



# The European Journal of Psychology Applied to Legal Context

<https://journals.comadrid.org/ejpalc>



## Truth or Lie: Ability of Listeners to Detect Deceptive Emergency Calls of Missing Children

Daniel E. O'Donnell<sup>1</sup>, Michelle C. Huffman<sup>2</sup>, Taylor E. Burd<sup>3</sup>, and Colleen L. O'Shea<sup>3</sup>

<sup>1</sup>Behavioral Analysis Unit 3, National Center for the Analysis of Violent Crime, Federal Bureau of Investigation, Quantico, Virginia, USA; <sup>2</sup>Behavioral Analysis Unit 5, National Center for the Analysis of Violent Crime, Federal Bureau of Investigation, Quantico, Virginia, USA; <sup>3</sup>Oak Ridge, Institute for Science and Education, Oak Ridge, Tennessee, USA.

### ARTICLE INFO

#### Article history:

Received 4 April 2024  
Accepted 18 June 2024

#### Keywords:

Emergency calls  
911  
Child  
Deception  
Homicide

### ABSTRACT

**Background:** Emergency calls may help law enforcement determine the proper response and provide investigative leads. Time may be wasted and appropriate resources misallocated if callers provide untruthful information. However, human ability to detect deception is generally weak. **Objectives:** We compared the abilities of law enforcement officers and non-law enforcement staff abilities to correctly identify truthful or deceptive emergency calls reporting missing children using Grice's maxims of communication (quantity, manner, relation, and quality of information). **Method:** Forty participants listened to 32 emergency calls reporting a missing child. Sixteen callers truthfully reported not knowing the child's whereabouts, and sixteen were responsible for killing the child before falsely reporting the child missing. Participants rated the quantity (insufficient, appropriate, excessive), manner (clear/orderly, unclear/disorderly), relation (relevant, irrelevant), and quality (truthful, deceptive) of information. Participants also provided a written narrative of their impressions of the call. **Results:** Accuracy in identifying truthful and deceptive callers was consistent with prior research, with sworn law enforcement slightly outperforming non-sworn staff. Participant agreement on Grice's maxims was poor. Ratings of quantity, manner, and relation of information predicted judgments of call quality, but were not associated with accurately identifying calls. Participant narratives describing reasons for judging a call to be truthful or deceptive were also not associated with accurate identification. **Conclusions:** Our findings do not support the use of Grice's maxims for determining deception in emergency calls. Although law enforcement officers outperformed non-sworn staff, both groups showed inconsistent rationales to support veracity judgments and relied on cues not associated with accuracy.

## Verdad o mentira: la capacidad de los teleoperadores para detectar llamadas falsas de emergencia de niños desaparecidos

### RESUMEN

**Antecedentes:** Las llamadas de emergencia pueden servir para que la policía responda adecuadamente además de proporcionar pistas a la investigación. Puede perderse tiempo y desperdiciarse recursos si los que llaman dan información errónea. No obstante, la capacidad humana para detectar el engaño es baja. **Objetivos:** Comparamos la capacidad de la policía y de otro personal para identificar adecuadamente llamadas de emergencia verdaderas o falsas referidas a niños desaparecidos utilizando las máximas de comunicación de Grice (cantidad, modo, relación y calidad de la información). **Método:** Se presentó a 40 participantes 32 llamadas de emergencia referidas a niños desaparecidos. Dieciséis de las personas que llamaban decían sinceramente que no sabían del paradero del niño y 16 eran responsables de haber matado al niño antes de informar falsamente de su desaparición. Los participantes valoraron la cantidad (insuficiente, adecuada, excesiva), modo (clara/organizadamente), relación (pertinente, no pertinente) y calidad (verdadero, falso) de la información. También facilitaron un relato escrito de sus impresiones acerca de la llamada. **Resultados:** La precisión en la detección de qué personas de las que llamaban decían la verdad y quiénes mentían concordaba con la de investigaciones previas, siendo la policía ligeramente mejor que el otro personal. El acuerdo de los participantes sobre las máximas de Grice era bajo. La precisión en la cantidad, modo y relación de la información predecía la valoración de la calidad de las llamadas, pero no guardaba relación con la precisión de la identificación de las llamadas. Los relatos de los participantes en los que describían los motivos por los que decían que una llamada era verdadera o falsa no guardaban relación con una identificación exacta de la calidad de la llamada. **Conclusiones:** Los resultados no avalan el uso de las máximas de Grice para detectar el engaño en las llamadas de emergencia. Aunque la policía era mejor que el otro personal, ambos grupos presentaban argumentos incoherentes que permitían emitir juicios de veracidad y descansaban en pistas que no estaban relacionadas con la precisión.

#### Palabras clave:

Llamadas de emergencia  
911 [112]  
Niño  
Engaño  
Homicidio

Cite this article as: O'Donnell, D. E., Huffman, M. C., Burd, T. E., & O'Shea, C. L. (2024). Truth or lie: Ability of listeners to detect deceptive emergency calls of missing children. *European Journal of Psychology Applied to Legal Context*, 16(2), 97-108. <https://doi.org/10.5093/ejpalc2024a9>

Correspondence: [deodonnell@fbi.gov](mailto:deodonnell@fbi.gov) (D. E. O'Donnell).

ISSN: 1889-1861/© 2024 Colegio Oficial de la Psicología de Madrid. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

The majority of children reported as missing each year in the United States result from non-criminal incidents, such as the child running away or becoming lost, and most children are recovered alive in relatively short order (Sedlak et al., 2017). Criminal incidents, such as sexually-motivated child abductions, are less common but may result in higher incidences of serious physical injury or death (Warren et al., 2020). A common feature among missing child cases is that law enforcement typically has little information about the circumstances when receiving initial reports. Early decisions must rely on preliminary details provided by reporting parties and these reports often come in the way of emergency calls.

Examination of information provided during missing child emergency calls may be helpful for first responders. However, misrepresentation of details by reporting parties or incorrect evaluation of the reporting parties' truthfulness may negatively impact the investigative decisions that follow. The current study assesses the ability of law enforcement officers and non-law enforcement personnel to identify truthful or deceptive emergency calls of missing children using Grice's maxims of conversation. We first describe various types of missing child investigations before discussing previous research into emergency calls and the ability to detect indicators of veracity and deception.

### Child Abduction

Perpetrators of most child abductions are caregivers who abduct children during custody disputes (Finkelhor et al., 1991; Sedlak et al., 2002). Most cases resolve with no criminal charges being filed against the caregivers (Grasso et al., 2001; Johnston & Girdner, 2001). These cases also differ from non-custodial abductions in that the perpetrators of custodial abductions are generally known by law enforcement and the reporting party when the abduction occurs (Hilts et al., 2015).

Perpetrators of non-custodial child abductions, on the other hand, are typically committed without the caregiver's knowledge by family acquaintances, strangers, or other relatives of the child. Motivations for these types of abductions also differ from the interpersonal conflict that is characteristic of custodial abductions. Examples include situations in which a female abducts a child to keep as her own (e.g., maternal desire), circumstances in which a child is abducted and held for financial gain (e.g., ransom), and incidents in which a child is abducted to satisfy sexual needs (e.g., sexual gratification; Beyer & Beasley, 2003; Boudreaux et al., 2000; Warren et al., 2016).

Initial reports of non-custodial child abductions often lack specific details, as caregivers may not possess knowledge of the abduction and may simply report the child as missing (Brown et al., 2006). Further, neither caregivers nor law enforcement may be able to identify eyewitnesses or determine an observable crime scene. The absence of detailed information heightens the need to develop substantive investigative leads to help recover the child, collect evidence, and prosecute the offender. Thus, reviewing the content of emergency calls may help investigators identify crucial elements early in an investigation to appropriately prioritize resources.

### False Allegations of Child Abduction

In certain cases, a caregiver with knowledge of the circumstances of the child's disappearance withholds this information from law enforcement. Cases in which a caregiver kills the child and disposes of the child's remains, or has knowledge of either but falsely reports the child as missing or abducted to conceal these details, are referred to as false allegations of child abductions (Canning et al., 2011). False allegation cases typically arise from a caregiver's ongoing physical abuse or neglect of the child or a caregiver's perception of the child being a burden. As such, the motivations leading to false allegation cases differ from the sexual motivation characteristic of many

non-custodial child abductions. Canning et al. (2011) describe other differences as well. For example, although male offenders are typical in non-custodial child abductions, females are common offenders in false allegation cases. Further, false allegation cases generally involve children five-years-old or younger, whereas sexually motivated child abductions usually involve children older than five.

Regardless of these differences, information available in the early stages of missing child investigations is generally limited and law enforcement must rely on this information to make critical timely decisions. Analyzing information contained in emergency calls may therefore assist law enforcement when deploying initial resources. As emergency calls can be recorded and preserved, the contents can be reviewed for investigative leads, as well as to corroborate truthful details and refute false information.

### Detecting Deception in Emergency Calls

Research into indicators of veracity and deception has expanded in recent years (Markey et al., 2022; Miller et al., 2020). Several of these studies have attempted to replicate findings of a study conducted by Harpster et al. (2009), in which the authors examined one hundred emergency calls made in the United States by persons reporting a homicide. Half of the calls were made by innocent parties, whereas the other half were made by individuals involved in the homicide but concealed their involvement from dispatchers. Harpster et al. (2009) theorized that the content and emphasis of information provided by innocent callers would differ from that of perpetrators and that these differences would result in observable indicators of veracity and deception.

Indeed, Harpster et al. (2009) reported numerous distinctions between the two types of callers. These differences, along with a subsequent publication (Harpster et al., 2017), led to the creation of the 911 Considering Offender Probability in Statements (COPS) Scale©. The 911 COPS Scale© contains 15 "innocent indicators" and 38 "guilty indicators." Innocent indicators include the caller issuing a plea for help, exhibiting a sense of urgency, fearing for his/her safety, providing relevant information, and focusing on the victim. In contrast, guilty indicators include the caller not pleading for help, exhibiting urgency, or fearing for his/her safety, providing extraneous information, and focusing on him/herself, among many others.

Later studies attempted to replicate the findings of Harpster et al. (2009) and the resulting 911 COPS Scale©, with little success and equivocal findings. For instance, Cromer et al. (2018) examined whether 18 indicators, nine of which were based upon Harpster et al. (2009), could correctly identify fifty emergency calls of homicides and suicides. Only two of the nine indicators proposed by Harpster et al. (2009) correctly identified truthful and deceptive callers (extraneous information and conflicting facts), with a third variable approaching significance (possession of the problem). None of the remaining variables distinguished between veracity and deception.

