AMERICAN PSYCHOLOGICAL ASSOCIATION

Psychology, Public Policy, and Law

In the public domain
ISSN: 1076-8971

2025, Vol. 31, No. 1, 6–21
https://doi.org/10.1037/law0000440

Evaluation of the 911 Considering Offender Probability in Statements Scale as a Deception Detection Method for 911 Calls

Daniel E. O'Donnell[1], Taylor E. Burd[2], and Michelle C. Huffman[3]

[1] Behavioral Analysis Unit-3, National Center for the Analysis of Violent Crime, Federal Bureau of Investigation

[2] Oak Ridge Institute for Science and Education, Oak Ridge, Tennessee, United States

[3] Behavioral Analysis Unit-5, National Center for the Analysis of Violent Crime, Federal Bureau of Investigation

911 calls are often the catalyst for initiating a medical or law enforcement response to critically injured or deceased persons and may contain important details to assist first responders in quickly prioritizing resources. Callers providing deceptive information may impede these efforts. The current study examined the efficacy of the Considering Offender Probability in Statements (COPS) Scale in judging whether a 911 caller was or was not involved in the death or serious physical injury of the individual whom they were reporting. Specifically, the current study examined whether sworn law enforcement officers using the COPS Scale could correctly distinguish 911 callers who did not cause the person's death from 911 callers who did cause the person's death. Agreement among participants on the presence of COPS Scale indicators in calls was generally poor. All guilty and innocent indicators were predictive of participant judgments of callers' guilt or innocence. However, most indicators were not predictive of guilty or innocent calls, and several were more common in the opposite type of call. The innocence or guilt of 911 callers was mischaracterized in approximately four out of every 10 calls and also influenced questions the law enforcement officers would ask of the caller. Our findings do not support the use of the COPS Scale as a measure of innocence or guilt among 911 callers and suggest its use could detrimentally influence investigative strategies.

Keywords: emergency calls, 911, veracity, death, deception,

Supplemental materials: https://doi.org/10.1037/law0000440.supp

Death investigations are among the most challenging types of investigations faced by law enforcement, who must piece together evidence without the ability to question the deceased while attempting to extract information from grieving family members and other parties (Carpenter et al., 2016). Two kinds of death inquiries encountered by medicolegal death investigators are homicide and suicide investigations. Among the challenges in these types of investigations is the lack of complete information that may be available to investigators when initially responding to a report of a death or serious physical injury (Craun et al., 2022; Pizarro et al., 2020). This early information, however imperfect, is often provided through 911 calls. Although limited in scope, details provided in 911 calls may assist the decision-making process of first responders and can be compared to additional information as it is developed in later stages of an investigation. However, callers who provide untruthful information may thwart first responder efforts or divert available resources away from productive investigative paths. The ability to gain insight into a caller's potential veracity or deception may therefore assist investigators when determining appropriate investigative steps. The current study examines the efficacy of the 911 Considering Offender Probability in Statements (COPS) Scale in assisting law enforcement judgments of 911 callers' guilt or innocence, as well as the effect these judgments have on subsequent interview questions. The types of questions asked by law enforcement during subsequent interviews of callers may impact the degree to which callers cooperate and the details they provide. This can affect law enforcement's understanding of the incident and influence other investigative decisions. We first describe two concepts associated with these investigations: equivocal deaths and crime

This article was published Online First February 27, 2025.

Jason Chin served as action editor.

Daniel E. O'Donnell https://orcid.org/0000-0002-6613-9588

All research procedures were approved by the Federal Bureau of Investigation (FBI)'s Institutional Review Board (Protocol 671-22) and conformed to U.S. Federal Policy for the Protection of Human Subjects. American Psychological Association ethical standards were followed throughout the study. 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. The authors have no known conflicts of interest to disclose. Research data are not shared.

Daniel E. O'Donnell served as lead for conceptualization, investigation, project administration, resources, supervision, and writing–original draft, contributed equally to writing–review and editing, and served in a supporting role for formal analysis and methodology. Taylor E. Burd contributed equally to formal analysis and validation, data curation, and served in a supporting role for investigation, methodology, project administration, resources, writing–original draft, and writing–review and editing. Michelle C. Huffman served as lead for methodology, contributed equally to formal analysis and validation, and served in a supporting role for data curation, investigation, writing–original draft, and writing–review and editing.

Correspondence concerning this article should be addressed to Daniel E. O'Donnell, Behavioral Analysis Unit-3, National Center for the Analysis of Violent Crime, Federal Bureau of Investigation, FBI Academy, 1 Range Road, Quantico, VA 22135, United States. Email: deodonnell@fbi.gov

scene staging. We then briefly summarize prior deception research, including previous studies examining 911 calls, and describe elements of the COPS Scale.

Equivocal Deaths and Crime Scene Staging

Medicolegal death investigators responding to reports of a homicide, suicide, or natural death must navigate a variety of potential obstacles when evaluating cause of death. These may include a lack of eyewitnesses, inconsistent witness statements, decayed or deficient physical evidence, unknown motivations, or refusal of witnesses to cooperate (Hough, 2019; Keel et al., 2009; Schroeder & White, 2009). Among the most challenging types of death investigations are equivocal deaths, in which uncertainty exists as to whether the death was deliberate or unintentional and whether it had been perpetrated by a third party or was self-inflicted (Lacks et al., 2008; Puranik et al., 2015; Saxena & Saini, 2017; Schneidman, 1981). This uncertainty can be created by offenders who intentionally alter elements of a crime scene to mislead investigators as to the true nature of the crime (Ferguson & Petherick, 2016; Visentin et al., 2019). This deceptive behavior is known as crime scene staging (Beauregard & Bouchard, 2010). Staging can result from physical behaviors, such as repositioning a body and destruction of property, or verbal behaviors, in which the offender changes some portion of the narrative to obscure or conceal the circumstances or sequencing of events (Ferguson, 2015; Ferguson & McKinley, 2019; Geberth, 2010; Schlesinger et al., 2014). Verbal staging can take a variety of forms, including omission of key facts, projecting responsibility for the crime onto another individual, or blaming the victim for what occurred. Deception is another type of verbal staging and may encompass actions such as the filing of a false police report or lying during an interview. The verbal aspect of staging is highly relevant to the analysis of 911 calls, as they typically contain the initial details available to law enforcement. This information can influence first responder decisions or interpretations of other evidence discovered by investigators. Offenders seeking to avoid detection may deliberately initiate contact with law enforcement and provide false or incomplete information to misdirect the subsequent investigation (Miller et al., 2021). Thus, the examination of 911 calls may provide insight to investigators when evaluating the circumstances of a reported death.

Cues to Detecting Deception in 911 Calls

Research into the ability of humans to detect deception is extensive, with decades of studies demonstrating that people generally perform near or slightly above chance and often rely on misguided, stereotypical beliefs about lying (Akehurst et al., 1996; Bond & DePaulo, 2008; Masip et al., 2005). The inability to detect deception may be explained, in part, by prior research findings that most cues to deception are fairly weak, despite people's tendency to be overconfident in their own skills (Aamodt & Custer, 2006; Bond & DePaulo, 2006; DePaulo et al., 2003). Although prior literature has suggested that reliable indicators of deception exist, these studies tend to suffer from lack of replicability or low power (Luke, 2019). Further, the greater reliance on behavioral indicators over verbal cues to detect deception can lead to erroneous interpretations of an individual's veracity (Hartwig et al., 2010; Hauch et al., 2016; Strömwall et al., 2006). Behaviors signaling nervousness, excitement, anger, or frustration may be incorrectly perceived by observers as signs of deception when such behaviors may be caused by other circumstances (Vrij et al., 2000). Further, not only are erroneous beliefs about deception cues resistant to change, but they are also widespread among both laypersons and law enforcement professionals. Although several studies have found law enforcement officers perform better than laypersons in detecting deception, both groups largely perform at or around chance levels (Garrido et al., 2004; Mann et al., 2004; Strömwall et al., 2004; Vrij et al., 2006).

Attempts to identify effective means to distinguish truthful and deceptive statements have resulted in the development of several theoretical approaches, including criteria-based content analysis (CBCA; Steller & Köhnken, 1989), reality monitoring (RM; Johnson & Raye, 1981), and scientific content analysis (SCAN; Nahari et al., 2012). Of these methods, prior research has generally found support for CBCA and RM in discriminating between truthful and fabricated narratives, but not for SCAN (Bogaard et al., 2016; Maier et al., 2018; Masip et al., 2005). However, the effectiveness of CBCA and RM is limited by the lack of deceptive criteria, as both contain largely truthful indicators (Amado et al., 2016; Sporer, 1997; Vrij et al., 2019). Additionally, the utilization of CBCA and RM was primarily intended for open-ended narratives, not other types of statements (Vrij, 2005; Vrij et al., 2000). Also, as noted by Luke (2019), the actual effectiveness of CBCA may be less than believed due to reporting and methodological limitations in the existing literature. The dynamic nature of 911 calls, which often involve hasty and fragmented interactions under stressful or traumatic conditions, does not typically allow for long, uninterrupted open narratives. Further, the relatively short duration of 911 calls may inhibit opportunities for certain CBCA or RM criteria to appear in caller narratives regardless of whether the statements are truthful or deceptive (Vrij et al., 2023). This is problematic, as the absence of truthful CBCA and RM criteria within a statement does not necessarily suggest deception (Masip, 2017; Nahari et al., 2019; Vrij & Leal, 2020).

Another problem with using CBCA or RM to evaluate 911 calls for veracity is that people in stressful or traumatic situations may present differently than they would in alternate settings, such as an interview. For example, a deceptive individual may attempt to remain calm and composed during an interview to avoid appearing nervous and arousing suspicion from the interviewer, whereas the same person may exhibit strong emotions during a 911 call to mimic the reaction of a genuinely panicked caller (Strömwall et al., 2004). The attempts of deceptive persons to manipulate their behavior to match the intended perception is known as the self-presentational perspective on deception (DePaulo, 1992). To maintain credibility, deceptive individuals may engage in different impression management and information management strategies depending on the context of the situation (Hartwig et al., 2007, 2010; Volbert & Steller, 2014). The distinct circumstances of 911 calls may therefore present difficulties when utilizing CBCA or RM criteria to examine the content of the calls.

