonstrating inaccuracy of beliefs about cues to deception, but as previously argued, findings from mostly low stakes research may not necessarily be applicable in some high stakes contexts;

indeed, in the stimulus materials used in the present study, gaze aversion was related to deception. Furthermore, as previously discussed, smiles may be indicative of either honesty or deception, depending on the context. In the present study, five participants mentioned smiles, and all related them to deception; this would be classified as an inaccurate belief if compared to meta-analytic findings, but is an accurate belief in the context used in this study. It is possible, then, that officer's understanding of behaviours related to deception may be flexible and take account of specific contexts, and that this more nuanced approach is not captured by research that takes a more blanket approach to classifying cues as correct or incorrect.

Although previous research suggests that nervousness often follows gaze aversion as a reported cue to deception, in the present study, relatively few officers specifically self-reported using nervousness as a cue; although it is unknown if some officers may have included this in the unspecified body language cue (which was also commonly mentioned). The second most spontaneously reported cue to deception was face touching, with almost half of participants mentioning this behaviour. It may be that face touching was regarded as being indicative of nervousness, but frequent reference to covering or blocking the face or mouth suggests that the behaviour may also have been regarded as indicative of distancing or hiding. Three emotion-related cues were each spontaneously reported by almost a third of the officers (fake emotion, lack of emotion, and genuine emotion), and a vocal characteristic related to emotion (discernible emotion in voice) was also relatively frequently mentioned, suggesting that emotion was important in the decision-making of the officers. Interestingly, in Mann et al.'s study (2004) of spontaneously reported cues, there was no category of cues related to emotion. Furthermore, verbal cues in that study were categorised as story cues, whereas in the present study verbal content cues were not necessarily related to the story. These differences between the findings may be related to the nature of the stimulus materials used; in the present study, the videos were a specific context that might be expected to be highly emotional, and as pleas do not necessarily contain a story, story cues may not be available. Again, this suggests that police officers may be flexible in their approach, and consider the relevance of specific contexts. This suggestion is further supported by the inclusion of a number of behaviours that are not usually investigated in deception research and that may be context specific (for example, focus on relative, and using past tense)

External cues were rarely mentioned, a finding that it is not surprising considering the nature of the stimulus materials, in which very few external cues were available. Vocal characteristics were spontaneously mentioned relatively infrequently; however, the descriptive statistics indicate that when specifically asked about vocal cues, a large majority of officers reported using them, and the presence of vocal cues in the pleas was positively related to accurate group classification of the plea. It would appear, then, that officers were using vocal information, even though they were spontaneously reporting using cues in this category comparatively infrequently.

In the investigation of police officers' reported use of prescribed cues previously found to discriminate between truth and deception in the stimulus materials used in the present study, all the prescribed cues were used correctly by at least half of the officers, and most were used correctly by a very large percentage. All the officers in the present study reported using fake emotion and fake facial expression as cues to deception, and the descriptive data indicate that very large percentages of the officers reported using emotion cues, vocal characteristics, and subjective global responses. A large percentage of officers also reported using equivocal language as a cue to deception, which is a behaviour that has been demonstrated to relate to deception not only in the type of stimulus materials used in the present study (ten Brinke & Porter, 2012; Wright Whelan et al., 2014), but also in other high stakes contexts (for example, Adams & Jarvis, 2006; Fuller, Biros, Burgoon & Nunamaker, 2013). In contrast to the spontaneously reported cues, gaze aversion was one of the least frequently mentioned of the prescribed cues (although still used by the majority of officers), whereas emotion-related cues and vocal characteristics were very frequently mentioned. There were no relationships between categories of prescribed cues, and deception detection accuracy, but overall there was very high use of cues known to occur in this data set, including some cues that are likely to be context-specific and so not typically investigated in deception research (for example, positive towards relative, norms of emotion/behaviour, lack of sympathy, feel pleader's pain, feel sorry for pleader). This fits with Hartwig and Bond's contention that "people rarely rely on the wrong cues" (2011, p.1), and again suggests that officers' approaches to detecting deception may be more nuanced than some of the research in this area.