A larger 2020 study examined 175 emergency calls of reported homicides and suicides and found only four of 28 indicators based on Harpster et al. (2009) discriminated between truthful and deceptive callers. Consistent with Harpster et al. (2009), deceptive callers repeated the word "just" more times throughout the call than did truthful callers (Miller et al., 2020). However, other effects were not in the expected direction. For example, contrary to Harpster et al.'s (2009) findings, voice modulation (i.e., change in intensity and pitch of voice) was more common with deceptive callers, whereas simply notifying dispatchers of a dead body was more likely among truthful callers.

O'Donnell et al. (2022) further attempted to replicate the findings of Harpster et al. (2009) using 70 calls made by caregivers of missing children. These calls included caregivers who were unaware of the circumstances of the child's disappearance, as well as those who were

responsible for the disappearance but concealed their involvement from dispatchers. The authors' findings were largely consistent with that of Miller et al. (2020). Of the forty-three 911 COPS Scale© variables tested, only six were supported. Moreover, as the authors explained, two of these may have resulted from idiosyncrasies in coding. The consistent findings between O'Donnell et al. (2022) and Miller et al. (2020) across different call types further calls into question the efficacy of Harpster et al. (2009) and the 911 COPS Scale©.

In another study, Markey et al. (2022) examined 86 deception cues across 146 calls. While not a direct comparison to the 911 COPS Scale©, Markey et al. (2022) found that deceptive callers tended to exaggerate their emotions, acted in a reckless manner, obstructed the victim from receiving help, became self-defensive, and provided evasive responses, among others. Truthful callers, on the other hand, were forthright, focused on the event, were helpful, corrected any errors, and relayed a plausible message. Finally, Markey et al. (2023) found that truthful emergency callers were more helpful than deceptive callers and behaved less emotionally. Further, the emotional behaviors of deceptive callers increased throughout the call, in contrast to helpful behaviors, which decreased as the call progressed.

Thus, identification of deception cues that can reliably distinguish veracity and deception in emergency calls has been elusive. The inconsistencies in the literature could be due to several reasons, including variations in the methodologies of available studies, the dynamic and fragmented nature of emergency calls, the lack of substantial studies investigating the ways in which dispatcher behavior may influence caller responses, and the absence of studies using criteria from empirically supported methods such as criteria-based content analysis (CBCA; Steller & Köhnken, 1989).

### Grice's Maxims and Deception Detection

In his seminal essay "Logic and conversation," Grice (1975) outlined four maxims of conversation essential for communication: quantity, quality, relation, and manner. According to Grice, the maxim of quantity holds that a person's communication should not be more informative than required, whereas quality specifies that statements should not include information a person knows to be false, or which lack adequate evidence. The maxim of relation means simply to be relevant when communicating, while manner refers to how information is communicated (e.g., information should be orderly, brief, and unambiguous). The four maxims support the cooperative principle, which describes how people utilize mutually beneficial conversation to communicate effectively. Violations of this principle may result in various communication problems, such as one-party misinterpreting what the other is stating or one party misleading the other.

Although Grice's (1975) work was not focused on deception detection, his four maxims highlight the important role verbal content plays in human interactions and interpretations of verbal information. Verbal cues to deception have been studied for decades and are generally considered to be better indicators of veracity and deception than behavioral cues (Hauch et al., 2016; Strömwall et al., 2006). Physiological and behavioral manifestations may suggest nervousness, excitement, anger, or frustration in individuals, but these can be exhibited by truth-tellers as well as liars (Vrij & Fisher, 2020). However, verbal indicators of deception are not infallible either, as valid cues – verbal or otherwise – are notoriously weak (Bond & DePaulo, 2006; DePaulo et al., 2003).

Literature on deception conflicts with people's beliefs regarding deception detection, as individuals tend to be overconfident in their skills to discriminate between truth and lies (Aamodt & Custer, 2006). In actuality, humans tend to perform no better than slightly above chance, with mean truth-lie discrimination being

approximately 54% (Bond & DePaulo, 2006; Bond & DePaulo, 2008). This poor overall performance may be due in part to misguided reliance on stereotypical beliefs or cues about lying (Hartwig et al., 2010). Although these beliefs often lack empirical support, they are nonetheless widespread and resistant to change (Strömwall et al., 2004).

Erroneous beliefs concerning verbal and behavioral deception cues extend beyond laypersons to professionals involved in deception detection, such as police officers (Akehurst et al., 1996). This may explain why law enforcement professionals generally perform no better than laypersons at detecting deception (Garrido et al., 2004). Further, although methods of detecting deception, such as CBCA and Reality Monitoring (RM), enjoy empirical support from researchers and typically increase performance to above chance levels, these methods are not widely known by laypersons or utilized among law enforcement (Amado et al., 2015; Amado et al., 2016; Gancedo et al., 2021).

Differences between laypersons and law enforcement professionals do exist, however. Laypersons tend to perform better at identifying truthful statements than at detecting false statements due to a truth bias (Vrij, 2008). As noted by Levine (2014), people are generally more exposed to truthful than deceptive statements in daily life, resulting in reliance on heuristic modes of thinking when evaluating the truthfulness of a statement. By contrast, law enforcement professionals often demonstrate a lie bias and may assume a suspect is guilty (Kassin, 2005; Meissner & Kassin, 2002). This can produce a focused attention on cues that confirm these assumptions.

Still, some studies have shown that law enforcement professionals can discriminate between truthful and deceptive statements at levels above chance. In a study of police officers, Mann et al. (2004) found truth and lie accuracy to be 65%. A subsequent study involving police officers found overall accuracy to be 72% (Vrij et al., 2006). These are among the highest rates ever recorded in the literature, indicating that under certain conditions, law enforcement professionals may perform better than other research suggests. Multiple explanations for this high performance exist. As Vrij et al. (2006) described, officers with experience in interviewing may be familiar with the types of lies individuals tell during police interviews. The lies told in these studies were also higher stake lies than are typically told in laboratory settings.

Indeed, some researchers have claimed that lies told in laboratory settings are not generalizable to real-life settings (Buckley, 2012) and that law enforcement professionals perform better at high-stakes lies than low-stake lies (O'Sullivan et al., 2012). However, a meta-analysis by Hartwig and Bond, (2014) found lie detectability to be stable across contexts. This analysis included situations concerning traumatic experiences or negative life events, which were expected to involve strong emotion. Further, Vrij et al. (2006) noted several methodological limitations of Mann et al. (2004). First, the police officers did not actually conduct the interviews, but instead watched recordings of interviews conducted by others. Previous research found that passive observers may perform better at detecting truth and lies than interviewers (Burgoon et al., 2001). Second, the officers were only exposed to a small portion of the overall interview, as ground truth could not be established for much of the remaining portions of the interviews. Third, officers did not know any of the specific case facts in the experiment, which may differ from real-life police interviews.

### The Present Study

The current study sought to examine whether Grice's (1975) maxims of quantity, manner, and relation of information are effective for distinguishing between emergency callers who truthfully report not knowing the whereabouts of the child (TRC) and callers who

falsely report not knowing the child's whereabouts, but who either have direct knowledge of or was involved in the disappearance of the child (FRC). We also examined accuracy rates for distinguishing between truthful and deceptive callers and compared accuracy rates between experienced sworn law enforcement officers and non-sworn staff members.

We hypothesized the following:

1. First, we hypothesized that regardless of participant status as sworn law enforcement or non-sworn staff, overall accuracy rates for correctly classifying callers as TRC or FAC would perform similarly to previous research and would not differ from each other.
2. Second, we hypothesized that, regardless of participant status as sworn law enforcement or non-sworn staff, callers judged to have provided an appropriate quantity of information offered relevant as opposed to irrelevant information or described information in a clear and unambiguous manner would be more likely classified by participants as TRC than FAC. Similarly, we hypothesized that TRCs would contain more instances of appropriate information, relevant information, and clear and unambiguous information than FACs.
3. Third, we hypothesized that participants' narratives describing their reasons for judging a call to be truthful or deceptive would not be significantly associated with correct identification of TRCs and FACs. We also hypothesized that sworn law enforcement officers and non-sworn staff would be similar in the themes and caller characteristics used to judge a call as truthful or deceptive.
4. Fourth, we hypothesized that non-sworn staff members would demonstrate a truth bias when evaluating calls, while sworn law enforcement officers would demonstrate a lie bias.

## Method

### Participants

A total of 40 participants from 19 states in the United States were recruited from a federal law enforcement agency to assess the emergency calls. Because we were interested in examining how law enforcement officers performed in identifying deception in the emergency calls, two groups of participants were recruited. Twenty participants were sworn law enforcement officers with experience and training in conducting investigations of criminal activity. The remaining twenty were non-sworn professional staff members of the law enforcement agency. Non-sworn staff did not have any experience as a sworn law enforcement officer and served in a variety of support positions for the law enforcement agency.

All participants completed the informed consent process in accordance with human subject protection regulations. All procedures were reviewed and approved by an institutional review board.