Due to the unique nature of 911 calls, researchers have recently sought alternative methods to examine calls for indicators of veracity and deception. Although still limited, this field of study has increased over the past decade (Cromer et al., 2019; Elms & Keatley, 2023; Markey et al., 2022). The primary catalyst for this recent interest can largely be attributed to a 2009 study that explored 911 calls involving reports of homicides (Harpster et al., 2009). Specifically, the study examined differences between callers who were involved in the homicide and callers who were not. Harpster et al. (2009) hypothesized that although both types of callers might be influenced by the emotional impact of witnessing or engaging in violence, the verbal content of truthful callers would differ from that of deceptive

callers. Harpster et al. (2009) based this idea, in part, on previous research conducted by DePaulo and Morris (2004), which found that truthful individuals were more likely than deceptive persons to correct previous erroneous statements, while deceivers repeated words or phrases more often than truthful persons. Harpster et al. (2009) also relied on Olsson (2004), who suggested that truthful callers reporting fire emergencies displayed a sense of urgency, were cooperative, and provided relevant information with appropriate voice inflection. Hoax callers, on the other hand, stalled before providing details and lacked cooperation. Harpster et al. (2009) found that truthful and deceptive callers (referred to as "innocent" and "guilty" callers) differed in several ways. Innocent callers exhibited a greater sense of urgency, demanded help, immediately provided the location where help was needed, corrected statements after learning additional information, and exhibited voice modulation (i.e., a normal and audible voice with clear inflection and variance in tone). By contrast, guilty callers were more likely to provide extraneous information, invoke ingratiating language, repeat words and phrases, resist answering questions, and provide conflicting facts. Harpster et al. (2009) also found that innocent callers were more likely than guilty callers to refuse to accept the victim's death, as postulated by the five stages of grief, in which denial is the first stage (Kübler-Ross & Kessler, 2005).

Harpster and Adams (2017) then identified additional indicators to supplement their original 2009 study and developed a checklist entitled, "911 Considering Offender Probability in Statements (COPS) Scale." As a foundation for these additional indicators, Harpster and Adams (2017) relied partially upon Grice's (1975) four maxims of communication (manner, quantity, quality, and relation) and DePaulo et al.'s (2003) findings that deceptive individuals were less committed to their statements than truthtellers. Several minor revisions after 2017 led to additional indicators being included, and, as a result, the final iteration of the COPS Scale included a total of 18 "innocent" indicators and 40 "guilty" indicators.

After its publication in 2017, several studies attempted to examine the efficacy of items listed on the COPS Scale and identified by Harpster et al. (2009) as distinguishing between innocent and guilty callers. However, these studies were largely unsuccessful. Cromer et al. (2019) evaluated fifty 911 calls containing reports of homicides and suicides and found that seven of the nine indicators proposed by Harpster et al. (2009) did not discriminate between guilty and innocent callers. Further, the two indicators that were found to be associated with guilty callers (conflicting facts and extraneous information) were only present in up approximately 10% of all calls, suggesting these indicators have limited practical value.

Subsequent to Cromer et al. (2019), Miller et al. (2021) analyzed 175 homicide and suicide calls using indicators from Harpster et al. (2009). Three of the indicators examined (use of the word "just" as a minimizer, urgency to the plea for help, and voice modulation) differentiated between innocent and guilty callers. However, two of the three (urgency to the plea for help and voice modulation), predicted the opposite type of caller than the COPS Scale would suggest. O'Donnell et al. (2022) investigated 43 of the COPS Scale indicators across seventy 911 calls involving truthful and deceptive callers in missing child cases, where truthful callers reported a child missing but were not involved in the disappearance, and deceptive callers falsely reported a child missing or abducted and were responsible for the disappearance. Only six of the 43 indicators (attempts to convince, initial delays, inappropriate order, evasion, equivocation, and awkward phrases) demonstrated an ability to discriminate between callers. However, the authors noted that effects for two of these (equivocation and awkward phrases) may have been due to the coding methods utilized, as calls containing a single instance of equivocation were coded the same as calls containing multiple instances, and the subjective definition of awkward phrases may have produced variations in coder interpretations.

The above studies were limited in that they only tested a portion of the COPS Scale indicators identified by Harpster and Adams (2017) and Harpster et al. (2009)or tested call types that differed from the homicide calls explored by Harpster et al. (2009; Cromer et al., 2019; Miller et al., 2021; O'Donnell et al., 2022). Other studies have examined alternative methods of distinguishing between truthful and deceptive 911 calls with varying success, but none have tested the ability of the COPS Scale in its entirety to discriminate between innocent and guilty callers (Markey et al., 2022, 2023; O'Donnell et al., 2023). This is problematic, as the COPS Scale is offered to law enforcement as an investigative tool for differentiating innocent from guilty 911 callers (Droney, 2023; Harpster, 2024).

Present Study

The lack of empirical data supporting the ability of the COPS Scale to discriminate between innocent and guilty callers has important implications for its use in law enforcement investigations. 911 calls may contain valuable details that can be compared to other evidence or subsequent interviews. However, improper judgments of a caller's innocence or guilt can have negative consequences for an investigation and may influence the types of interview questions asked by law enforcement or the prioritization of limited resources. The current study aimed to examine the efficacy of indicators on the COPS Scale in discriminating between innocent and guilty 911 callers reporting homicides, suicides, or serious physical injuries that ultimately resulted in fatalities. The current study also explored the impact participant judgments had on the development of interview questions to emulate what occurs during actual investigations. To mirror the study by Harpster et al. (2009), only calls reporting the death, or serious physical injury resulting in death, of another person were included and were divided into two call types. A true report caller (TRC) legitimately reported a homicide, suicide, or serious injury leading to death, and was not responsible for the injury or death. By contrast, a false report caller (FRC) was responsible for the individual's death but concealed this information, blamed the death on another individual, and/or staged the death to appear as a suicide. The current study, therefore, had three primary objectives:

1. The first objective was to examine the extent to which law enforcement officers agree on the presence of COPS Scale indicators within 911 calls.
2. The second objective was to examine associations between the indicators on the COPS Scale and participant judgments of callers as guilty or innocent, along with the ability of the indicators to discriminate between TRCs and FRCs.
3. The third objective was to explore how judgments of innocence or guilt impacted potential interview questions asked by law enforcement.

Method

Participants

Forty-eight sworn federal, state, and local law enforcement officers assigned to federal task forces and/or field divisions of a federal law

enforcement agency across the United States were recruited for this study. Because the research took place in a law enforcement agency, our sample size was based on resource limitations, as we were required to minimize the number of participants and the amount of time spent by each participant on the study to prevent disruption with operational needs and active investigations. All officers volunteered to participate by responding to a canvass to field divisions of a federal law enforcement agency requesting volunteers to assist with research involving 911 calls. All participants had professional experience conducting investigations of a variety of criminal violations, including violent crime. On average, participants had 14.13 ($SD = 7.31$) years of total law enforcement experience, with a range of 1 year to 32 years. None of the participants advised they had previously received training on the COPS Scale. All participants completed the informed consent process in compliance with human subject protection regulations, and all procedures were reviewed and approved by an institutional review board. This study was not preregistered.

Materials

Calls

911 calls were identified through various sources, including federal, state, local, and military law enforcement agencies in the United States. Calls were also obtained from the Internet via news agencies (such as local affiliates of NBC, CBS, and ABC news organizations) and other online media (e.g., YouTube), wherein the call was found to be complete and unredacted. Calls were only included in the sample when the resolution of the associated case was verifiable. Call types included in the current study fit one of two potential outcomes: homicides or suicides in which the caller was not involved (TRC), and homicides in which the caller was responsible (FRC) and either concealed this from dispatchers, placed blame on other individuals, and/or staged the homicides to appear as suicides. The sample included calls in which the subject of the call (the victim) was still alive at the time the call was placed, but who subsequently died from the injuries sustained, as well as calls in which the victim was already deceased at the time the call was placed.

To determine whether a call represented a TRC or FRC, a combination of external sources was reviewed. These sources included official court records, law enforcement reports, and media reports. To be considered an FRC, the caller must have been criminally convicted in a court of law for specific offenses related to the death of the victim in the call. In four cases, a call was deemed to be an FRC after the caller was arrested but prior to conviction. In these limited circumstances, this decision was based upon the totality of available records, such as forensic evidence, physical evidence, and subsequent caller statements to law enforcement. For TRCs, either a third party was convicted of charges related to the homicide in which the caller was not involved, or the caller was exonerated by law enforcement through verification of alibis, forensic evidence, and/or other physical evidence. This method for classifying calls as TRCs and FRCs is consistent with past 911 research (Cromer et al., 2019; Markey et al., 2022, 2023; Miller et al., 2021; O'Donnell et al., 2022, 2023).

A total of 60 calls were identified for use as stimuli, 30 of which were TRCs and 30 of which were FRCs. The number of calls in the study was determined by resource limitations, as we were limited in the number of available 911 calls, including those in which a caller was involved in the reported death. During the course of data collection, four calls were excluded from the sample. One was excluded due to the caller having admitted to being responsible for the homicide in the call; a second was excluded due to the victim being a nationally recognized figure and the circumstances of his death being widely published in national media outlets. A third call was excluded due to conflicting decisions by juries in separate criminal trial proceedings and a fourth call was excluded as it was a duplicate of another call in the sample. This resulted in a final sample of 56 calls, which included 27 FRCs (48%) and 29 TRCs (52%).

COPS Scale

For each assigned call, participants were asked to complete the COPS Scale to aid in determining the innocence or guilt of the 911 callers. The revised COPS Scale consists of 18 innocent indicators and 40 guilty indicators divided into multiple sections that assess different aspects of the call. Sixteen innocent indicators each have a single corresponding guilty indicator, while two (Aggressive Demands and Free Flow of Information) do not. The remainder of the COPS Scale consists of several guilty indicators with no corresponding single innocent indicator. Each section, with the exception of "Additional Guilty Indicators," also includes a "not applicable (N/A)" option for any variables that do not apply to the call. Examples of the 16 forced choice indicators include Immediate Plea for Help/No Immediate Plea for Help; Focus on Victim/Focus on Caller; Relevant Information/Extraneous Information; and Urgency/Lack of Urgency. Examples of the remaining guilty indicators (i.e., those without corresponding innocent indicators) are Minimizes, Equivocation, Self-interruption, Attempts to Convince, Awkward Phrases, and Conflicting Facts. See the Appendix for a complete list of COPS Scale indicators.