The lack of difference in reported use of cues between good and poor lie detectors raises the question of how good lie detectors were able to achieve higher accuracy.

Informal, post-study interviews with three of the most accurate participants (all 80%+), suggested that they looked for disparity between channels of behaviours (verbal, nonverbal, and vocal), and deviations from norms and expected behaviour, but also that they applied a subtle, nuanced, and often idiosyncratic approach that may be difficult to capture when investigating the use of individual cues. This is in line with suggestions that accuracy may be increased by using behaviours in relation to each other, rather than behaviours in isolation (DePaulo et al., 2003; O'Sullivan & Ekman, 2004), and that some manifestations of deception are idiosyncratic and difficult to capture in quantitative research (Porter & ten Brinke, 2010). This is clearly an area that would benefit from further research.

Pleas that contained more of the emotion cues, vocal characteristic cues, and subjective global response cues (previously found to discriminate between truthful and deceptive pleas), were more accurately categorised by the group, with moderate to strong relationships: this suggests that the officers may have been correctly using these categories of cues to inform their decisions. Furthermore, these are the same three categories of cues that the descriptive data indicate the police officers reported frequently using when asked about their use of prescribed cues. Looking at the overall picture from the descriptive data, and the significant relationships, it appears that emotion-related cues, vocal characteristics, and emotional responses and impressions evoked in observers (all subjectively measured), were an important element in the decision-making processes of the participants. This is in line with meta-analytic findings, which suggest that truths and lies may be more powerfully discriminated by using subjective measures rather than objective ones (DePaulo et al., 2003; Hartwig & Bond, 2011). Interestingly, although eye contact was the most frequently reported cue to deception spontaneously mentioned by the officers, there was no relationship between gaze aversion in the pleas, and group accuracy in categorising the pleas as truthful or deceptive. This suggests that officers may use this behaviour less frequently than they believe. Furthermore, the lack of relationships between gaze aversion and verbal content cues in the pleas (both categories of cues which had been measured with frequency counts rather than subjectively), and group accuracy in categorising the pleas as truthful or deceptive, has two implications: first, there was reliable information available in the pleas that the participants were not generally using; and second, that participants were using subjective impressions rather than specific, individual behaviours that may be more objectively measured, to make correct veracity decisions. However, when

using subjective cues such as fake emotion, and performance, it may be that people use constellations of behaviours to formulate these more global judgements; these types of cues are very likely to include combinations of behaviours, rather than a single verbal or nonverbal behaviour. It is possible, then, that specific, individual, measurable behaviours form part of these subjective global impressions; it would be interesting for future research to attempt to unpick this.

Despite leading researchers noting that context may be important in the investigation of cues to deception (DePaulo et al., 2003), there is a tendency for it to be overlooked in deception research (although, see Harpster et al., 2009; Hunt & Bull, 2012; ten Brinke & Porter, 2012; Wright Whelan et al., 2014; Wright Whelan et al, 2015b). However, based on the present findings, it would appear that police officers are sensitive to the role that context may play on behaviours indicative of deception. When asked to spontaneously report cues used to detect deception after viewing pleas for help with missing and murdered relatives, officers mentioned cues that may apply across contexts (for example, vocal pitch, directness), but also context-specific cues, and cues that are not routinely investigated in deception research (for example, focus on relative, focus on story, using past tense). Moreover, when asked about their use of cues that have been identified as discriminating between truthful and deceptive pleas in previous research, officers overwhelmingly reported using the cues. It would appear that asking participants only about their use of cues found to relate to deception in very general circumstances, may not capture the flexibility and the accuracy with which people use cues to deception in the real world.

In sum, the findings in the present study offer general support for Hartwig and Bond's (2011) proposition that people may rely on useful behaviours when detecting deception. In the investigation of spontaneously reported cues, the use of prescribed cues, and the relationship between available cues and group accuracy in detecting deception, emotion and subjective impressions played a prominent role; this suggests that there is some useful information available in these types of cues, and that it is correctly used by some observers.