### Emergency Calls

Emergency calls from the United States reporting the disappearance of a child were obtained from the Internet via news agencies (e.g., local affiliates of NBC, CBS, and ABC news organizations), along with other online media sources (e.g., YouTube, police department releases). In instances in which the full, unredacted call was not publicly available, the authors requested and received the full version of the call from the investigating law enforcement agency. Emergency calls pertaining to three case outcome types were included as stimuli: 1) non-criminal event (e.g., runaway, wandering child),

**Table 1.** Descriptions and Intraclass Correlation Values of Themes/Call Characteristics in Participant Narratives

Theme or Caller Characteristic	Description	ICC
Plausible Story	The participant describes the narrative provided by the caller as being plausible, consistent/not changing during the call, "making sense," and/or clear.	.90
Implausible Story	The participant describes the caller as providing a narrative that is implausible, does not make sense, or seems "far-fetched." Also included are instances in which the participant described the narrative as disjointed, confusing, odd, unusual, or bizarre.	.87
Irrelevant Information	The participant describes the caller as providing additional information that was not relevant or not related to the incident or child being missing.	.86
Excessive Information or Detail	The participant describes the caller as providing excessive detail or amount of information. However, the excessive information is <i>not</i> described as irrelevant, unnecessary, or off topic. For example, a caller may provide excessive detail about a vehicle seen in the area of the child's disappearance.	.88
Insufficient or Missing Information	The participant states that the caller did not provide sufficient information or failed to provide key information. It should be noted that the reason for not providing the information was not considered. For example, the caller could genuinely have not known the information, or the participant could have felt that the caller was deliberately withholding information.	.89
Remained on Topic	The participant described the caller as remaining focused on the child in the information provided to the dispatcher, without providing of-topic, irrelevant, or unneeded information.	.93
Contradiction or Lack of Expected Information	The participant states that the caller contradicts themselves in the information they provide or within the narrative given to the dispatcher, or the caller does not know information that the participant felt should be known (e.g., the caller does not know the child's age or spelling of the child's name).	.85
Sense of Urgency or Concern	The participant stated that the caller showed concern or urgency regarding the child being missing.	.84
Lack of Urgency or Concern	The participant stated that the caller did not sound or convey a sense of concern or urgency over the child being missing. This includes statements from the participant about the caller waiting to or delaying reporting of the missing child to authorities.	.91
High Emotion	References to the caller demonstrating heightened emotion, being upset, or being distressed.	.90
Low or No Emotion	References to the caller demonstrating low emotion, no emotion, or being calm.	.92
Fake Emotion	The participant describes the caller as "faking" or not being genuine in their emotion, "putting on a show," or "displaying crocodile tears" in regards to the caller's emotional reaction or crying.	.88
Searched for Child	The participant mentions that the caller physically searched for the child or took steps to locate the child (e.g., calling friends/family/school).	.95
Linguistic Abnormality	The participant describes the specific wording or phrasing used by the caller as being unusual or suspicious (e.g., the caller says "the child" instead of "my child").	.95

2) non-caregiver or stranger abduction, and 3) false allegation, in which the caller was a caregiver who caused the death of the child, believed the child was dead, or had knowledge of the child's death prior to making the emergency call to report the child's disappearance. For the non-criminal events and non-caregiver or stranger abductions, the caller truthfully reported to the dispatcher that they were unaware of the location of their child (true report call, or TRC). False allegation calls (FAC) were marked by the caller being deceptive in reporting the child as missing but either having direct knowledge of or being involved in the disappearance of the child. A total of 32 emergency calls were included as stimuli, with 16 of the calls being TRC and the other 16 being FAC.

## Procedure

Participants were informed that the current study intended to determine if proposed indicators of veracity could be used to distinguish between truthful emergency calls reporting a missing child and false allegations in which the caller was responsible for or had knowledge of the child's death prior to making the call. Participants were instructed to listen to each call once and then answer four questions related to Grice's (1975) maxims. Specifically, participants were asked to select whether the information provided by each caller was insufficient, appropriate, or excessive (quantity); relevant or irrelevant (relation); clear, unambiguous, and orderly or unclear, ambiguous, and disorderly (manner); and truthful or not truthful (quality). Participants were instructed to only select one response from each category and record their selection on a scoring sheet.

Participants were then asked to provide a brief written narrative regarding their impressions of the caller, their interpretation of the information provided, and why they determined the call to be truthful or deceptive. These narratives were coded for themes and caller characteristics described by participants. Themes and caller characteristics were abstracted via narrative analysis. Researchers coded predominant and recurring expressions, words, and ideas expressed by participants in the written narratives. The themes were not predetermined, but were created after the fact based on observed reoccurring patterns. The calls were divided amongst two researchers for coding of themes and call characteristics. To test for interrater reliability, a third coder was assigned one-third of all calls. See Table 1 for a list and description of the themes and caller characteristics.

Each participant listened to the same 32 calls, for a total of 1,280 possible ratings of calls. Participants were instructed to not respond to any calls with which they had familiarity of the case. A total of six responses to calls by four participants were missing due to reported familiarity with the case. Three of the observations were for a high-profile case in which the caller states the missing child's name. The participants who reported familiarity with this case recognized the missing child's name due to publicity and/or had received a brief on the investigation. For another call, one participant stated they were familiar with the case because it occurred in their region. The other two responses excluded due to participant familiarity with the case did not provide additional details regarding level of familiarity. Two participants did not complete all of the emergency calls for a total of 28 missing ratings of calls. The total number of responses to calls by participants was 1,246. The calls were presented in a random order and participants were blind as to which calls were TRC and FAC. Participants were instructed verbally and in writing that the calls were presented in random order and to not assume an even number of TRCs and FACs.

## Data Analysis

Agreement among participants in ratings of quantity, manner, relation, and quality was assessed via Krippendorff's alpha (Hayes

& Krippendorff, 2007). To compare agreement between TRCs and FACs and between law enforcement officers and non-sworn staff, Krippendorff's alpha was bootstrapped with 1,000 iterations to generate 95% confidence intervals (Krippendorff, 2016).

Because ratings of each call were nested within participants, we used multilevel binomial logistic regression to assess associations between type of call (TRC versus FAC) and participant ratings of call quality (truthful or deceptive), relation (relevant or irrelevant), and manner (clear/organized/unambiguous or unclear/disorganized/ambiguous). Quantity of information was separated into two hierarchical binary logistic regression models (insufficient versus appropriate, excessive versus appropriate) as described by Begg and Gray (1984). Employee type (non-sworn staff versus sworn law enforcement officers) was included in the model to explore differences in judgments and accuracy in classifying calls as truthful or deceptive.

For the themes and caller characteristics coded from participant narratives, interrater reliability was calculated via intraclass correlation (ICC). Multilevel binomial logistic regression was used to examine associations between themes/characteristics and participant ratings of truthfulness and deception (Quality), and the association between type of call (TRC or FAC) and themes/caller characteristics. Differences between non-sworn staff and law enforcement officers and between TRCs and FACs in the total number of attributes or characteristics contained in narratives were examined using hierarchical ordinal regression.

To correct for multiple comparisons, we applied the Benjamini-Hochberg procedure with a 10% false discovery rate. Adjusted *p*-values are presented throughout the results.

All analyses were conducted using R (R Core Team, 2023). Krippendorff's alpha and intraclass correlation were run using the *irr* (Gamer et al., 2019) package. Bootstrapping of Krippendorff's alpha was conducted using the *kripp.boot* (Proutskova & Gruszczynski, 2020) package. Hierarchical binomial logistic regressions were run using the *lme4* package (Bates et al., 2015); Hierarchical ordinal regressions were computed using the "ordinal" package (Christensen, 2023).

## Results

### Agreement among Participants in Ratings of Verbal Characteristics

There was poor agreement amongst participants for ratings of quantity ( $\alpha = .18$ , 95% CI [.02, .33]), manner ( $\alpha = .31$ , 95% CI [.11, .49]), and relation ( $\alpha = .13$ , 95% CI [-.20, .40]) of information in the calls, as well as for judgments of whether the calls were truthful or deceptive ( $\alpha = .17$ , 95% CI [-.04, .39]). Although the agreement for quantity of information and manner were better than chance, both characteristics were still below the accepted level of .80 for interrater agreement. Agreement coefficients did not differ between TRCs and FACs, nor between law enforcement officers and non-sworn staff. See Table 2 for Krippendorff alpha and 95% CIs for all verbal characteristics.

As a result of the poor participant agreement, caution should be used when interpreting the analyses pertaining to Grice's maxims. We present the results of the analyses because the purpose of the current study was to examine the relationship between participant ratings of Grice's maxims and ratings of callers as truthful or deceptive, the relationship between type of call and ratings of Grice's maxims, and differences between sworn law enforcement personnel and professional staff.

### Accuracy of Classifying Calls as TRC and FAC

Overall, participants correctly identified 58% of calls as truthful or deceptive. TRCs (69%) were more likely to be correctly identified

than FACs, 46%; Wald's  $\chi^2(1) = 75.80, p = .020, OR = 2.82, 95\% CI [2.22, 3.58]$ , and law enforcement officers were more likely than non-sworn staff to correctly identify calls as truthful or deceptive. Wald's  $\chi^2(1) = 14.52, p = .041, OR = 1.78, 95\% CI [1.36, 2.33]$ .

Across both types of calls, non-sworn staff correctly identified 51% of calls. Non-sworn staff had 4.6 times greater odds of identifying TRCs than FACs, with TRCs and FACs being correctly identified 69% and 33% of the time respectively. Overall, law enforcement officers correctly identified 64% of calls, with TRCs having nearly two times greater odds of correct identification over FACs (70% versus 58% correctly identified respectively, Wald's  $\chi^2(1) = 16.24, p = .034, OR = 0.35, 95\% CI [0.23, 0.61]$  (see Table 3.)

**Table 2.** Krippendorff Alpha and 95% Confidence Interval Values for Ratings of Grice's (1975) Maxims

Maxim	$\alpha$	Bootstrapped Average $\alpha$	95% CI
Quantity	.17	.18	.02, .33
LEO – TRC	.22	.22	.08, .35
LEO – FAC	.11	.11	-.04, .25
NSS – TRC	.27	.27	.13, .40
NSS – FAC	.12	.12	-.08, .29
Relation	.13	.13	-.20, .40
LEO – TRC	.10	.09	-.24, .40
LEO – FAC	.17	.19	-.15, .39
NSS – TRC	.09	.10	-.23, .40
NSS – FAC	.19	.19	-.14, .49
Manner	.31	.31	.11, .49
LEO – TRC	.28	.29	.10, .46
LEO – FAC	.29	.28	.09, .47
NSS – TRC	.29	.29	.09, .49
NSS – FAC	.36	.37	.14, .56
Quality	.17	.16	-.04, .39
LEO – TRC	.14	.14	-.10, .38
LEO – FAC	.09	.09	-.13, .28
NSS – TRC	.25	.25	.03, .44
NSS – FAC	.23	.23	.01, .44

Note. LEO = law enforcement officer; NSS = non-sworn staff; TRC = true report call; FAC = false allegation call.

**Bias in Ratings of Truthfulness or Deception**

On average, participants classified 62% of calls as truthful, with a standard deviation of 15%. 55% and 68% of calls were classified as truthful by law enforcement officers and non-sworn staff respectively. Across participants, the range of calls classified as truthful was between 34% to 91% (34-84% for law enforcement officers, 44-91% for non-sworn staff).