Procedure

Participants were informed that the purpose of the study was to examine the ability of the COPS Scale to differentiate between innocent and guilty callers. Participants completed a training provided by the research team on the appropriate completion procedure of the COPS Scale. Prior to developing this training, members of the research team obtained training regarding the findings of Harpster et al. (2009), consulted with other law enforcement personnel who were trained on the COPS Scale, reviewed investigative case files in which the COPS Scale had been utilized in criminal investigations, and studied the publications by Harpster and Adams (2017) and Harpster et al. (2009). Participants were provided with a written copy of each indicator's definition, in accordance with the COPS Scale definitions utilized by Harpster and Adams (2017). Participants could refer to these definitions at any time to ensure consistent interpretation of the indicators. Participants were instructed to check the associated innocent or guilty indicator if a specific variable was present in the call. If a specific variable did not apply to a call, participants were instructed to check "N/A" in accordance with the procedures specified by Harpster and Adams (2017).

A nested design was used to maximize the number of observations and calls used as stimuli. Following the completion of the training, participants were randomly assigned to one of eight groups of six. The 60 available calls were divided randomly among the six groups,

for a total of 15 calls per group. This resulted in different proportions of TRCs and FRCs between groups of calls. A portion of each set of the 15 calls was assigned to two overlapping groups. Thus, each call was heard by participants from two different groups. This was done to reduce within-group effects. See the online supplemental materials for a figure depicting assignment of calls to groups.

All participants were blind to whether the assigned calls were TRCs or FRCs. Participants were instructed that the calls were randomly assigned and to not assume an even number of TRCs and FRCs. Participants were free to choose the order in which they responded to their assigned calls. Additionally, participants were told they were not required to complete all of the calls at the same time but could respond to the calls over the course of several days. This was done to prevent fatigue and to minimize operational disruptions. Each participant only had access to his or her assigned calls and rated the calls independently from other participants in their assigned group. Participants were instructed to listen to each call and complete a corresponding COPS Scale form for the call. Call transcripts were also made available to participants to aid in their assessments. Participants were instructed to not assess any calls with which they had familiarity of the associated case. Two participants attrited from the study, resulting in two groups of five and six groups of six participants. A total of 644 observations were available for analysis after accounting for removal of calls, participant attrition, and participant case familiarity. See the online supplemental materials for a breakdown of the total number of observations available for analysis.

After completing the COPS Scale, participants were asked to provide a guilty or innocent judgment of each caller based on the totality of the completed COPS Scale (i.e., whether more guilty or innocent indicators were marked). In support of their determination of guilt or innocence, participants were also asked to list five indicators from the COPS Scale upon which they relied most heavily when making their determination of guilt or innocence. Finally, participants were asked to provide five questions they would ask the caller in a subsequent interview. From these hypothetical questions, two coders from the research team determined the focus of the question (caller, victim, third party, or scene), timeframe of interest (incident or general background), and type of question (open-ended, probing, or closed). One researcher coded 100% of the questions, while the second coded a random subset of 33% of the questions. Questions that were coded by both researchers were reviewed and any disagreements in coding were resolved via discussion.

Data Analysis

Fleiss' kappa was used to determine the agreement between participants in their ratings of indicator presence and guilt determination. Fleiss' kappa was bootstrapped with 1,000 iterations to generate 95% confidence intervals (CIs). Percent agreement was also calculated due to potential bias in Fleiss' kappa for indicators with low frequency.

Due to the design of the COPS Scale, different strategies were used for indicators that had a forced choice/contrasting indicator, and indicators that did not have a forced choice/contrasting item. Indicators with forced choice/contrasting indicator were directly compared, with the innocent indicator serving as the reference group and the guilty indicator serving as the comparison. For these indicators, selection of "N/A" was treated as missing. It should be noted there was a total of six observations in which the participant selected both the guilty and innocent indicators (four observations for Bleeding Comments/Blood/Brains Comments; and two observations for Relevant Information/Extraneous Information), thus not allowing for a comparison of the indicators. All six observations were treated as missing. Indicators without a forced choice/contrasting indicator were analyzed individually, with selection of the indicator (present) coded as 1 and no selection of the indicator (absent) coded as 0. If the participant did not select any innocent or guilty indicators and instead selected "N/A," these instances were treated as missing. For both participant judgments of callers as innocent or guilty and type of call, innocent/TRC served as the reference group and guilty/FRC served as the comparison.

We used binomial logistic regression to examine accuracy of participant judgments of callers as innocent or guilty, associations between participant selection of COPS Scale indicators and ratings of callers as innocent or guilty, and whether COPS Scale indicators were more common in TRCs or FRCs. For analyses examining participant accuracy, correct identification of caller guilt/innocence was regressed on type of call. The intercept model was used to examine whether overall accuracy was statistically better than chance. To explore associations between COPS Scale indicators and participant judgment of callers as innocent or guilty, participant judgment of calls was regressed on each indicator. For analyses examining whether COPS Scale items were more common in TRCs or FRCs, type of call was regressed on each indicator. It should be noted that we were unable to run analyses on the Lack of Contractions indicator due to it only being selected a total of three times across observations. Binomial logistic regression was also used to assess the associations between interview question characteristics and participant ratings of truthfulness and deception, with participant judgments of callers predicting question characteristics.

All analyses were conducted using R (R Core Team, 2023). Fleiss' kappa and percent agreement were run using the irr (Gamer et al., 2019) package. Bootstrapping of Fleiss' kappa was conducted using the boot (Proutskova & Gruszczynski, 2020) package. Binomial logistic regressions were run using the glm function in R. Statistical significance was set at $p < .05$.

Transparency Statement

Regarding the materials used in the study, copyright restrictions do not allow us to publicly archive the 911 COPS Scale; readers seeking access are advised to contact the Copyright Clearance Center or Taylor & Francis Group. Legal restrictions prevent the authors from publicly releasing a majority of the 911 calls used as stimuli, as the calls remain the property of state and local law enforcement agencies as part of investigative case files. For calls found via open-source methods, readers seeking access should contact the lead author.

Research data are not shared. The raw/processed data are part of an ongoing study. There are also legal and ethical restrictions on the sharing of the data, as the data contain details regarding criminal investigations that are not public. Study analysis code is publicly available on the Open Science Framework (https://osf.io/3gkpd/; O'Donnell et al., 2024).

Results

Participant Agreement in Indicator Presence and Guilt Determination

There was slight agreement among participants for 23 of the 40 guilty indicators ($\kappa < .20$). Although the agreement for the remaining

indicators was either fair or moderate, all were still below the accepted level of .60 for adequate agreement ($\kappa \leq .49$; Landis & Koch, 1977). It should be noted that the Fleiss' kappa for the indicator lack of contractions was at chance levels, but the percent agreement was over 90%; this discrepancy was due to the indicator only being selected three times total across all observations. Similarly, participant judgments of callers as guilty or innocent showed fair agreement ($\kappa = .31$) but was still below the accepted level for reliability. See Table 1 for Fleiss' kappa and bootstrapped 95% CIs and percent agreement for all indicators and for guilt determination.

As a result of the poor agreement, caution should be used when interpreting the results related to validity of indicators and participants' judgments of callers as guilty or innocent. We present the results of the analyses because the purpose of the current study was to examine both the reliability and validity of COPS Scale indicators.

Accuracy of Participant Judgments of Callers as Guilty or Innocent

Overall, participants correctly identified the veracity of 58.0% of calls ($z = 4.01$, $p < .001$, ORs $= 1.39$), with no differences in accuracy of identification between TRCs and FRCs, Wald's $\chi^2(1) = 0.12$, $p = .724$, $OR = 0.95$, 95% CI [0.69, 1.29].

COPS Scale Indicators' Association With Judgments of Guilt or Innocence and Type of Call

Who Is the Call About?

All guilty indicators in the "Who Is the Call About?" section of the COPS Scale were associated with increased likelihood of participants

Table 1
Participant Agreement in Ratings of Indicators and Guilt

Indicator	κ	95% CI	% agreement
Guilty indicators			
No Immediate Plea for Help for Victim	.20	[0.11, 0.29]	17.9
No Immediate Assessment of Victim	.27	[0.20, 0.35]	21.7
Focus on Caller	.24	[0.15, 0.34]	22.2
Acceptance of Victim's Death	.35	[0.28, 0.42]	22.9
No Aid Provided to Victim	.46	[0.37, 0.56]	39.3
Extraneous Information	.17	[0.09, 0.26]	22.5
No Immediate Sensory Details	.24	[0.16, 0.33]	15.9
Inappropriate Order	.20	[0.13, 0.28]	13.6
Blood/Brains Comments	.49	[0.42, 0.57]	39.7
Inappropriate Delivery	.28	[0.19, 0.37]	22.9
Lack of Urgency	.43	[0.33, 0.52]	34.9
No Fear for Caller's Safety	.22	[0.13, 0.31]	32.1
No Proximity to Victim	.33	[0.24, 0.43]	30.5
Initial Delays	.14	[0.08, 0.19]	8.5
Failure to Address or Coach the Victim	.25	[0.17, 0.33]	12.0
Defendant Mentality	.12	[0.05, 0.19]	5.4
Ingratiating Remarks	.10	[0.04, 0.17]	8.1
Mental Miscues	.11	[0.05, 0.18]	5.4
Minimizes	.11	[0.06, 0.17]	4.1
Diversion	.01	[−0.04, 0.06]	42.2
Equivocation	.04	[−0.04, 0.12]	30.1
Evasion	.10	[0.01, 0.20]	31.3
Hangs Up	.12	[0.01, 0.25]	53.0
Only Answers What's Asked	.15	[0.07, 0.23]	26.5
Pauses	.15	[0.04, 0.25]	33.7
Repetition	.16	[0.08, 0.26]	39.8
Self-interruption	.04	[−0.03, 0.13]	54.2
Short Answers	.15	[0.05, 0.26]	33.7
Unintelligible Comments	.17	[0.09, 0.25]	26.5
Dispatcher Confusion	.23	[0.13, 0.34]	25.0
Attempts to Convince	.21	[0.08, 0.37]	41.7
Awkward Phrases	.09	[0.01, 0.19]	36.9
Conflicting Facts	.14	[0.06, 0.24]	48.8
"Huh?" Factor	.14	[0.01, 0.29]	48.8
"I Don't Know"	.31	[0.20, 0.42]	35.7
Insult/Blame Victim	.30	[0.14, 0.48]	67.9
Isolated "Please"	.12	[0.04, 0.22]	72.6
Lack of Contractions	−.01	[−0.02, 0.00]	90.5
No Modulation	.16	[0.08, 0.25]	47.6
Unexplained Knowledge	.12	[0.03, 0.21]	57.1
Innocent indicators			
Aggressive Demands	.17	[0.10, 0.24]	6.8
Free Flow of Information	.23	[0.14, 0.32]	20.5
Guilt determination	.31	[0.22, 0.41]	24.4

Note. CI = confidence interval.

classifying callers as guilty instead of innocent, Wald's $\chi^2(1) \geq 28.01$, $p < .001$, $ORs \geq 2.60$.