References

- Adams, S. H. & Jarvis, J. P. (2006). Indicators of veracity and deception: an analysis of written statements made to police. *The International Journal of Speech, language and the Law*, 13 (1), 1-22
- Akehurst, L., Kohnken, G., Vrij, A. & Bull, R. (1996). Laypersons' and police officers' beliefs regarding deceptive behaviour. *Applied Cognitive Psychology*, 10, 461-471
- Bond, C. F. & DePaulo, B. M. (2006). Accuracy of deception judgements. *Personality and Social Psychology Review*, 10 (3), 214-234. doi: 10.1207/s15327957pspr1003_2
- DePaulo, B. M., Kashy, D. A., Kirkendol, S. E., Wyer, M. M. & Epstein, J. A. (1996). Lying in everyday life. *Journal of Personality and Social Psychology*, 70, 979-995
- DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K. and Cooper, H. (2003). Cues to Deception. *Psychological Bulletin*, 129 (1), 74-118
- DePaulo, B. M. & Morris, W. L. (2004). Discerning lies from truths: behavioural cues to deception and the indirect pathway of intuition. In Granhag, P. A. & Stromwall, L. A. (Eds), *The Detection of Deception in Forensic Contexts* (pp.15-40). Cambridge: Cambridge University Press
- DePaulo, B. M. & Pfeifer, R. L. (1986). On-the-job experience and skill at detecting deception. *Journal of Applied Social Psychology*, 16, 249-267.
- Ekman, P. & O'Sullivan, M. (1991). Who can catch a liar? *American Psychologist*, 46, 913-920.
- Frank, M. G. & Svieteva, E. (2012). Lies worth catching involve both emotion and cognition: Commentary on Vrij and Granhag (2012). *Journal of Applied Research in Memory and Cognition*, 1 (2), 131-133
- Fuller, C. M., Biros, D. P., Burgoon, J. & Nunamaker, J. (2013). An Examination and Validation of Linguistic Constructs for Studying High-Stakes Deception. *Group Decision and Negotiation*, 22 (1), 117-134
- The Global Deception Research Team (2006). A world of lies. *Journal of Cross-Cultural Psychology*, 37 (1), 60-74. doi:10.1177/0022022105282295
- Granhag, P. A. & Stromwall, L. A. (2004) (Eds). *The Detection of Deception in Forensic Contexts* (pp.15-40). Cambridge: Cambridge University Press
- Harpster, T., Adams, S. H. and Jarvis, J. P. (2009). Analyzing 911 homicide calls for indicators of guilt or innocence). *Homicide Studies*, 13 (1), 69-93