Overall, calls were more likely to be judged as truthful than deceptive ( $z = 5.54, p = .032, OR = 1.70, 95\% CI [1.37, 2.10]$ ). This bias did not differ between non-sworn staff and law enforcement officers, Wald's  $\chi^2(1) = 6.76, p = .055, OR = 0.58, 95\% CI [0.39, 0.86]$  (see Table 4).

**Relationships between Grice's Maxims, Ratings of Truthfulness versus Deception, and Accuracy in Identifying TRCs and FACs**

**Quantity**

Calls classified as having an appropriate quantity of information were more likely to be judged as being truthful than if the call was classified as having insufficient information, Wald's  $\chi^2(1) = 193.97, p = .011, OR = 7.33, 95\% CI [5.43, 9.91]$ , or excessive information, Wald's  $\chi^2(1) = 207.89, p = .007; OR = 26.91, 95\% CI [15.45, 46.88]$ . Specifically, 84% of calls classified as having an appropriate amount of information were judged to be truthful. In contrast, 56% of calls classified as having insufficient information and 80% of those classified as having excessive information were judged as deceptive. This pattern was observed for both employee types : appropriate versus insufficient, Wald's  $\chi^2(1) = 0.72, p = .082, OR = 1.29, 95\% CI [0.72, 2.34]$ , and appropriate versus excessive, Wald's  $\chi^2(1) = 0.15, p = .093, OR = 1.22, 95\% CI [0.44, 3.35]$ .

Regarding the relationship between quantity of information and type of call, TRCs were equally likely to be rated by participants as having insufficient or appropriate quantity of information. However, FACs were more likely to be classified as having appropriate than insufficient information, Wald's  $\chi^2(1) = 11.72, p = .043, OR = 0.65, 95\% CI [0.51, 0.83]$ . TRCs and FACs were equally likely to be classified as having appropriate and excessive quantity

**Table 3.** Percentages of Participant Judgments of Emergency Calls as Truthful or Deceptive based on Grice's Maxims and Relationship between Type of Call and Maxims

Predictor	Participant Ratings					Type of Call				
	Truthful (n = 771)	Deceptive (n = 475)	OR	95% CI LL	95% CI UL	TRC (n = 608)	FAC (n = 611)	OR	95% CI LL	95% CI UL
Quantity			7.33*	5.43	9.91			0.65*	0.51	0.83
Appropriate	84.2%	15.8%				47.3%	52.7%			
Insufficient	43.8%	56.2%				57.0%	43.0%			
Quantity			26.91*	15.45	46.88			1.79	1.19	2.69
Appropriate	84.2%	15.8%				47.3%	52.7%			
Excessive	20.2%	79.9%				35.8%	64.2%			
Relation of Information			14.19*	9.12	22.09			1.12	0.82	1.55
Relevant	70.0%	30.0%				50.1%	49.9%			
Not relevant	16.9%	83.1%				48.5%	51.5%			
Manner of Information			5.51*	4.24	7.17			0.82	0.65	1.03
Clear/unambiguous/orderly	76.1%	23.9%				47.7%	52.3%			
Unclear/ambiguous/disorderly	38.3%	61.7%				53.2%	46.8%			

Note. OR = odds ratio; CI = confidence interval; LL = lower limit; UL = upper limit. \* $p < .05$ .

**Table 4.** Themes and Caller Characteristics in Participant Narratives and Their Association with Participant Judgments of Calls as Truthful or Deceptive and with Type of Call

Predictor	Participant Ratings					Type of Call				
	Truthful (n = 744)	Deceptive (n = 467)	OR	LL	UL	TRC (n = 608)	FAC (n = 611)	OR	LL	UL
Plausibility of Story	17.3%	4.9%	4.37*	2.70	7.08	12.8%	12.1%	1.06	0.75	1.50
Law enforcement officer	12.5%	4.7%				8.5%	9.4%			
Non-sworn staff	21.3%	5.1%				17.3%	14.8%			
Implausibility of Story	12.1%	32.1%	0.25*	0.18	0.34	18.8%	20.6%	0.88	0.66	1.17
Law enforcement officer	11.9%	29.8%				17.9%	21.5%			
Non-sworn staff	12.3%	35.4%				19.6%	19.7%			
Contradictory/Lacking Info.	2.8%	17.1%	0.11*	0.07	0.19	8.2%	8.3%	0.98	0.65	1.48
Law enforcement officer	2.4%	9.2%				4.2%	6.5%			
Non-sworn staff	4.2%	28.2%				12.3%	10.2%			
Insufficient/Missing Info.	16.3%	22.5%	0.62*	0.45	0.85	23.8%	13.4%	2.13*	1.56	2.90
Law enforcement officer	13.1%	17.3%				18.6%	11.4%			
Non-sworn staff	18.9%	29.7%				29.2%	15.5%			
Irrelevant/Unnecessary Info.	2.7%	13.3%	0.15*	0.09	0.27	6.6%	6.9%	0.91	0.58	1.45
Law enforcement officer	2.4%	12.5%				6.2%	7.5%			
Non-sworn staff	2.9%	14.4%				7.0%	6.3%			
Excessive Info. or Detail	3.1%	6.9%	0.42*	0.23	0.75	3.0%	6.2%	0.46*	0.26	0.83
Law enforcement officer	3.3%	7.4%				3.6%	6.8%			
Non-sworn staff	2.9%	6.2%				2.3%	5.6%			
Remained On Topic	14.1%	2.8%	6.37*	3.40	11.94	8.9%	10.5%	0.79	0.52	1.21
Law enforcement officer	9.2%	2.2%				6.2%	5.9%			
Non-sworn staff	18.1%	3.6%				11.6%	15.1%			
Urgency/Concern	7.7%	1.5%	7.47*	3.25	17.18	6.7%	3.8%	1.96*	1.13	3.41
Law enforcement officer	4.2%	0.7%				4.6%	0.7%			
Non-sworn staff	10.5%	2.6%				9.0%	6.9%			
Lacking Urgency/Concern <sup>1</sup>	5.5%	0.4%	0.17*	0.12	0.25	15.1%	13.7%	1.13	0.82	1.55
Law enforcement officer	3.9%	0.7%				12.7%	11.1%			
Non-sworn staff	6.9%	0.0%				17.6%	16.4%			
High Emotion <sup>1</sup>	31.5%	16.9%	2.25*	1.67	3.05	22.4%	29.0%	0.69*	0.53	0.90
Law enforcement officer	22.6%	16.9%				18.6%	21.2%			
Non-sworn staff	38.7%	16.9%				26.2%	36.8%			
Low or No Emotion <sup>1</sup>	12.8%	23.6%	0.41*	0.29	0.57	16.3%	17.3%	0.93	0.68	1.26
Law enforcement officer	14.6%	21.0%				16.6%	17.9%			
Non-sworn staff	11.3%	27.2%				15.9%	16.8%			
Fake/Feigned Emotion	1.1%	7.9%	0.14*	0.06	0.30	1.0%	6.4%	0.13*	0.06	0.32
Law enforcement officer	2.1%	12.1%				1.6%	11.4%			
Non-sworn staff	0.2%	2.1%				0.3%	1.3%			
Linguistic Abnormality	3.5%	14.3%	0.22*	0.14	0.37	7.7%	7.5%	1.03	0.67	1.60
Law enforcement officer	6.3%	18.8%				10.7%	12.7%			
Non-sworn staff	1.2%	8.2%				4.7%	2.3%			
Searched for Child	8.7%	6.6%	1.66	1.03	2.68	6.3%	9.5%	0.62*	0.40	0.96
Law enforcement officer	10.4%	8.5%				7.5%	11.4%			
Non-sworn staff	7.4%	4.1%				5.0%	7.6%			

Note. OR = odds ratio; CI = confidence interval; LL = lower limit; UL = upper limit.

<sup>1</sup>Indicates ratings of themes/caller characteristics differing by employee type.

\*  $p < .05$ .

of information, Wald's  $\chi^2(1) = 7.82, p = .050, OR = 1.79, 95\% CI [1.19, 2.69]$ . These patterns were similar for both non-sworn staff and law enforcement officers: appropriate versus insufficient, Wald's  $\chi^2(1) = 3.60, p = .057, OR = 1.61, 95\% CI [0.98, 2.63]$ , and appropriate versus excessive, Wald's  $\chi^2(1) = 1.26, p = .077, OR = 1.61, 95\% CI [0.71, 3.64]$ .

### Relation

Calls classified by participants as containing relevant information had over 14 times greater odds of being judged as truthful than deceptive, with 70% of calls classified as providing relevant information being labeled as truthful. Conversely, 87% of calls classified as having irrelevant information were classified as being deceptive,

Wald's  $\chi^2(1) = 193.41, p = .011, OR = 14.19, 95\% CI [9.12, 22.09]$ . This pattern did not differ by job type, Wald's  $\chi^2(1) = 3.03, p = .061, OR = 1.49, 95\% CI [0.94, 2.36]$ .

However, TRCs and FACs were equally likely to be classified as having relevant or irrelevant information, Wald's  $\chi^2(1) = 0.49, p = .089, OR = 1.12, 95\% CI [0.82, 1.55]$ , regardless of employee type, Wald's  $\chi^2(1) = 2.89, p = .070, OR = 1.76, 95\% CI [0.93, 3.36]$ .

### Manner

Calls categorized as having clear/organized/unambiguous information had 5.5 times greater odds of being classified as being truthful, with 76% of calls categorized as having clear/organized/

unambiguous information being classified as truthful. Conversely, 62% of calls labeled as having unclear/disorganized/ambiguous information were judged to be deceptive, Wald's  $\chi^2(1) = 175.17$ ,  $p = .018$ ,  $OR = 5.51$ , 95% CI [4.24, 7.17], regardless of employee type, Wald's  $\chi^2(1) = 0.08$ ,  $p = .100$ ,  $OR = 1.08$ , 95% CI [0.64, 1.82].

For both law enforcement officers and non-sworn staff, Wald's  $\chi^2(1) = 0.46$ ,  $p = .091$ ,  $OR = 1.17$ , 95% CI [0.74, 1.87], ratings of call manner were not significantly associated with whether the call was a TRC or FAC, Wald's  $\chi^2(1) = 2.97$ ,  $p = .066$ ,  $OR = 0.82$ , 95% CI [0.65, 1.03].

See Table 3 for participant judgments of emergency calls based on Grice's maxims and type of call (TRC versus FAC).