However, only one guilty indicator was more likely to be present in FRCs than TRCs—Focus on Caller, Wald's $\chi^2(1) = 8.09$, $p = .005$, $OR = 1.64$, 95% CI [1.17, 2.32]. Acceptance of Victim's Death and No Aid Provided to Victim were more likely to appear in TRCs than FRCs—Acceptance of Victim's Death: Wald's $\chi^2(1) = 22.76$, $p < .001$, $OR = 0.42$, 95% CI [0.30, 0.60]; No Aid Provided to Victim: Wald's $\chi^2(1) = 20.46$, $p < .001$, $OR = 0.43$, 95% CI [0.30, 0.62]. No Immediate Plea for Help for Victim and No Immediate Assessment of Victim were no more likely to be present in FRCs than TRCs—Wald's $\chi^2(1) \leq 0.25$, $p \geq .614$.

What Is the Call About?

All "What Is the Call About?" guilty indicators were associated with greater odds of a caller being classified as guilty over innocent—Wald's $\chi^2(1) \geq 35.59$, $p < .001$, $ORs \geq 5.12$.

Regarding associations with type of call, two guilty indicators were more likely to be present in FRCs than TRCs: Extraneous Information, Wald's $\chi^2(1) = 10.36$, $p = .001$, $OR = 1.77$, 95% CI [1.25, 2.52], and Inappropriate Order, Wald's $\chi^2(1) = 6.61$, $p = .010$, $OR = 1.54$, 95% CI [1.11, 2.16]. No Immediate Sensory Details and Blood/Brains Comments were equally likely to appear in FRCs and TRCs, Wald's $\chi^2(1) \leq 3.20$, $p \geq .074$.

How Is the Call Made?

All guilty indicators in the "How Is the Call Made?" section were associated with increased likelihood of participants judging a caller as guilty instead of innocent, Wald's $\chi^2(1) \geq 7.49$, $p \leq .006$, $ORs \geq 2.19$.

However, only one guilty indicator, Inappropriate Delivery, was more likely to be present in FRCs than TRCs, Wald's $\chi^2(1) = 7.05$, $p = .008$, $OR = 1.54$, 95% CI [1.12, 2.13]. All other guilty indicators in this section (Lack of Urgency, No Fear for Caller's Safety, Initial Delays, No Proximity to Victim, and Failure to Address or Coach the Victim) were equally likely to be present in FRCs and TRCs, Wald's $\chi^2(1) \leq 3.10$, $p \geq .079$.

Passive Defenses

All guilty indicators under "Passive Defenses" were significantly associated with increased likelihood of judging callers to be guilty, Wald's $\chi^2(1) \geq 19.63$, $p < .001$, $ORs \geq 4.47$.

Defendant Mentality was the only guilty indicator more likely to appear in FRCs than TRCs, Wald's $\chi^2(1) = 16.45$, $p < .001$, $OR = 3.20$, 95% CI [1.80, 5.91]. The remaining guilty indicators (Ingratiating Remarks, Mental Miscues, and Minimizes) were no more likely to appear in FRCs than TRCs, Wald's $\chi^2(1) \leq 1.59$, $p \geq .210$.

Withholding Information

All indicators in the "Withholding Information" section significantly predicted participants judging callers to be guilty over innocent, Wald's $\chi^2(1) \geq 12.75$, $p < .001$, $ORs \geq 2.81$.

Three guilty indicators were more likely to be present in FRCs than TRCs: Repetition, Wald's $\chi^2(1) = 8.06$, $p = .005$, $OR = 1.88$, 95% CI [1.21, 2.94]; Evasion, Wald's $\chi^2(1) = 6.19$, $p = .014$, $OR = 1.77$, 95% CI [1.13, 2.81]; and Short Answers, Wald's $\chi^2(1) = 5.54$, $p = .019$, $OR = 1.65$, 95% CI [1.09, 2.52]. One guilty indicator, Hangs Up, was more common in TRCs than FRCs, Wald's $\chi^2(1) = 4.58$, $p = .037$, $OR = 0.51$, 95% CI [0.26, 0.95]. All other guilty indicators (Equivocation, Diversion, Only Answers What's Asked, Pauses, Self-interruption, and Unintelligible Comments) were no more likely to appear in FRCs than TRCs, Wald's $\chi^2(1) \leq 3.05$, $p \geq .084$.

Dispatcher Comprehension/Confusion

Although Dispatcher Confusion was associated with greater likelihood of judging a caller as guilty, Wald's $\chi^2(1) = 110.88$, $p < .001$, $OR = 8.47$, 95% CI [5.49, 13.44]), this indicator was equally likely to appear in FRCs and TRCs, Wald's $\chi^2(1) = 3.65$, $p = .056$.

Additional Guilty Indicators

All "Additional Guilty Indicators" were associated with greater odds of being judged as guilty over innocent, Wald's $\chi^2(1) \geq 7.61$, $p \leq .004$, $ORs \geq 2.95$.

Four indicators in this section were more likely to occur in FRCs than TRCs: Conflicting Facts, Wald's $\chi^2(1) = 18.87$, $p < .001$, $OR = 2.70$, 95% CI [1.71, 4.34]; Insult/Blame Victim, Wald's $\chi^2(1) = 18.48$, $p < .001$, $OR = 3.71$, 95% CI [2.00, 7.33]; Attempts to Convince, Wald's $\chi^2(1) = 15.28$, $p < .001$, $OR = 2.29$, 95% CI [1.51, 3.53]; and Awkward Phrases, Wald's $\chi^2(1) = 3.10$, $p = .014$, $OR = 1.70$, 95% CI [1.11, 2.61]. In contrast, Unexplained Knowledge was more likely to appear in TRCs than FRCs, Wald's $\chi^2(1) = 6.58$, $p = .012$, $OR = 0.50$, 95% CI [0.28, 3.820.85]. The remaining guilty indicators in this section ("Huh?" Factor, "I Don't Know," Isolated "Please," and No Modulation) were equally likely to appear in FRCs and TRCs, Wald's $\chi^2(1) \leq 2.62$, $p \geq .106$.

See Table 2 for odds ratios, 95% CIs, and probabilities of guilty indicators present in TRCs and FRCs and participant judgments of calls associated with each guilty indicator. See Figure 1 for plot of odds ratios and 95% CIs of COPS Scale guilty indicators' associations with participant judgments of callers as guilty or innocent and type of call.

Innocent Indicators

Aggressive Demands and Free Flow of Information predicted judgments of callers as innocent, Wald's $\chi^2(1) \geq 165.04$, $p < .001$, $ORs = 0.04$. Both were more likely to occur in TRCs than FRCs—Aggressive Demands: Wald's $\chi^2(1) = 5.98$, $p = .015$, $OR = 0.58$, 95% CI [0.37, 0.90]; Free Flow of Information: Wald's $\chi^2(1) = 6.58$, $p = .011$, $OR = 0.66$, 95% CI [0.48, 0.91].

See Table 3 for odds ratios, 95% CIs, and probabilities of innocent indicators present in TRCs and FRCs and participant judgments of calls associated with the innocent indicators.

Most Common and Most Influential Indicators

Across calls, the most commonly identified innocent indicators were Dispatcher Comprehension ($N = 444$), Relevant Information ($N = 437$), Focus on Victim ($N = 395$), Immediate Assessment of Victim ($N = 389$), and Proximity to Victim ($N = 367$). Of these, Focus on Victim, Relevant Information, and Immediate Assessment of Victim were also indicated by participants as being most influential on their guilt determinations ($N = 204$, $N = 189$,

Table 2
Associations Between Guilty Indicators and Participant Judgment of Callers as Guilty or Innocent and Type of Call