- Hartwig, M. & Bond, C. F. Jr (2011). Why do lie-catchers fail? A lens model meta-analysis of human lie judgements. *Psychological Bulletin*, 137 (4), 643-659
- Hunt, L. & Bull, R. (2012). Differentiating genuine and false rape allegations: a model to aid rape investigations. *Psychiatry, Psychology and Law*, 19 (5), 682-691
- Koper, R. J. & Sahlman, J. M. (1991). The behavioral correlates of real-world deceptive communication. Paper presented at the Annual Meeting of the International Communication Association. Retrieved online 13 March 2009, from http://www.eric.ed.gov/ERICDocs/data/ericdocs2sql/content_storage_01/0000019b/80/24/18/ae.pdf
- Mann, S. & Vrij, A. (2006). Police officers' judgement of veracity, tenseness, cognitive load and attempted behavioural control in real-life police interviews. *Psychology, crime & Law*, 12 (3), 307-319
- Mann, S., Vrij, A. & Bull, R. (2004). Detecting true lies: Police officers' ability to detect suspects' lies. *Journal of Applied Psychology*, 89, (1), 137-149. doi:10.1037/0021-9010.89.1.137
- Mann, S., Vrij, A. & Bull, R. (2006). Looking through the eyes of an accurate lie detector. *The Journal of Credibility Assessment and Witness Psychology*, 7 (1), 1-16
- Masip, J. & Herrero, C. (2015). Police detection of deception: Beliefs about behavioural cues to deception are strong even though contextual evidence is more useful. *Journal of Communication*, 65, 125-145. doi: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, 469-480.
- Miller, G. R. & Stiff, J. B. (1993). *Deceptive Communication*. Newbury Park, CA: Sage
- O'Sullivan, M. & Ekman, P. (2004). The wizards of deception detection. In Granhag, P. A. & Stromwall, L. A. (Eds), *The Detection of Deception in Forensic Contexts* (pp.269-286). Cambridge: Cambridge University Press
- Porter, S. and ten Brinke, L. (2008). Reading between the lies: Identifying concealed and falsified emotions in universal facial expressions. *Psychological Science*, 19 (5), 508-514
- Porter, S. and ten Brinke, L. (2010). The truth about lies: what works in detecting high-stakes deception? *Legal and Criminological Psychology*, 15, 57-75
- Serota, K. B., Levine, T. R. & Boster, F. J. (2010). The prevalence of lying in America: Three

- studies of self-reported lies. *Human Communication Research*, 36, 2-25.
doi:10.1111/j.1468-2958.2009.01366.x
- Shuy, R.W. (1998). *The Language of Confession, Interrogation and Deception*. Thousand Oaks, California: Sage Publications Inc
- Stromwall, L. A , Granhag, P. A, & Hartwig, M. (2004). Practitioners' beliefs about deception. In Granhag, P. A. & Stromwall, L. A. (Eds), *The Detection of Deception in Forensic Contexts* (pp.15-40). Cambridge: Cambridge University Press
- ten Brinke, L., & Porter, S. (2012). Cry me a river: Identifying the behavioral consequences of extremely high-stakes interpersonal deception. *Law and Human Behavior*, 36 (6), 469-477
- Vrij, A. (1993). Credibility judgments of detectives: The impact of nonverbal behavior, social skills, and physical characteristics on impression formation. *Journal of Social Psychology*, 133 (5), 601-610
- Vrij, A. (2004). Guidelines to catch a liar. In Granhag, P. A. & Stromwall, L. A. (Eds), *The Detection of Deception in Forensic Contexts* (pp.287-314). Cambridge: Cambridge University Press
- Vrij, A. & Mann, S. (2001). Telling and detecting lies in a high-stake situation: the case of a convicted murderer. *Applied Cognitive Psychology*, 15, 187-203
- Vrij, A., Mann, S., Robbins, E. & Robinson, M. (2006). Police officers ability to detect deception in high stakes situations and in repeated lie detection tests. *Applied Cognitive Psychology*, 20, 741-755. doi:10.1002/acp.1200
- Vrij, A. & Semin, G. R. (1996). Lie experts' beliefs about nonverbal indicators of deception. *Journal of Nonverbal Behavior*, 20, 65-80
- Wright Whelan, C., Wagstaff, G. F. & Wheatcroft, J. M. (2014). High stakes lies: verbal and nonverbal cues to deception in public appeals for help with missing or murdered relatives. *Psychiatry, Psychology and Law*, 21 (4), 523-537
- Wright Whelan, C., Wagstaff, G. F. & Wheatcroft, J. M. (2015a). High stakes lies: police and non-police accuracy in detecting deception. *Psychology, Crime and Law*, 21 (2), 127-138
- Wright Whelan, C., Wagstaff, G. F. & Wheatcroft, J. M. (2015b). Subjective cues to deception/honesty in a high stakes situation: an exploratory approach. *The Journal of Psychology: Interdisciplinary and Applied*, 149 (5), 517-534

Tables

Table 1: Self-report cues; cue categories, cues, descriptions of cues, and frequency of reported cue use (expressed as a percentage of participants)