### Themes and Caller Characteristics within Participant Narratives

The average number of themes/caller characteristics described in participant narratives was  $1.72 \pm 1.18$  per call, with a range of 0 to 7 and a median of 2 themes/caller characteristics per call. Using hierarchical ordinal regression, there were no differences between non-sworn staff and law enforcement officers, Wald's  $\chi^2(1) = 2.63$ ,  $p = .105$ ,  $OR = 0.58$ , 95% CI [0.30, 1.11], or between TRCs and FACs, Wald's  $\chi^2(1) = 0.42$ ,  $p = .515$ ,  $OR = 0.93$ , 95% CI [0.76, 1.15], in the number of features or characteristics described in narratives.

### Plausibility of Story Provided by Caller

Call narratives characterized as being plausible had over four times greater odds of being classified as truthful than deceptive, Wald's  $\chi^2(1) = 44.45$ ,  $p = .008$ ,  $OR = 4.37$ , 95% CI [2.70, 7.08]; this pattern was similar for non-sworn staff and law enforcement officers, Wald's  $\chi^2(1) = 1.29$ ,  $p = .064$ ,  $OR = 0.57$ , 95% CI [0.22, 1.51]. However, TRCs and FACs were equally likely to be characterized as having a plausible narrative, Wald's  $\chi^2(1) = 0.10$ ,  $p = .091$ ,  $OR = 1.06$ , 95% CI [0.75, 1.50]. Non-sworn staff were more likely than law enforcement officers to describe a caller's story as plausible, Wald's  $\chi^2(1) = 4.95$ ,  $p = .045$ ,  $OR = 0.48$ , 95% CI [0.25, 0.90].

In contrast to narratives described as being plausible, narratives described as implausible had four times greater odds of being classified as deceptive over truthful, Wald's  $\chi^2(1) = 77.49$ ,  $p = .004$ ,  $OR = 0.25$ , 95% CI [0.18, 0.34], with the same pattern emerging in both non-sworn staff and law enforcement officers, Wald's  $\chi^2(1) = 1.85$ ,  $p = .060$ ,  $OR = 1.56$ , 95% CI [0.82, 2.96]. In actuality, there was no association with type of call and the narrative being characterized as implausible, Wald's  $\chi^2(1) = 0.78$ ,  $p = .072$ ,  $OR = 0.88$ , 95% CI [0.66, 1.17]. Non-sworn staff and law enforcement officers were equally likely to describe a caller's story as implausible, Wald's  $\chi^2(1) = 0.01$ ,  $p = .097$ ,  $OR = 1.02$ , 95% CI [0.64, 1.63].

Regardless of employee type, Wald's  $\chi^2(1) = 3.73$ ,  $p = .053$ ,  $OR = 2.95$ , 95% CI [1.01, 8.60], narratives described as being contradictory or not including expected information had over 9 times greater odds of being judged as deceptive than truthful, Wald's  $\chi^2(1) = 84.38$ ,  $p = .003$ ,  $OR = 0.11$ , 95% CI [0.07, 0.19]. However, TRCs and FACs were equally likely to be described as contradictory or not including expected information, Wald's  $\chi^2(1) = 0.01$ ,  $p = .098$ ,  $OR = 0.98$ , 95% CI [0.65, 1.48]. Non-sworn staff were more likely to describe calls as having contradictory or not including expected information than law enforcement officers, Wald's  $\chi^2(1) = 9.92$ ,  $p = .029$ ,  $OR = 0.45$ , 95% CI [0.28, 0.72].

### Quantity and Relevance of Information Provided by Caller

Calls described as having insufficient or missing information were more likely to be judged as deceptive, Wald's  $\chi^2(1) = 8.89$ ,  $p = .031$ ,  $OR = 0.62$ , 95% CI [0.45, 0.85], with no differences between

employee types, Wald's  $\chi^2(1) = 1.58$ ,  $p = .063$ ,  $OR = 1.50$ , 95% CI [0.80, 2.84]. Conversely, calls that were described as having insufficient or missing information were more likely to be present in TRCs than FACs, Wald's  $\chi^2(1) = 23.39$ ,  $p = .020$ ,  $OR = 2.13$ , 95% CI [1.56, 2.90]. Non-sworn staff and law enforcement officers were equally likely to describe caller narratives as containing insufficient information, Wald's  $\chi^2(1) = 3.11$ ,  $p = .055$ ,  $OR = 0.59$ , 95% CI [0.34, 1.04].

For both non-sworn staff and law enforcement officers, Wald's  $\chi^2(1) = 0.15$ ,  $p = .085$ ,  $OR = 0.80$ , 95% CI [0.26, 2.47], calls described as containing irrelevant or unnecessary information had over six times greater odds of being classified as deceptive than truthful, Wald's  $\chi^2(1) = 52.77$ ,  $p = .006$ ,  $OR = 0.15$ , 95% CI [0.09, 0.27]. However, there was no association between type of call and descriptions of the call having irrelevant or unnecessary information, Wald's  $\chi^2(1) = 0.14$ ,  $p = .088$ ,  $OR = 0.91$ , 95% CI [0.58, 1.45]. Non-sworn staff and law enforcement officers were equally likely to describe calls as containing irrelevant information, Wald's  $\chi^2(1) < 0.01$ ,  $p = .100$ ,  $OR = 0.99$ , 95% CI [0.51, 1.92].

Calls described by participants as having excessive information or detail had over twice the odds of being classified as deceptive than truthful, Wald's  $\chi^2(1) = 8.56$ ,  $p = .032$ ,  $OR = 0.42$ , 95% CI [0.23, 0.75]. This pattern did not differ between non-sworn staff and law enforcement officers, Wald's  $\chi^2(1) = 0.16$ ,  $p = .084$ ,  $OR = 0.79$ , 95% CI [0.24, 2.55]. FACs were more likely to be described as having excessive information or detail than TRCs, Wald's  $\chi^2(1) = 7.16$ ,  $p = .035$ ,  $OR = 0.46$ , 95% CI [0.26, 0.83]. Non-sworn staff and law enforcement officers were equally likely to describe a call as having excessive information or detail, Wald's  $\chi^2(1) = 0.85$ ,  $p = .068$ ,  $OR = 1.43$ , 95% CI [0.66, 3.11].

For both non-sworn staff and law enforcement officers, Wald's  $\chi^2(1) = 0.49$ ,  $p = .076$ ,  $OR = 0.64$ , 95% CI [0.18, 2.23], callers described by participants as remaining on topic had over six times greater odds of being classified as truthful than deceptive, Wald's  $\chi^2(1) = 45.82$ ,  $p = .007$ ,  $OR = 6.37$ , 95% CI [3.40, 11.94]. However, there was no association between descriptions of callers remaining on topic and whether the call was a TRC or FAC, Wald's  $\chi^2(1) = 1.18$ ,  $p = .066$ ,  $OR = 0.79$ , 95% CI [0.52, 1.21]. Non-sworn staff and law enforcement officers were equally likely to describe callers as remaining on topic, Wald's  $\chi^2(1) = 1.78$ ,  $p = .062$ ,  $OR = 0.50$ , 95% CI [0.19, 1.35].

### Sense of Urgency or Concern

Callers described by participants as having a sense of urgency or concern had over seven times greater odds of being rated as being truthful over deceptive, Wald's  $\chi^2(1) = 32.15$ ,  $p = .013$ ,  $OR = 7.47$ , 95% CI [3.25, 17.18], with no differences between employee types, Wald's  $\chi^2(1) = 0.10$ ,  $p = .089$ ,  $OR = 1.34$ , 95% CI [0.22, 8.24]. TRCs were more likely to be described as having a sense of urgency or concern than FACs, Wald's  $\chi^2(1) = 5.95$ ,  $p = .041$ ,  $OR = 1.96$ , 95% CI [1.13, 3.41]. Non-sworn staff and law enforcement officers were equally likely to describe callers as concerned, Wald's  $\chi^2(1) = 4.28$ ,  $p = .052$ ,  $OR = 0.33$ , 95% CI [0.12, 0.92].

Conversely, callers described by participants as lacking concern or urgency had nearly 6 times greater odds of being classified as deceptive than as truthful, Wald's  $\chi^2(1) = 92.02$ ,  $p = .003$ ,  $OR = 0.17$ , 95% CI [0.12, 0.25]. Although this pattern was similar for both non-sworn staff and sworn law enforcement officers, it was more pronounced in non-sworn staff ( $OR$  for non-sworn staff = 8.57) than in law enforcement officers,  $OR$  for law enforcement officers = 3.35, Wald's  $\chi^2(1) = 6.02$ ,  $p = .040$ ,  $OR = 2.55$ , 95% CI [1.21, 5.37]. However, there was no association between type of call and participants describing the call as having lack of urgency or concern, Wald's  $\chi^2(1) = 0.52$ ,  $p = .074$ ,  $OR = 1.13$ , 95% CI [0.82, 1.55]. Non-sworn staff and law enforcement officers were equally likely to describe callers as lacking concern or urgency, Wald's  $\chi^2(1) = 4.37$ ,  $p = .050$ ,  $OR = 0.64$ , 95% CI [0.42, 0.97].



### Emotion Expressed by Caller

Participants were more likely to classify callers with high emotion as being truthful than deceptive, Wald's  $\chi^2(1) = 29.46$ ,  $p = .015$ ,  $OR = 2.25$ , 95% CI [1.67, 3.05]. This effect was larger in non-sworn staff ( $OR = 3.76$ ) than in law enforcement officers ( $OR = 1.69$ ), Wald's  $\chi^2(1) = 6.18$ ,  $p = .037$ ,  $OR = 0.46$ , 95% CI [0.25, 0.85]. In actuality, calls described as having high emotions were more likely to be FACs than TRCs, Wald's  $\chi^2(1) = 7.39$ ,  $p = .034$ ,  $OR = 0.69$ , 95% CI [0.53, 0.90]. Non-sworn staff had nearly twice greater odds of describing a caller as having high emotion compared with law enforcement officers, Wald's  $\chi^2(1) = 10.13$ ,  $p = .028$ ,  $OR = 0.53$ , 95% CI [0.37, 0.76].