Guilty indicator	Participant judgments				Type of call			
	Pr guilty (%)	*Pr* innocent (%)	*OR*	95% CI	*Pr* false (%)	*Pr* true (%)	*OR*	95% CI
No Immediate Plea for Help for Victim	70.3	21.3	8.93*	[6.17, 13.11]	49.5	48.5	1.04	[0.76, 1.44]
No Immediate Assessment of Victim	82.3	29.1	11.26*	[7.62, 16.95]	49.5	47.4	1.09	[0.79, 1.50]
Focus on Caller	86.5	27.4	16.97*	[10.85, 27.41]	57.1	44.8	1.64*	[1.17, 2.32]
Acceptance of Victim's Death	60.9	37.5	2.60*	[1.82, 3.73]	36.3	57.6	0.42*,‡	[0.30, 0.60]
No Aid Provided to Victim	61.2	27.8	4.11*	[2.80, 6.11]	42.9	63.5	0.43*,‡	[0.30, 0.62]
Extraneous Information	91.8	29.8	26.36*	[15.43, 48.26]	58.0	43.8	1.77*	[1.25, 2.52]
No Immediate Sensory Details	75.8	24.8	9.46*	[6.53, 13.89]	52.8	45.4	1.35	[0.97, 1.87]
Inappropriate Order	90.1	20.8	34.79*	[21.48, 58.67]	54.1	43.5	1.54*	[1.11, 2.16]
Blood/Brains Comments	64.4	26.1	5.12*	[2.96, 9.02]	43.2	33.8	1.48	[0.87, 2.53]
Inappropriate Delivery	89.7	18.5	38.54*	[24.24, 63.39]	54.8	44.1	1.54*	[1.12, 2.13]
Lack of Urgency	81.0	26.0	12.14*	[8.32, 18.00]	49.8	47.9	1.08	[0.79, 1.49]
No Fear for Caller's Safety	56.8	37.5	2.19*	[1.25, 3.94]	46.2	44.4	1.07	[0.61, 1.88]
No Proximity to Victim	70.9	35.2	4.49*	[3.15, 6.44]	42.5	50.0	0.74	[0.53, 1.03]
Initial Delays	84.8	31.8	11.90*	[7.51, 19.46]	52.2	45.1	1.33	[0.92, 1.93]
Failure to Address or Coach the Victim	63.9	32.8	3.63*	[2.32, 5.76]	46.2	55.8	0.68	[0.44, 1.05]
Defendant Mentality	88.1	54.4	6.18*	[3.00, 14.45]	73.1	46.0	3.20*	[1.80, 5.91]
Ingratiating Remarks	85.0	55.9	4.47*	[2.21, 10.03]	46.8	52.4	0.80	[0.45, 1.39]
Mental Miscues	94.6	52.6	19.24*	[6.92, 80.06]	58.2	49.8	1.41	[0.83, 2.44]
Minimizes	83.9	48.6	5.49*	[3.24, 9.70]	56.9	66.7	1.13	[0.73, 1.76]
Diversion	93.5	44.8	17.63*	[6.34, 73.34]	60.9	47.6	1.72	[0.94, 3.23]
Equivocation	81.6	43.71	5.70*	[3.20, 10.84]	56.5	47.4	1.45	[0.90, 2.37]
Evasion	82.4	42.5	6.35*	[3.69, 11.56]	60.5	46.2	1.77*	[1.13, 2.81]
Hangs Up	71.1	46.7	2.81*	[1.48, 5.65]	34.6	51.0	0.51*,‡	[0.26, 0.95]
Only Answers What's Asked	82.1	37.4	7.68*	[4.93, 12.35]	49.0	48.2	1.03	[0.71, 1.48]
Pauses	80.2	42.2	5.55*	[3.36, 9.56]	53.9	47.9	1.27	[0.83, 1.95]
Repetition	86.9	41.2	9.45*	[5.32, 18.15]	61.7	46.2	1.88*	[1.21, 2.94]
Self-interruption	77.1	46.8	3.84*	[1.79, 9.18]	36.3	51.9	0.53	[0.25, 1.08]
Short Answers	78.9	41.9	5.19*	[3.22, 8.67]	59.0	46.5	1.65*	[1.09, 2.52]
Unintelligible Comments	68.9	42.9	2.96*	[1.97, 4.50]	53.1	47.4	1.26	[0.86, 1.86]
Dispatcher Confusion	82.1	35.1	8.47*	[5.49, 13.44]	55.2	46.5	1.42	[0.99, 2.04]
Attempts to Convince	88.4	40.4	11.24*	[6.37, 21.50]	65.2	45.1	2.29*	[1.51, 3.53]
Awkward Phrases	89.6	40.7	12.59*	[6.87, 25.42]	59.5	46.2	1.70*	[1.11, 2.61]
Conflicting Facts	88.4	42.3	7.98*	[4.55, 15.00]	68.1	44.1	2.70*	[1.71, 4.34]
"Huh?" Factor	92.9	43.4	16.96*	[7.42, 49.02]	48.5	48.5	1.00	[0.61, 1.65]
"I Don't Know"	74.6	39.2	4.54*	[3.10, 6.76]	53.7	46.5	1.33	[0.94, 1.90]
Insult/Blame Victim	81.5	45.8	5.21*	[2.68, 11.14]	75.9	46.0	3.71*	[2.00, 7.33]
Isolated "Please"	52.2	47.5	2.95*	[1.39, 6.79]	51.5	48.2	1.14	[0.56, 2.31]
Lack of Contractions[a]								
No Modulation	83.2	41.4	7.00*	[4.24, 12.13]	44.8	50.0	0.81	[0.54, 1.22]
Unexplained Knowledge	87.3	44.6	8.54	[4.23, 19.70]	33.3	50.0	0.50*,‡	[0.28, 0.85]

Note. *Pr* = probability; CI = confidence interval.
[a] Unable to complete analysis due to low frequency.
‡ Significant in opposite of predicted direction. * $p < .05$.

and $N = 135$, respectively). Participants also cited Urgency ($N = 163$) and Expected Delivery ($N = 147$) as highly influential indicators when determining guilt. Frequencies of all innocent indicators identified in TRCs and FRCs can be found in Table 4.

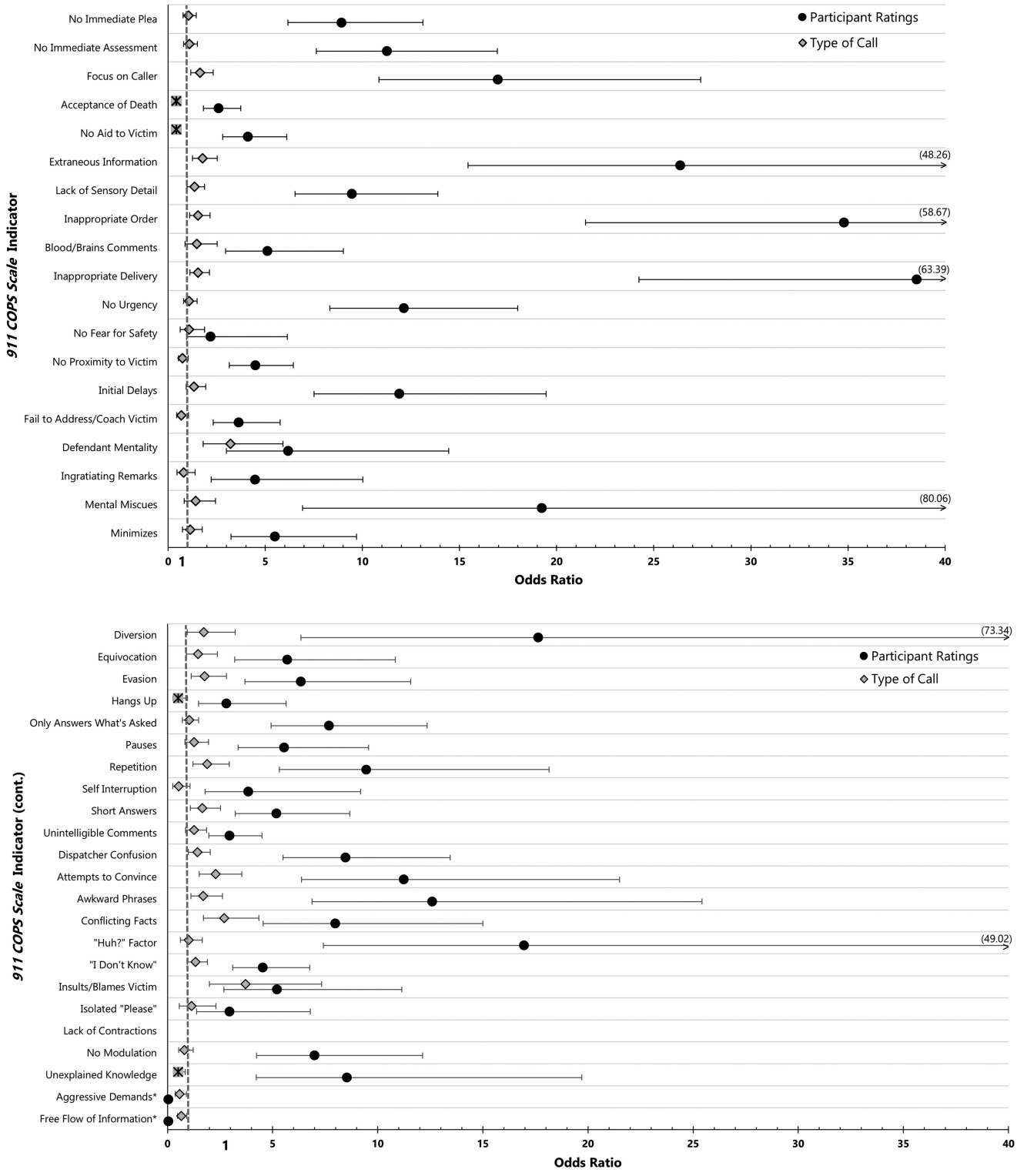
The most commonly identified guilty indicators were No Fear for Caller's Safety ($N = 458$), No Aid Provided to Victim ($N = 397$), No Immediate Plea for Help for Victim ($N = 368$), Failure to Address or Coach the Victim ($N = 300$), Acceptance of Victim's Death ($N = 299$), and No Immediate Sensory Details ($N = 299$). No Immediate Plea for Help for Victim, No Aid Provided to Victim, and Acceptance of Victim's Death were also three of the five indicators cited by participants as being most influential on their guilt determinations ($N = 161$, $N = 123$, and $N = 81$, respectively). The remaining two guilty indicators that most heavily influenced participants' guilt determinations were Lack of Urgency ($N = 154$) and Inappropriate Delivery ($N = 108$). Frequencies of all guilty indicators identified in TRCs and FRCs can be found in Table 5.