Category	Cue	Description	Freq.
Eyes	Eye contact	Including aversion, direct, covering eyes	75%
	Eye movements	Including blinking, flitting, darting	32%
	Eyes - general	Including unspecified, and lack of emotion	6%
Body language	Face touch	Including covering/blocking face/mouth	49%
	Face	Including expressions, smiling, lip-licking	16%
	Nervous	Including shifty, fidgeting, excessive hand gestures	17%
	Head	Including position, movement	6%
	Body language – general	Unspecified, or any other	40%
Emotion	Fake emotion	Including crocodile tears, too much emotion, inconsistent emotion	33%
	Lack of emotion	Including indifference, coldness	28%
	Genuine emotion	Including real tears, naturalness	30%
	General emotion	Unspecified, or any other	14%
Vocal characteristics	Pitch	Including high voice	12%
	Discernible emotion	Including shaky voice, conviction in voice	19%
	Lack of emotion	Including steady voice, flat voice	6%
	Voice - general	Unspecified, or any other	9%
Verbal content	Directness	Including clearly expressed, direct, unequivocal	6%
	Focus on self	Including self-references, concern with self	6%
	Speech dysfluency	Including stumbling over words, repetition, mumbling	4%
	Focus on relative	Including hope of finding, positive description, pleading for return	13%
	Avoidant	Including question resistance, reading statement, avoid relative, vague, irrelevant, impersonal, rambling	29%
	Focus on story	Including too much detail, focus on own version of events	12%
	Plausibility	Including lack of plausibility, consistency	4%
	Tense	Using past tense for missing relative	13%

	Language – general	Unspecified, or any other	22%
External cues	Family signals	Including behaviour of present family members	4%
	Appearance	Including teeth, general appearance	3%
	Circumstances	Including age, relationship to victim, time elapsed	4%

Table 2: Prescribed cues; cue categories, cues, and frequency (percentage of participants who reported using the cue correctly)

Category	Cue	Frequency
Eyes	Gaze aversion	61%
Emotion	Fake emotion	100%
	Fake facial expression	100%
	Performance	99%
	Sad	93%
	Sad eyes	87%
	Genuine/heartfelt	91%
	Urgency	51%
Vocal characteristics	Personal/expressive voice	84%
	Voice quiver with emotion	81%
Verbal content	Equivocal language	84%
	Positive towards relative	64%
	Hope	57%
	Norms of emotion/behaviour	61%
Subjective global responses	Unnatural	80%
	Creepy	88%
	Lack of sympathy	78%
	Dislike	75%
	Plausible	84%
	Normal	70%
	Feel pleader's pain	87%
	Feel sorry for pleader	83%

Table 3: Self-report cues; means and SDs of cue use for each cue category, for good and poor lie detectors, and in total.

Cue category	Good lie detector Mean (SD)	Poor lie detector Mean (SD)	All participants Mean (SD)
Eyes	1.08 (0.55)	1.19 (0.69)	1.13 (0.62)
Body language	1.22 (1.00)	1.19 (0.93)	1.20 (0.96)
Emotion	0.97 (0.65)	1.16 (1.05)	1.06 (0.86)
Vocal characteristics	0.43 (0.65)	0.47 (0.57)	0.45 (0.61)
Verbal content	1.19 (1.20)	0.97 (0.86)	1.09 (1.05)
External cues	0.14 (0.35)	0.09 (0.30)	0.12 (0.32)

Table 4: Prescribed cues; means and SDs of cue use for each cue category, for good and poor lie detectors.

Cue category	Good lie detector Mean (SD)	Poor lie detector Mean (SD)
Eyes	0.89 (0.31)	0.88 (0.34)
Emotion	6.08 (0.80)	6.38 (0.71)
Vocal characteristics	1.65 (0.59)	1.66 (0.48)
Verbal content	2.78 (1.13)	2.50 (1.11)
Subjective global responses	6.51 (1.45)	6.41 (1.70)