Callers described as showing little or no emotion were more likely to be judged as deceptive, Wald's  $\chi^2(1) = 28.37$ ,  $p = .016$ ,  $OR = 0.41$ , 95% CI [0.29, 0.57]. This pattern was more pronounced for non-sworn staff ( $OR$  for non-sworn staff = 0.27) than law enforcement officers,  $OR$  for law enforcement officers = 0.59, Wald's  $\chi^2(1) = 5.64$ ,  $p = .042$ ,  $OR = 2.23$ , 95% CI [1.15, 4.32]. However, there was no association between description of a caller having low or no emotion and whether the call was a TRC or FAC, Wald's  $\chi^2(1) = 0.23$ ,  $p = .081$ ,  $OR = 0.93$ , 95% CI [0.68, 1.26]. Non-sworn staff and law enforcement officers were equally likely to describe a caller as having low or no emotion, Wald's  $\chi^2(1) = 0.19$ ,  $p = .083$ ,  $OR = 1.15$ , 95% CI [0.62, 2.10].

Callers described as showing fake or feigned emotions were more likely to be rated as deceptive than truthful, Wald's  $\chi^2(1) = 31.43$ ,  $p = .014$ ,  $OR = 0.14$ , 95% CI [0.06, 0.30]. This pattern did not differ between non-sworn staff and law enforcement officers, Wald's  $\chi^2(1) = 0.06$ ,  $p = .092$ ,  $OR = 1.36$ , 95% CI [0.12, 15.60]. FACs were more likely to be described as having fake or feigned emotions than TRCs, Wald's  $\chi^2(1) = 29.17$ ,  $p = .016$ ,  $OR = 0.13$ , 95% CI [0.06, 0.32]. Law enforcement officers had over eight times greater odds of describing a caller as exhibiting fake or feigned emotion compared with non-sworn staff, Wald's  $\chi^2(1) = 20.21$ ,  $p = .022$ ,  $OR = 8.61$ , 95% CI [3.14, 23.59].

### Other Themes within Participant Narratives

Calls described by participants as having a linguistic abnormality had over four greater odds of being classified as deceptive than truthful, Wald's  $\chi^2(1) = 39.20$ ,  $p = .009$ ,  $OR = 0.22$ , 95% CI [0.14, 0.37], regardless of employee type, Wald's  $\chi^2(1) = 2.57$ ,  $p = .058$ ,  $OR = 2.59$ , 95% CI [0.78, 8.68]. However, there was no association between the description of a call as having a linguistic abnormality and whether the call was a TRC or FAC, Wald's  $\chi^2(1) = 0.02$ ,  $p = .094$ ,  $OR = 1.03$ , 95% CI [0.67, 1.60]. Law enforcement officers had over four times greater odds of describing the call as having a linguistic abnormality than non-sworn staff, Wald's  $\chi^2(1) = 14.84$ ,  $p = .025$ ,  $OR = 4.18$ , 95% CI [2.02, 8.65].

Participants who mentioned the caller having searched for the child were equally likely to judge the caller as truthful than deceptive, Wald's  $\chi^2(1) = 4.49$ ,  $p = .051$ ,  $OR = 1.66$ , 95% CI [1.03, 2.68], with no differences between non-sworn staff and law enforcement officers, Wald's  $\chi^2(1) = 0.50$ ,  $p = .077$ ,  $OR = 0.70$ , 95% CI [0.25, 1.92]. FRCs were more likely to have mentions of searching for the child than TRCs, Wald's  $\chi^2(1) = 4.68$ ,  $p = .047$ ,  $OR = 0.62$ , 95% CI [0.40, 0.96]. Non-sworn staff and law enforcement officers were equally likely to mention the caller searching for the child, Wald's  $\chi^2(1) = 0.82$ ,  $p = .071$ ,  $OR = 1.42$ , 95% CI [0.67, 3.01].

See Table 4 for presence of themes/caller characteristics in participant themes and associations with judgments of calls as truthful or deceptive and with type of call (TRC versus FAC).

### Discussion

Emergency calls differ from open-ended narratives in that they generally involve dynamic and highly fragmented communications

that are often conducted under stressful or traumatic conditions, which may not be suitable for certain methods of deception detection, such as CBCA or RM (Masip et al., 2005; Vrij, 2005; Vrij et al., 2018). Due to this, we examined whether Grice's (1975) four principles of verbal communication could be utilized to help distinguish between truthful (TRC) and deceptive (FAC) reports of a missing child. We further compared accuracy rates for identifying TRC or FAC between two groups of participants, sworn law enforcement officers and non-sworn staff members of a law enforcement agency. We also examined the reasoning of both groups for judging calls as truthful or deceptive.

Our first hypothesis was partially supported. The ability of sworn and non-sworn personnel to distinguish between TRC and FAC was just over 57%, on par with prior research demonstrating people typically perform only slightly better than chance at detecting deception (Bond & DePaulo, 2006; Bond & DePaulo, 2008). Both groups performed significantly above chance levels when identifying truthful calls (69% for non-sworn staff and 70% for sworn officers), but struggled to identify deceptive calls, with non-sworn staff correctly identifying FAC only 33% of the time, well below chance levels. Sworn officers performed better than non-sworn by correctly identifying 58% of FAC, but these results remain just above chance levels. Overall identification of TRC and FAC for each group demonstrated that sworn law enforcement officers outperformed non-sworn staff with total accuracy rates of 64% versus 50%. This overall rate for sworn officers is higher than is typically found in deception research, but consistent with a minority of studies demonstrating similar accuracy rates among police officers (Hartwig & Bond, 2014; Mann et al., 2004; Vrij et al., 2006).

However, the current study differed methodologically from Mann et al. (2004) and Vrij et al. (2006). In the current study, both sworn officers and non-sworn staff listened to the entirety of the call and were then asked to judge whether the call was TRC or FAC. This meant that participants were likely exposed to various truthful and deceptive statements within certain calls, even though ground truth for every single statement could not be established. This contrasts with Mann et al. (2004) and Vrij et al. (2006), where participants were asked to determine whether individual statements (in which ground truth had been ascertained) were truthful or deceptive. Thus, the actual ability of sworn and non-sworn personnel in the current study to differentiate between specific truthful and deceptive statements within each call is not known, as this was not tested.

Rather, the current study attempted to mirror what often occurs in real-life investigative settings. When forming initial opinions or making early investigative decisions, law enforcement may need to quickly evaluate and interpret the totality of information at-hand based on prior training and experience. This could explain why sworn personnel performed better at identifying TRC and FAC overall than non-sworn staff, even though they reported similar erroneous cues as non-sworn staff. First, sworn officers may have had more experience than non-sworn staff with the types of high stakes lies told by emergency callers who caused the death or disappearance of a child. It is unknown whether sworn personnel would have performed as well with lower stake lies. Second, the investigative experience of sworn officers may have led to a greater comprehension of the dynamics and circumstances of missing child cases than non-sworn staff. Sworn personnel were not asked to describe prior investigative experience in missing child cases, so although their previous exposure to these cases was unknown, their overall investigative experience may have given them an advantage over non-sworn staff, who did not possess similar backgrounds. It is not known whether sworn personnel would have performed as well in situations where they lacked experience. Likewise, it is unknown whether sworn personnel would have performed better than non-sworn staff in situations where both groups had similar experience. Third, because participants relied solely on audio recordings of the emergency calls, they did not have the opportunity to view caller behaviors. Previous research has

demonstrated that people generally perform poorly when relying on behavioral cues to detect deception (Hartwig et al., 2010). However, this would not fully explain the discrepancy in performance between sworn personnel and non-sworn staff, as both groups were limited to the audio recordings. It is likely that the fact that sworn officers had more previous exposure to emergency calls in general than non-sworn staff and this greater familiarity, combined with a truth bias among non-sworn staff, could have led to higher rates among sworn personnel (Vrij, 2008). Finally, law enforcement officers and non-sworn staff differed on the extent to which emotional cues impacted veracity judgements. This difference may also have contributed to the discrepancy in overall performance between the two groups.

Our hypothesis regarding the relationship between participant ratings of Grice's (1975) maxims and classifying calls as TRCs or FACs was also only partially supported. Participants who judged callers to have provided an appropriate quantity of information were more likely than those who judged the call to contain insufficient or excessive details to interpret the caller as being truthful rather than deceptive. Likewise, callers who were judged to have provided relevant information versus irrelevant information, and those who reported information in a clear and unambiguous manner, had a greater likelihood of being classified as TRC versus FAC. This was true for both sworn law enforcement and non-sworn staff. However, only insufficient quantity of information was predictive of FACs. Participant ratings of manner or relation of information were not predictive of type of call, nor was quantity of information predictive of TRCs. Thus, although participant's ratings of a call's quantity, manner, or relation of information predicted their perception of truthfulness or deception, Grice's (1975) maxims did not aid in correct identification of truthful and deceptive calls.

Additionally, agreement among participants was poor across all four maxims and was at chance levels for two of the four, specifically relation and quality (truth or deception). Participants' status as sworn or non-sworn had no impact on agreement among participants, as neither group reached acceptable levels. As a result, the outcome of the analyses exploring associations between participant ratings and type of call should be interpreted with caution. This indicates that an individual's perception of a communicator's conformity to or divergence from these general principles is likely partially unique to the individual. Factors such as personality, communication style, life experience, and education may influence this perception (Palena et al., 2021; Palena et al., 2022; Semrad & Scott-Parker, 2019).

Further, the context in which communication occurs may also be relevant. Deceptive persons are known to engage in impression management and information management strategies, and these may differ across environments (Hartwig et al., 2007; Volbert & Steller, 2014). For example, while a deceptive person may want to avoid appearing nervous or agitated during an interview, a deceptive emergency caller reporting a missing child may want to exhibit strong emotions while communicating to portray the appearance of a concerned caller (O'Donnell et al., 2023). Indeed, we found that participants who described callers as faking or feigning emotion were more likely to correctly identify FACs. However, some truthful callers may exhibit unexpected behaviors, such as lacking urgency when reporting information. These combined factors may lead to inconsistent judgements when relying on general communication principles to detect deception.