Focus of Potential Interview Questions

Participants' hypothetical interview questions were more likely to focus on the caller when they judged a caller to be guilty versus innocent, Wald's $\chi^2(1) = 36.61$, $p < .001$, $OR = 2.71$, 95% CI [1.96, 3.78]. Conversely, when participants rated the caller as innocent, their questions were more likely to focus on the victim, Wald's $\chi^2(1) = 7.12$, $p = .008$, $OR = 0.65$, 95% CI [0.47, 0.89]; a third party, Wald's $\chi^2(1) = 28.77$, $p < .001$, $OR = 0.36$, 95% CI [0.25, 0.53]; or the crime scene, Wald's $\chi^2(1) = 6.43$, $p = .011$, $OR = 0.40$, 95% CI [0.18, 0.81].

Figure 1

Forest Plot of ORs and 95% Confidence Intervals of COPS Scale Indicators' Associations With Participant Judgments of Callers as Guilty or Innocent and Type of Call



Note. *OR*s greater than 1 represent greater likelihood of judging a caller to be guilty/associated with FRCs. *OR*s less than 1 represent greater likelihood of judging a caller to be innocent/associated with TRCs. Grey boxes with a star denote indicators significant in the opposite direction predicted. Symbol * indicates innocent indicator. COPS = Considering Offender Probability in Statements; FRC = false report caller; TRC = true report caller.

Table 3
Associations Between Innocent Indicators and Participant Judgment of Callers as Guilty or Innocent and Type of Call

	Participant judgments				Type of call			
Innocent indicator	*Pr* guilty (%)	*Pr* innocent (%)	*OR*	95% CI	*Pr* false (%)	*Pr* true (%)	*OR*	95% CI
Aggressive Demands	19.7	87.4	0.04*	[0.02, 0.06]	56.9	69.5	0.58*	[0.37, 0.90]
Free Flow of Information	15.0	81.2	0.04*	[0.03, 0.06]	56.9	66.7	0.66*	[0.48, 0.91]

Note. *Pr* = probability; CI = confidence interval.
* $p < .05$.

Participants were more likely to ask closed (i.e., yes/no) questions when they judged the caller to be innocent, Wald's $\chi^2(1) = 16.72$, $p < .001$, $OR = 0.50$, 95% CI [0.35, 0.70]. In contrast, participants were more likely to ask probing questions when they judged the caller to be guilty, Wald's $\chi^2(1) = 17.35$, $p < .001$, $OR = 2.09$, 95% CI [1.47, 2.98]. Across calls, only 7.0% of questions asked by participants were open-ended, and participants were no more likely to ask open-ended questions when they judged the caller as guilty or innocent, Wald's $\chi^2(1) = 0.82$, $p = .366$.

When participants judged a caller to be innocent, they were more likely to ask general background questions, Wald's $\chi^2(1) = 19.72$, $p < .001$, $OR = 0.45$, 95% CI [0.32, 0.65]. Conversely, when participants judged callers to be guilty, their questions were more likely to surround the incident itself, Wald's $\chi^2(1) = 20.43$, $p < .001$, $OR = 2.29$, 95% CI [1.60, 3.33]. See Table 6 for probabilities, odds ratios, and 95% CIs for association between participant judgment of callers as guilty or innocent and question focus.

Discussion

The COPS Scale has been offered to law enforcement as a method of distinguishing between innocent and guilty 911 callers despite limited existing data that supports its use (Miller et al., 2021; O'Donnell et al., 2022). We therefore examined the reliability and validity of the COPS Scale to determine its utility in discriminating between TRCs and FRCs. We also explored how the COPS Scale indicators correlated with judgments of caller guilt or innocence and subsequent influence of these innocent and guilty judgments on hypothetical interview questions. Our results call into question the efficacy of the COPS Scale and demonstrate its limited utility in assisting judgments of callers' innocence or guilt.

With respect to our first objective of testing the reliability of COPS Scale items, problematic agreement among participants was observed for every indicator. Agreement for 23 of the 40 guilty indicators on the COPS Scale was below .20 ("slight agreement"; Landis & Koch, 1977). Of these 23 guilty indicators, 21 of them were standalone/

Table 4
Number of Innocent Indicators Identified by Participants

Innocent indicator	True calls	False calls	Total
Dispatcher Comprehension	239	205	444
Relevant Information[a]	245	192	437
Focus on Victim[a]	218	177	395
Immediate Assessment of Victim[a]	205	185	389
Proximity to Victim	183	184	367
Expected Delivery[a]	201	158	359
Prioritized Order	203	155	358
Urgency[a]	180	167	347
Initial Sounds or Comments	168	141	309
Free Flow of Information	171	130	301
Immediate Sensory Details	156	130	286
Immediate Plea for Help for Victim	131	123	254
No Acceptance of Death	93	127	220
Aid Provided to Victim	64	111	175
Bleeding Comments	91	47	138
Aggressive Demands	75	57	132
Address or Coach the Victim	51	65	116
Fear for Caller's Safety	31	25	56

[a] Innocent indicators that were most influential to participants' guilt determination.

Table 5
Number of Guilty Indicators Identified by Participants

Guilty indicator	False calls	True calls	Total
No Fear for Caller's Safety	212	246	458
No Aid Provided to Victim[a]	172	226	397
No Immediate Plea for Help for Victim[a]	182	186	368
Failure to Address or Coach the Victim	139	161	300
Acceptance of Victim's Death[a]	109	190	299
No Immediate Sensory Details	158	141	299
Lack of Urgency[a]	137	138	275
Inappropriate Delivery[a]	144	119	263
No Immediate Assessment of Victim	117	120	237
Inappropriate Order	126	107	233
No Proximity to Victim	97	132	229
Focus on Caller	114	87	201
Extraneous Information	105	75	180
Initial Delays	93	85	178
"I Don't Know"	93	80	173
Dispatcher Confusion	90	73	163
Only Answers What's Asked	75	78	153
Unintelligible Comments	70	62	132
Minimizes	67	57	124
No Modulation	51	63	114
Attempts to Convince	73	39	112
Short Answers	65	45	110
Awkward Phrases	63	43	106
Pauses	55	47	102
Repetition	61	38	99
Blood/Brains Comments	42	55	97
Conflicting Facts	66	31	97
Evasion	55	36	91
Equivocation	43	33	76
"Huh?" Factor	34	36	70
Defendant Mentality	49	18	67
Mental Miscues	39	28	67
Unexplained Knowledge	21	42	63
Ingratiating Remarks	28	32	60
Insult/Blame Victim	41	13	54
Diversion	28	18	46
Hangs Up	16	30	46
Self-interruption	13	23	36
Isolated "Please"	17	16	33
Lack of Contractions	1	2	3

[a] Guilty indicators that were most influential to participants' guilt determination.

Table 6
Interview Question Focus and Type by Guilt Determination

Question focus/type	Pr guilty (%)	Pr innocent (%)	*OR*	95% CI
Question focus				
Caller	72.7	50.5	2.71*	[1.96, 3.78]
Victim	31.8	41.9	0.65*	[0.47, 0.89]
Third party	15.6	33.6	0.36*	[0.25, 0.53]
Scene	22.6	7.6	0.40*	[0.18, 0.81]
Time period captured				
Incident	81.9	66.4	2.29*	[1.60, 3.33]
General background	20.6	36.4	0.45*	[0.32, 0.65]
Question type				
Open-ended	7.9	6.1	1.32	[0.72, 2.46]
Probing	78.7	36.2	2.09*	[1.47, 2.98]
Closed	24.4	39.5	0.50*	[0.35, 0.70]

Note. *Pr* = probability; CI = confidence interval.
* $p < .05$.

nonforced choice guilty indicators. Similarly, agreement for both standalone innocent indicators was low. Further, of the three guilty indicators with the highest levels of agreement (Blood/Brain Comments, No Aid Provided to Victim, and Lack of Urgency), two did not distinguish between TRCs and FRCs, and one was associated with TRCs instead of FRCs, opposite of COPS Scale expectations. The inadequate interrater reliability for every indicator on the COPS Scale calls into question the utility of the scale when applied to real cases in which law enforcement officers do not know the final outcome. It also hinders our ability to confidently determine the validity of individual indicators and the COPS Scale as a whole.

Our second objective was to examine whether the indicators assisted sworn law enforcement officers in judging innocence and guilt among callers and whether the indicators discriminated TRCs from FRCs. A small number of previous studies have shown that law enforcement officers may be able to discriminate between truthful and deceptive statements above chance levels in particular settings, with some accuracy rates as high as 65% to 72% (Mann et al., 2004; Vrij et al., 2006). This was not found in the present study, as use of the COPS Scale to judge TRCs and FRCs led to an overall accuracy rate of 58% among participants and accuracy rates did not differ between call types. Although statistically different from chance, the 58% accuracy rate does not reflect a practical improvement in distinguishing between innocent and guilty callers.

In terms of the indicators themselves, 24 of the 40 guilty indicators did not distinguish TRCs from FRCs, and four of these indicators (No Aid Provided to Victim, Acceptance of Death, Unexplained Knowledge, and Hangs Up) were opposite of COPS Scale expectations. Both standalone innocent indicators (Free Flow of Information and Aggressive Demands) were associated with both participant judgments of callers as being innocent and TRCs. Further, there was a wide discrepancy between each indicator's impact on participant judgments of caller innocence and guilt and the indicator's actual ability to differentiate TRC from FRC. Even for the minority of COPS Scale indicators that did discern TRC and FRC, the effect sizes were much smaller than the effect they had on participant judgments, and most were not large enough to be of practical use for investigators. Since 911 calls are often the catalyst for a law enforcement response in death investigations, little corroborating evidence may exist at the time of the call. Misjudging an individual to be innocent or guilty in the early stages of an investigation can negatively influence all future aspects of the investigation, from the interpretation of additional evidence as it is obtained, to the types of questions asked. With such small effect sizes, use of these indicators in practical settings could result in misinterpretation of a caller's role in the victim's death, which can damage other portions of the investigative process. These findings are consistent with prior research on COPS Scale indicators, which has largely been unable to replicate the results of Harpster et al. (2009) or find support for the majority of COPS Scale indicators to differentiate between guilty and innocent callers (Cromer et al., 2019; Miller et al., 2021; O'Donnell et al., 2022).