This is evidenced by the narrative responses provided by participants. Our third hypothesis was partially supported. Several themes reported by law enforcement officers and non-sworn staff were offered to support each group's veracity judgements, but these judgements were not necessarily accurate. For example, both groups of participants were more likely to rate a call as FAC than TRC if the call contained contradictions or unexpected information, insufficient or missing information, irrelevant,

unnecessary, or excessive information, a lack of urgency, or if the narrative was implausible. However, there was no actual association between TRC and FAC for four of these (contradictions/unexpected information, irrelevant/unnecessary information, lack of urgency, and implausibility), and more TRCs than FACs contained descriptions of insufficient/missing information. Similarly, both groups of participants were more likely than not to judge a call as TRC than FAC when the caller referenced searching for the child. In actuality, more FACs than TRCs included this information. The finding that plausibility was not effective in distinguishing TRC from FAC contrasts with previous research (Vrij et al., 2021). Two possible explanations may account for this discrepancy. First, Vrij et al. (2021) found that plausibility was correlated with details, complications, and verifiable sources. However, the extent to which this information is available to emergency callers reporting missing persons may be limited as the circumstances of these disappearances are often unknown at the time of the call (O'Donnell et al., 2023). As noted by Vrij et al. (2016), the nature of the event is relevant to the type and degree of details provided. Second, the dynamic and fragmented nature of emergency calls, which are also typically short in duration, might constrict the ability of callers to provide structured narratives or sufficient details. This could result in some caller statements being misinterpreted as implausible accounts.

Themes of emotion were also prevalent within participant narratives, but several differences emerged between law enforcement officers and non-sworn staff. Non-sworn staff were more likely than law enforcement officers to rate calls with high emotion as TRC and calls lacking emotion as FAC. However, neither of these assessments were accurate, as TRC and FAC were equally likely to contain descriptions of limited emotion and FAC were actually more likely than TRC to include instances of high emotion. By contrast, although law enforcement officers and non-sworn staff were both more likely to rate calls as FAC when instances of fake or feigned emotion were identified, law enforcement officers characterized calls as containing feigned emotion at a far greater rate than non-sworn staff. This was consistent with the actual calls, as FAC were more likely than TRC to include descriptions of feigned emotion.

Our final hypothesis was not fully supported. Both sworn law enforcement officers and non-sworn staff were more likely to view a caller as truthful than as deceptive. This contrasts with prior research findings showing laypersons often display a truth bias, whereas law enforcement may exhibit a bias towards deception (Kassin, 2005; Vrij, 2008). It is possible the unique context of emergency calls reporting a missing child could have accounted for this discrepancy. Information and impression management strategies utilized during emergency calls may have influenced sworn law enforcement differently than if the same strategies or behaviors were exhibited during interviews (O'Donnell et al., 2022). Alternatively, some participants may have assumed an equal number of TRC and FRC in the sample of calls, which might have artificially inflated the number of calls identified as TRC.

Regardless of these factors, the results of the current study reinforce the difficulty encountered when attempting to discriminate between veracity and deception and the risks of relying on stereotypical cues such as emotional responses when judging veracity. While sworn personnel performed at levels higher than typically seen in deception research, their ability to correctly identify TRC and FAC was still well below perfection. This has important implications. Although sworn personnel correctly identified over two-thirds of TRC, they incorrectly judged almost one-third of truthful callers to be deceptive, thereby inferring these callers were responsible for the child's disappearance when they were not. Further, sworn personnel correctly identified only slightly over half of deceptive calls. While emergency calls comprise

only a small piece of a larger investigative puzzle involving many other sources of evidence, these calls can provide valuable insight as they typically comprise the initial details available to investigators. This information can help inform first responders, generate future leads, and be compared to future interviews, as well as to other physical and forensic evidence. However, the potential to misinterpret a caller as truthful or deceptive in the early stages of an investigation can negatively influence other investigative decisions, resource planning, and questioning tactics. Distinguishing valid cues to veracity and deception from erroneous indicators can help prevent misinterpretation of information contained in emergency calls. Future research should continue to examine indicators of veracity and deception in emergency calls, as their fragmented and dynamic nature requires caution when evaluating their content.

### Limitations

Most selected calls were identified through law enforcement or open media sources, representing only cases that rise to the level of intense police response and/or wider media attention, which may not generalize to other missing child cases. Similarly, the results of this study may not generalize to other types of emergency calls, which can vary extensively in circumstance. The type of information provided in missing child emergency calls may be dissimilar to many other types of calls.

The current study did not examine how differences between emergency call operators might have affected the verbal communication of callers or overall judgements by participants. It is possible that variations among operator questioning styles or demeanors might have influenced how callers communicated and/or how participants interpreted this information.

Although non-sworn staff did not possess law enforcement backgrounds and were not responsible for investigating crimes, their employment at a law enforcement agency may have exposed them to at least some similar information as sworn personnel, along with exposure to law enforcement investigative techniques. Thus, their performance in the current study may not be indicative of other laypersons with no affiliation with law enforcement.

Finally, methodological limitations prevented the authors from examining the ability of participants to detect individual lies within each call as ground truth could not be established for every individual statement within a call. Whether this ability would have impacted overall judgements of caller involvement in the child's disappearance is unknown. It is possible that callers with no involvement in the child's disappearance may still have made deceptive statements during the call. Likewise, callers who were responsible for the child's disappearance may have included some truthful details throughout the call. Thus, it is unknown whether sworn law enforcement and non-sworn staff would have performed similarly had they been asked to identify specific truthful or deceptive statements.

### Conclusion

The present study examined the ability of sworn law enforcement personnel and non-sworn staff of a law enforcement agency to correctly identify truthful and deceptive emergency calls reporting a missing child using Grice's (1975) four maxims of communication. Results demonstrated sworn personnel outperformed non-sworn staff, with both groups displaying inconsistent rationale supporting their respective judgements and relying on deception detecting cues that are not predictive of accuracy. This has implications for investigative settings, as law enforcement must evaluate and interpret information contained in emergency calls when

determining how to respond. Initial assessments of these calls may influence future investigative decisions, interactions, or prioritization of resources. Understanding the rationale individuals utilize to form judgements of veracity and deception in emergency calls may enhance these efforts.

### Conflict of Interest

The authors of this article declare no conflict of interest.

### Acknowledgments

The authors would like to thank the law enforcement agencies that participated in this study and the FBI's Critical Incident Response Group, Investigative and Operations Support Section for supporting this project. The authors also would like to acknowledge the efforts and contributions of FBI Crime Analyst Joy Shelton and Supervisory Special Agent Tom Seftick.

### Data Availability

Research data are not shared due to legal constraints.

### Informed Consent

Informed consent was obtained from all subjects involved in the study.

### Institutional Review Board Statement

This project was approved by the FBI Institutional Review Board (#672-22).

### References

- Aamodt, M., & Custer, H. (2006). Who can best catch a liar? A meta-analysis of individual differences in detecting deception. *The Forensic Examiner*, 15(1), 6-11.
- 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\(199612\)10:6<461::AID-ACP413>3.0.CO;2-2](https://doi.org/10.1002/(SICI)1099-0720(199612)10:6<461::AID-ACP413>3.0.CO;2-2)
- Amado, B. G., Arce, R., & Fariña, F. (2015). Undeutsch hypothesis and Criteria Based Content Analysis: A meta-analytic review. *European Journal of Psychology Applied to Legal Context*, 7(1), 3-12. <https://doi.org/10.1016/j.ejpal.2014.11.002>
- Amado, B. G., Arce, R., Farina, F., & Vilarino, M. (2016). Criteria-based content analysis (CBCA) reality criteria in adults: A meta-analytic review. *International Journal of Clinical and Health Psychology*, 16(2), 201-210. <https://doi.org/10.1016/j.ijchp.2016.01.002>
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1-48. <https://doi.org/10.18637/jss.v067.i01>
- Begg, C. B., & Gray, R. (1984). Calculation of polychotomous logistic regression parameters using individualized regressions. *Biometrika*, 71(1), 11-18. <https://doi.org/10.2307/2336391>
- Beyer, K. R., & Beasley, J. O. (2003). Nonfamily child abductors who murder their victims: Offender demographics from interviews with incarcerated offenders. *Journal of Interpersonal Violence*, 18(10), 1167-1188. <https://doi.org/10.1177/0886260503255556>
- 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\\_2](https://doi.org/10.1207/s15327957pspr1003_2)
- Bond, C. F., & DePaulo, B. M. (2008). Individual differences in judging deception: Accuracy and bias. *Psychological Bulletin*, 134(4), 477-492. <https://doi.org/10.1037/0033-2909.134.4.477>
- Boudreaux, M. C., Lord, W. D., & Etter, S. E. (2000). Child abduction: An overview of current and historical perspectives. *Child Maltreatment*, 5(1), 63-71. <https://doi.org/10.1177/107755950005001008>
- Brown, K., Keppel, R., Weis, J., & Skeen, M. (2006). *Investigative case management for missing child investigations: Report II*. Attorney General of Washington and Office of Juvenile Justice and Delinquency Prevention, U.S. Department of Justice.

- Buckley, J. P. (2012). Detection of deception researchers need to collaborate with experienced practitioners. *Journal of Applied Research in Memory and Cognition*, 1(2), 126-127. <https://doi.org/10.1016/j.jarmac.2012.04.002>
- Burgoon, J. K., Buller, D. B., & Floyd, K. (2001). Does participation affect deception success? A test of the interactivity principle. *Human Communication Research*, 27(4), 503-534. <https://doi.org/10.1093/hcr/27.4.503>
- Canning, K. E., Hiltz, M. A., & Muirhead, Y. E. (2011). False allegation of child abduction. *Journal of Forensic Sciences*, 56(3), 794-802. <https://doi.org/10.1111/j.1556-4029.2011.01715.x>
- Christensen, R. (2023). *Ordinal – regression models for ordinal data*. R package version 2023-12.4. <https://cran.r-project.org/package=ordinal>
- Cromer, J. D., Brewster, J., Fogler, K., & Stoloff, M. (2018). 911 calls in homicide cases: What does the verbal behavior of the caller reveal? *Journal of Police and Criminal Psychology*, 34(2), 156-164. <https://doi.org/10.1007/s11896-018-9282-0>
- 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>
- Finkelhor, D., Hotaling, G., & Sedlak, A. (1991). Children abducted by family members: A national household survey of incidence and episode characteristics. *Journal of Marriage and the Family*, 53(3), 805-817. <https://doi.org/10.2307/352753>
- Gamer, M., Lemon, J., Fellows, I., & Singh, P. (2019). *irr: Various coefficients of interrater reliability and agreement*. R package version 0.84.1. <https://CRAN.R-project.org/package=irr>
- Gancedo, Y., Fariña, F., Seijo, D., Vilariño, M., & Arce, R. (2021). Reality monitoring: A meta-analytical review for forensic practice. *European Journal of Psychology Applied to Legal Context*, 13(2), 99-110. <https://doi.org/10.5093/ejpalc2021a10>
- 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/10.1080/0020759034400041>
- Grasso, K. L., Sedlak, A., Chiancone, J. L., Gragg, F., Schultz, D., & Ryan, J. F. (2001). *The criminal justice system's response to parental abduction*. Juvenile Justice Bulletin, Office of Juvenile Justice and Delinquency Prevention, U.S. Department of Justice
- Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Speech acts* (pp. 41-58). Brill Publishers.
- Harpster, T., & Adams, S. H. (2017). *Analyzing 911 homicide calls*. CRC Press. <https://doi.org/10.1201/9781315386508>
- Harpster, T., Adams, S. H., & Jarvis, J. P. (2009). Analyzing 911 homicide calls for indicators of guilt or innocence: An exploratory analysis. *Homicide Studies*, 13(1), 69-93. <https://doi.org/10.1177/1088767908328073>
- 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., & Strömwall, L. A. (2007). Guilty and innocent suspects' strategies during interrogations. *Psychology, Crime & Law*, 13(2), 213-227. <https://doi.org/10.1080/10683160600750264>
- Hartwig, M., Granhag, P. A., Strömwall, L. A., & Doering, N. (2010). Impression and information management: On the strategic self-regulation of innocent and guilty suspects. *The Open Criminology Journal*, 3, 10-16. <https://doi.org/10.2174/1874917801003010010>
- Hauch, V., Sporer, S. L., Michael, S. W., & Meissner, C. A. (2016). Does training improve the detection of deception? A meta-analysis. *Communication Research*, 43(3) 283-343. <https://doi.org/10.1177/0093650214534974>
- Hayes, A. F., & Krippendorff, K. (2007). Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures*, 1(1), 77-89. <https://doi.org/10.1080/19312450709336664>
- Hiltz, M. A., Donaldson, W. H., MacKizer, M., Slater, K. E., & Sloan, W. (2015). *Understanding child abduction*. In U.S. Department of Justice, *Crimes against children: Behavioral and investigative perspectives from the FBI's Behavioral Analysis Unit* (pp. 3-16). Behavioral Analysis Unit III, Critical Incident Response Group, Federal Bureau of Investigation.
- Johnston, J. R., & Girdner, L. K. (2001). Family abductors: Descriptive profiles and preventative interventions. *Juvenile Justice Bulletin*, 43. Office of Juvenile Justice and Delinquency Prevention, U.S. Department of Justice.
- Kassin, S. M. (2005). On the psychology of confessions: Does innocence put innocents at risk? *American Psychologist*, 60(3), 215-28. <https://doi.org/10.1037/0003-066X.60.3.215>
- Krippendorff, K. (2016). *Bootstrapping distributions for Krippendorff's alpha for coding predefined units: Single-valued alpha and multi-valued alpha*. University of Pennsylvania Annenberg School for Communication. <https://www.asc.upenn.edu/sites/default/files/documents/boot.c-alpha>
- Levine, T. (2014). Truth-Default Theory (TDT): A theory of human deception and deception detection. *Journal of Language and Social Psychology*, 33(4), 378-392. <https://doi.org/10.1177/0261927X14535916>
- Mann, S., Vrij, A., & Bull, R. (2004). Detecting true lies: Police officers' ability to detect suspects' lies. *Journal of Applied Psychology*, 89(1), 137-49. <https://doi.org/10.1037/0021-9010.89.1.137>
- Markey, P. M., Feeney, E., Berry, B., Hopkins, L., & Creedon, I. (2022). Deception cues during high-risk situations: 911 homicide calls. *Psychological Science*, 33(7), 1040-1047. <https://doi.org/10.1177/0956797622107721>
- Markey, P., Martin, A., Berry, B., Feeney, E., & Slotter, E. (2023). The continuous expression of emotional and helpful behavior during high-stake deception: 911 homicide calls. *Journal of Police and Criminal Psychology*, 38, 519-527. <https://doi.org/10.1007/s11896-022-09567-x>
- Masip, J., Sporer, S. L., Garrido, E., & Herrero, C. (2005). The detection of deception with the reality monitoring approach: A review of the empirical evidence. *Psychology, Crime & Law*, 11(1), 99-122. <https://doi.org/10.1080/10683160410001726356>
- 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>
- Miller, M. L., Merola, M. A., Opanashuk, L., Robins, C. J., Chancellor, An. S., & Craun, S. W. (2020). 911 what's your emergency?: Deception in 911 homicide and suicide staged as homicide calls. *Homicide Studies*, 25(2), 189-189. <https://doi.org/10.1177/1088767920948242>
- O'Donnell, D. E., Shelton, J. L., Huffman, M. C., Porter, K., & Miller, M. (2023). 911 calls in mysterious disappearances of children: Indicators of veracity and deception. *Applied Cognitive Psychology*, 37(3), 578-589. <https://doi.org/10.1002/acp.4063>
- O'Donnell, D. E., Shelton, J. L., Shaffer, S. A., Isom, A., Bowlin, J., & Wood, E. (2022). "My child is missing": 911 calls in mysterious disappearances of children. *Aggression and Violent Behavior*, 67, Article 101795. <https://doi.org/10.1016/j.avb.2022.101795>
- 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>
- Palena, N., Caso, L., Cavagnis, L., & Greco, A. (2021). Profiling the interrogee: Applying the person-centered approach in investigative interviewing research. *Frontiers in Psychology*, 12, Article 722893. <https://doi.org/10.3389/fpsyg.2021.722893>
- Palena, N., Caso, L., Cavagnis, L., Greco, A., & Vrij, A. (2022). Exploring the relationship between personality, morality and lying: A study based on the person-centered approach. *Current Psychology*, 42, 20502-20514. <https://doi.org/10.1007/s12144-022-03132-9>
- Proutskova, P., & Gruszczynski, M. (2020). *kripp.boot: Bootstrap Krippendorff's alpha intercoder reliability statistic*. R package version 1.0.0. <https://github.com/MikeGruz/kripp.boot>
- R Core Team. (2023). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*. Vienna, Austria. <https://www.R-project.org>
- Sedlak, A. J., Finkelhor, D., & Brick, J. M., (2017). *National estimates of missing children: updated findings from a survey of parents and other primary caregivers*. Office of Juvenile Justice and Delinquency Prevention, U.S. Department of Justice.
- Sedlak, A. J., Finkelhor, D., Hammer, H., & Shultz, D. J. (2002). *National Incidence Studies of Missing, Abducted, Runaway, and Thrownaway Children (NIS/MART) national estimates of missing children: An overview*. Office of Juvenile Justice and Delinquency Prevention, U.S. Department of Justice.
- Semrad, M., & Scott-Parker, B. (2019). Police, personality and the ability to deceive. *International Journal of Police Science & Management*, 22(1), 50-61. <https://doi.org/10.1177/1461355719880568>
- Steller, M., & Köhnen, G. (1989). Criteria-Based Content Analysis. In D. C. Raskin (Ed.), *Psychological methods in criminal investigation and evidence* (pp. 217-245). Springer-Verlag.
- Strömwall, L. A., Granhag, P. A., & Hartwig, M. (2004). Practitioners' beliefs about deception. In P. A. Granhag & L. A. Strömwall (Eds.), *The detection of deception in forensic contexts* (pp. 229-250). Cambridge University Press
- Strömwall, L. A., Hartwig, M., & Granhag, P. A. (2006). To act truthfully: Non-verbal behaviour and strategies during a police interrogation. *Psychology, Crime & Law*, 12(2), 207-219. <https://doi.org/10.1080/10683160512331331328>
- Volbert, R., & Steller, M. (2014). Is this testimony truthful, fabricated, or based on false memory? *European Psychologist*, 19(3), 207-220. <https://doi.org/10.1027/1016-9040/a000200>
- Vrij, A. (2005). Criteria-Based Content Analysis: A qualitative review of the first 37 studies. *Psychology, Public Policy, and Law*, 11(1), 3-41. <https://doi.org/10.1037/1076-8971.11.1.3>
- Vrij, A. (2008). *Detecting lies and deceit: Pitfalls and opportunities* (2nd edition). Wiley.
- Vrij, A., Deeb, H., Leal, S., Granhag, P. A., & Fisher, R. P. (2021). Plausibility: A verbal cue to veracity worth examining? *European Journal of Psychology Applied to Legal Context*, 13(2), 47-53. <https://doi.org/10.5093/ejpalc2021a4>
- Vrij, A., & Fisher, R. P. (2020). Unraveling the misconception about deception and nervous behavior. *Frontiers in Psychology*, 11, Article 1377. <https://doi.org/10.3389/fpsyg.2020.01377>
- Vrij, A., Leal, S., Jupe, L., & Harvey, A. (2018). Within-subjects verbal lie detection measures: A comparison between total detail and proportion of complications. *Legal and Criminological Psychology*, 23(2), 265-279. <https://doi.org/10.1111/lcrp.12126>
- Vrij, A., Mann, S., Robbins, E., & Robinson, M. (2006). Police officers ability to detect deception in high stakes situations and in repeated lie detection test. *Applied Cognitive Psychology*, 20(6), 741-755. <https://doi.org/10.1002/acp.1200>

- Vrij, A., Nahari, G., Isitt, R., & Leal, S. (2016). Using the verifiability lie detection approach in an insurance claim setting. *Journal of Investigative Psychology and Offender Profiling*, 13(3), 183-197. <https://doi.org/10.1002/jip.1458>
- Warren, J. I., Wellbeloved-Stone, J. M., Hiltz, M. A., Donaldson, W. H., Muirhead, Y. E., Craun, S. W., Burnette, A. G., & Millspaugh, S. B., (2016). An investigative analysis of 463 incidents of single-victim child abductions identified through Federal Law Enforcement. *Aggression and Violent Behavior*, 30, 59-67. <https://doi.org/10.1016/j.avb.2016.07.006>
- Warren, J. I., Reed, J., Leviton, A. C. R., Millspaugh, S. B., Dietz, P., Grabowska, A. A., Isom, A. N., Shelton, J. L. E., & Lybert, K. (2020). The lethality of non-family child abductions: Characteristics and outcomes of 565 incidents involving youth under the age of 18 years. *Behavioral Sciences & the Law*, 39(3), 262-278. <https://doi.org/10.1002/bsl.2495>