Several possible explanations exist for why use of the COPS Scale resulted in the mischaracterization of approximately four out of every 10 calls. First, the COPS Scale contains a number of indicators that have intuitive appeal or mirror widespread beliefs about how people are supposed to act during emergency situations. However, stereotypical beliefs about lying are often misguided (Hartwig et al., 2010; Strömwall et al., 2004). For example, the COPS Scale includes Lack of Urgency as an indicator of guilt, whereas Urgency is considered an innocent indicator. However, an innocent caller may not exhibit a sense of urgency for a variety of reasons other than guilt. Personality characteristics, prior experience with traumatic events, or current relationship status with the deceased individual may all impact the general demeanor of a caller. Indeed, although Lack of Urgency was the second most influential indicator of guilt among participants in the current study, and Urgency was the third most influential innocent indicator, neither were associated with type of call.

Additional guilty indicators with intuitive appeal yielded similar results. The items No Aid Provided to Victim and Acceptance of Death were both within the top five most influential indicators of guilt, but both were more likely to occur in TRCs than FRCs. The underlying assumption for the inclusion of these items on the COPS Scale is that an innocent caller will "cling to hope" the victim will survive, especially when a close relationship exists between the two (Harpster & Adams, 2017). This assumption is based upon the five stages of grief, in which denial is the first stage (Kübler-Ross & Kessler, 2005). However, these stages are often criticized as being oversimplistic, lacking empirical support, and being destructive to those who do not grieve in the presumed order (Avis et al., 2021; Stroebe et al., 2017). Even the original authors of the five stages of grief later acknowledged that "… there is not a typical response to loss as there is no typical loss. Our grief is as individual as our lives … not everyone goes through all of them or in a prescribed order" (Kübler-Ross & Kessler, 2009). Thus, the use of these two items as signs of guilt may cause investigators to incorrectly judge an innocent caller as being responsible for the victim's death. Indeed, not only were the associations of these two indicators in the opposite direction of COPS Scale expectations, but they were also some of the strongest effects of any indicator on the entire scale.

Reliance on stereotypical cues to deception was reflected in the COPS Scale indicators commonly cited by participants. Of the 10 most commonly selected indicators of guilt on the COPS Scale, only two (Inappropriate Delivery and Inappropriate Order) were significantly associated with FRCs. The remaining eight indicators either had no association with call type or were more common in TRCs. Despite the overall inability to discriminate between TRCs and FRCs, the 10 most commonly cited indicators on the COPS Scale occurred more often than all other guilty indicators combined.

Similarly, six of the top 10 indicators named by participants as most influential to their judgments of guilt were not actually

associated with guilty callers, and two of these six were more likely to occur in TRCs. Notably, of the top five indicators named as most influential to judgments of guilt, two had no association with call type (No Immediate Plea for Help for Victim and Lack of Urgency) and two were more common in TRCs (No Aid Provided to the Victim and Acceptance of Death).

A second explanation for the COPS Scale's poor performance is that multiple innocent indicators (e.g., Immediate Plea for Help for Victim, Immediate Assessment of Victim, Urgency, Proximity to Victim, Aggressive Demands, Fear for Caller's Safety, and Initial Sounds or Comments) appear susceptible to manipulation by deceptive callers attempting to create the perception of a panicked and distraught caller. Impression management and information management strategies are commonly utilized by deceivers to maintain credibility (Hartwig et al., 2007). On the other hand, truthful callers may be more concerned with relaying accurate information regardless of how it is perceived (Vrij, Leal, Mann, et al., 2018; Vrij, Leal, Jupe, & Harvey, 2018). These motivational differences among callers may result in COPS Scale users erroneously applying guilty indicators to innocent callers and vice versa. Further, an innocent caller may be untruthful about aspects of the incident to minimize embarrassment, conceal other criminal activity, or prevent discovery of other noncriminal behavior. The COPS Scale does not address these possibilities. For instance, none of the above innocent indicators on the COPS Scale were accurately associated with TRCs, and participants were no more accurate in judgments of innocent callers than of guilty callers.

Flaws in the design of the COPS Scale may have contributed to the issues described above. The COPS Scale contrasts with other empirically supported deception detection methods in that it contains far more guilty indicators than innocent. For instance, CBCA and RM contain mostly truthful indicators, and although the presence of these variables may signify truthfulness, their absence does not necessarily indicate deception (Amado et al., 2016; Sporer, 1997). Instructions for the COPS Scale, however, advise investigators that "… a majority of checks on either side of the COPS Scale …" signifies the call is consistent with innocence or guilt (Harpster & Adams, 2017). This is problematic, as guilty indicators outnumber innocent indicators by a margin of more than two-to-one. Innocent indicators potentially associated with TRCs may be overwhelmed by the greater volume of guilty indicators, regardless of whether the guilty items are actually associated with FRCs. Further, most of the top 10 commonly cited and influential variables were those in which a one-to-one ratio existed between guilty and innocent indicators. The forced choice nature of these items may cause COPS Scale users to weigh these more heavily than indicators with no opposing counterpart. The COPS Scale does not address this, but rather infers users should rely on a majority of either guilty or innocent indicators. However, with no standardized scoring method, the degree to which certain indicators are relied upon is left to the discretion of each user. Likewise, no minimum number of indicators are required to be present before assessing a caller as innocent or guilty. Thus, investigators' prior experience and training may lead to wide variances in the interpretation and application of COPS Scale criteria.

Third, the COPS Scale does not account for how 911 operators might impact the behavior of callers. Operators' tone, demeanor, personality, prior experience, training, and types of questions asked may all influence caller attitudes and the extent of information provided (O'Donnell et al., 2023). 911 calls typically do not allow for long uninterrupted narratives by callers, as operators are attempting to quickly assess the situation, ascertain immediate needs, and determine the appropriate response. These factors, combined with the relatively short duration of most 911 calls and the stressful environment in which calls are often made, might impact the ability of callers to process and relay information. This could also result in responses that are contrary to COPS Scale expectations. For instance, previous research has found that truthtellers tend to be forthcoming in providing details, whereas deceivers attempt to keep their stories simple (Leal et al., 2023; Vrij et al., 2019, 2022). However, the time constraints and fragmented nature of emergency calls may inhibit truthful callers from providing as many details as they might in other settings, such as interviews. Future research should examine the nature of interactions between operators and callers, and whether alternative questioning tactics may assist in obtaining additional details.

Use of the COPS Scale may also influence the future opinions, interpretations of evidence, and subsequent actions of investigators. Our third objective examined how judgments of innocence or guilt might impact the types and nature of follow-up questions asked by investigators. Participants differed in multiple aspects depending on whether the COPS Scale led them to view the caller as innocent or guilty. Participants who assessed the caller to be guilty were more likely to focus questions on the caller's behaviors, intentions, or thoughts, rather than those of the victim or a third party or aspects of the scene. Question focus can impact an interview in a variety of ways, including the types of responses provided and specific topics discussed. Although questions related to caller activities may be necessary, an overconcentration on caller behavior to the detriment of other topics could result in an incomplete understanding of relevant details. Callers, regardless of whether they are actually innocent or guilty, may also exhibit a defensive posture if they perceive themselves to be targeted by the interviewer. These factors may lead to confirmation bias among interviewers, as they might interpret defensiveness or nervousness as signs that individuals are withholding or fabricating information.

Additional question characteristics were impacted as well. Callers judged to be guilty were more likely than callers judged as innocent to be asked questions about the incident. Conversely, participants were less likely to ask general background questions of callers judged to be guilty than those assessed as innocent. These findings have critical implications for investigations. Details tangential to the incident may be overlooked or minimized when callers are thought to be guilty. This can delay or prevent investigators from learning relevant contextual information and generating appropriate investigative leads. Likewise, questions primarily isolated to the incident may infer to callers that background details are not relevant, which may decrease the likelihood they will volunteer such information. This, in turn, may suggest to investigators that callers do not possess these details and may preclude them from sufficiently probing for more information. By contrast, callers perceived as innocent may not be asked adequate questions regarding the incident and valuable information might go undetected by investigators.

Interestingly, participants were more likely to ask closed-ended questions of callers judged to be innocent than those judged to be guilty and more likely to ask those judged to be guilty probing questions than those judged as innocent. Why this occurred is unknown, although a possible explanation is that participants may have inferred innocent callers did not possess as many details about the incident than guilty callers. This might have led to less intrusive questioning styles. Although probing questions (i.e., who, what, where, why,

when, and how) are generally more effective in gathering information than closed questioning styles, the belief that a person is guilty may influence the range of questions asked. This, again, has potential ramifications for investigations as neither TRCs nor FRCs generated many open-ended questions, which may be able to help mitigate detrimental effects of closed-ended or direct questions influenced by perceptions of innocence or guilt.

In totality, our findings suggest the COPS Scale should not be used to evaluate potential innocence or guilt among 911 callers. Despite its lack of effectiveness, though, the COPS Scale may still have investigative value to law enforcement, similar to SCAN. Although SCAN does not have empirical support for detecting deception, some researchers have acknowledged its potential limited use in assisting with future interviews by identifying portions of a statement requiring clarification or more detail (Vrij, 2008). Likewise, COPS Scale criteria may help investigators to elicit additional details during interviews by identifying unclear, confusing, or ambiguous information within 911 calls. For example, while Lack of Urgency may not be an indicator of guilt, seeking clarification as to why a caller exhibited this trait may help investigators better understand the caller's personality, relationship with the victim, current life circumstances, or state of mind at the time of the call. This information can be highly relevant to the investigation, regardless of whether the individual is telling the truth or lying. These details may help generate leads that investigators can further explore or provide insight into the thoughts, feelings, motivations, or behaviors of callers. Utilizing particular criterion from the COPS Scale to seek greater insight therefore appears more appropriate than using the COPS Scale to form opinions of guilt or innocence.

Limitations

Several limitations should be considered. First, calls were obtained primarily from law enforcement agencies or public media sources, which may not generalize to other calls that do not lead to media attention or a significant police response. The collected calls were also specific to those involving deaths or serious physical injuries resulting in death. Although this is consistent with the intended use of the COPS Scale, the results from this study may not generalize to other types of 911 calls. Another limitation is that the current study did not explore the potential impact different 911 operators might have on caller responses, nor did the study examine how these differences might subsequently affect veracity judgments of callers by users of the COPS Scale. It is possible the general demeanor, tone, experience, training, and types of questions asked by operators could influence both the information provided by callers and the manner in which this information is interpreted by COPS Scale users. Further, FRC and TRC call status was determined by the judicial and investigative outcome for each case. While consistent with prior 911 call research, caution should be exercised as ground truth could not be established with absolute certainty.

Methodological and practical limitations of the current study did not allow participants the opportunity to engage in actual interviews when evaluating questioning styles, but rather required participants to write down the primary questions they would ask if they were conducting an interview. Participants, who were all sworn law enforcement officers, would likely have asked a multitude of additional questions during actual interviews. These may have differed from their written questions and might have varied depending on interviewee responses. However, the written questions permitted the authors to evaluate the initial relevant questions considered by participants as these likely would have been asked in the early stages of interviews. This is important to explore, as initial questions and topics introduced by investigators may influence the overall direction of the interview.

Finally, all participants were trained on the COPS Scale using the book by Harpster and Adams (2017), the original study by Harpster et al. (2009), and information obtained by the research team through prior training and consultations with other trained law enforcement. It is possible this training was not sufficient for participants to gain proficiency in use of the COPS Scale. However, this appears unlikely as the book describes each indicator in detail, along with specific case examples demonstrating the application of the COPS Scale. Further, the back cover of the book states that it provides "… a practical method of analyzing 911 homicide calls to uncover the truth …" and "… lays out a framework to analyze the call to determine truth from fiction" (Harpster & Adams, 2017). Participants in the current study were also provided copies of each indicator's definition, to which they could refer at any time. If developers of the COPS Scale assert that specific training courses are needed to become proficient in its use, this should be examined in future studies.

Conclusion

The present study examined the ability of the COPS Scale indicators to assist law enforcement personnel in discriminating between innocent and guilty callers in cases involving fatalities. Our results question the effectiveness of the COPS Scale in distinguishing guilty from innocent callers, as participants mischaracterized approximately four out of 10 calls. Opinions of guilt based upon the COPS Scale also influenced the types of questions asked by investigators, as well as the manner in which the questions were posed. These factors have implications for law enforcement, as investigative decisions and prioritization of limited resources may be impacted by investigators' perceptions of whether 911 callers are involved in a crime or not. Although individual criterion within the COPS Scale may be useful for seeking additional insight into caller statements, the COPS Scale as a whole does not appear to be an effective instrument to judge innocence or guilt among callers.

References

Aamodt, M. G., & 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., Kohnken, 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

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

Avis, K. A., Stroebe, M., & Schut, H. (2021). Stages of grief portrayed on the internet: A systematic analysis and critical appraisal. *Frontiers in Psychology*, *12*, Article 772696. https://doi.org/10.3389/fpsyg.2021.772696

Beauregard, E., & Bouchard, M. (2010). Cleaning up your act: Forensic awareness as a detection avoidance strategy. *Journal of Criminal Justice*, *38*(6), 1160–1166. https://doi.org/10.1016/j.jcrimjus.2010.09.004

Bogaard, G., Meijer, E. H., Vrij, A., & Merckelbach, H. (2016). Scientific Content Analysis (SCAN) cannot distinguish between truthful and fabricated accounts of a negative event. *Frontiers in Psychology*, *7*, Article 243. https://doi.org/10.3389/fpsyg.2016.00243

Bond, C. F., & DePaulo, B. (2006). Accuracy of deception judgments. *Personality and Social Psychology Review*, *10*(3), 214–234. 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

Carpenter, B., Tait, G., Quadrelli, C., & Thompson, I. (2016). Investigating death: The emotional and cultural challenges for police. *Policing and Society*, *26*(6), 698–712. https://doi.org/10.1080/10439463.2015.1016024

Craun, S. W., Tanner, L., Clausen, V., Merola, M. A., Opanashuk, L., & Keel, T. G. (2022). Homicide or suicide: How nudity factors into this determination. *Homicide Studies*, *26*(3), 292–307. https://doi.org/10.1177/10887679211013071

Cromer, J., Brewster, J., Fogler, K., & Stoloff, M. (2019). 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. (1992). Nonverbal behavior and self-presentation. *Psychological Bulletin*, *111*(2), 203–243. https://doi.org/10.1037/0033-2909.111.2.203

DePaulo, B. M., Lindsay, J. J., Malone, B., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological Bulletin*, *129*(1), 74–118. https://doi.org/10.1037/0033-2909.129.1.74

DePaulo, B. M., & Morris, W. L. (2004). Discerning lies from truths: Behavioral cues to deception and the indirect pathway of intuition. In P. A. Granhag & L. Stromwoll (Eds.), *Deception detection in forensic contexts* (pp. 15–40). Cambridge University Press. https://doi.org/10.1017/CBO9780511490071.002

Droney, P. (2023). *Soros-funded left-wing media outlet launches attack on respected law enforcement professionals—gets humiliated*. Law Enforcement Today Page. Retrieved November 22, 2023, from https://lawenforcementtoday.com/critical-homicide-investigative-tool-overcomes-unfounded-left-wing-media-attack

Elms, S., & Keatley, D. (2023). The linguistic disposition of emergency callers: Using proximity coefficients to identify the sentiments of guilty and innocent emergency callers. *Journal of Police and Criminal Psychology*. Advance online publication. https://doi.org/10.1007/s11896-023-09601-6

Ferguson, C. (2015). Staged homicides: An examination of common features of faked burglaries, suicides, accidents and car accidents. *Journal of Police and Criminal Psychology*, *30*(3), 139–157. https://doi.org/10.1007/s11896-014-9154-1

Ferguson, C., & McKinley, A. (2019). Detection avoidance and mis/unclassified, unsolved homicides in Australia. *Journal of Criminal Psychology*, *10*(2), 113–122. https://doi.org/10.1108/JCP-09-2019-0030

Ferguson, C., & Petherick, W. (2016). Getting away with murder: An examination of detected homicides staged as suicides. *Homicide Studies*, *20*(1), 3–24. https://doi.org/10.1177/1088767914553099

Gamer, M., Lemon, J., Fellows, I., & Singh, P. (2019). *irr: Various coefficients of interrater reliability and agreement* (R package Version 0.84.1) [Computer software]. https://CRAN.R-project.org/package=irr

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

Geberth, V. J. (2010). Crime scene staging: An exploratory study of the frequency and characteristics of sexual posing in homicides. *Investigative Sciences Journal*, *2*(2), 1–19.

Grice, H. P. (1975). Logic and conversation. In P. Cole, J. L. Morgan (Eds.), *Speech acts* (pp. 41–58). Brill.

Harpster, T. (2024). *Home [LinkedIn]*. LinkedIn. Retrieved March 18, 2024, from https://www.linkedin.com/in/tracy-harpster-830ab12a

Harpster, T., & Adams, S. (2017). *Analyzing 911 homicide calls: Practical aspects and applications*. CRC Press. https://doi.org/10.1201/9781315386508

Harpster, T., Adams, S., & Jarvis, J. (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., Granhag, P., & Strömwall, L. (2007). Guilty and innocent suspects' strategies during police interrogations. *Psychology, Crime & Law*, *13*(2), 213–227. https://doi.org/10.1080/10683160600750264

Hartwig, M., Granhag, P., Strömwall, L., & Doering, N. (2010). Impression and information management: On the strategic self-regulation of innocent and guilty suspects. *The Open Criminology Journal*, *3*(2), 10–16. https://doi.org/10.2174/1874917801003020010

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

Hough, R. M. (2019). The investigation of homicide. *Homicide Studies*, *23*(2), 87–92. https://doi.org/10.1177/1088767919827348

Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological Review*, *88*(1), 67–85. https://doi.org/10.1037/0033-295X.88.1.67

Keel, T. G., Jarvis, J. P., & Muirhead, Y. E. (2009). An exploratory analysis of factors affecting homicide investigations: Examining the dynamics of murder clearance rates. *Homicide Studies*, *13*(1), 50–68. https://doi.org/10.1177/1088767908326903

Kübler-Ross, E., & Kessler, D. (2005). *On grief and grieving: Finding the meaning of grief through the five stages of loss*. Simon & Schuster.

Kübler-Ross, E., & Kessler, D. (2009). The five stages of grief. In Library of Congress Catalogin in Publication Data. (Ed.), *On grief and grieving* (pp. 7–30).

Lacks, R. D., Westveer, A. E., Dibble, A., & Clemente, J. (2008). Equivocal death investigation: Case study analyses. *Victims & Offenders*, *3*(2–3), 150–164. https://doi.org/10.1080/15564880801938292

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), Article 159. https://doi.org/10.2307/2529310

Leal, S., Vrij, A., Deeb, H., & Fisher, R. P. (2023). Interviewing to detect omission lies. *Applied Cognitive Psychology*, *37*(1), 26–41. https://doi.org/10.1002/acp.4020

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

Maier, B. G., Niehaus, S., Wachholz, S., & Volbert, R. (2018). The strategic meaning of CBCA criteria from the perspective of deceivers. *Frontiers in Psychology*, *9*, Article 855. https://doi.org/10.3389/fpsyg.2018.00855

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. 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/10.1177/09567976221077216

Markey, P. M., 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*(2), 519–527. https://doi.org/10.1007/s11896-022-09567-x

Masip, J. (2017). Deception detection: State of the art and future prospects. *Psicothema*, *29*(2), 149–159. https://doi.org/10.7334/psicothema2017.34

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

Miller, M. L., Merola, M. A., Opanashuk, L., Robins, C. J., Chancellor, A. S., & Craun, S. W. (2021). "911 what's your emergency?": Deception in 911 homicide and homicide staged as suicide calls. *Homicide Studies*, *25*(3), 256–272. https://doi.org/10.1177/1088767920948242

Nahari, G., Ashkenazi, T., Fisher, R., Granhag, P., Hershkowitz, I., Masip, J., Meijer, E., Nisin, Z., Sarid, N., Taylor, P., Verschuere, B., & Vrij, A.

(2019). "Language of lies": Urgent issues and prospects in verbal lie detection research. *Legal and Criminological Psychology*, *24*(1), 1–23. https://doi.org/10.1111/lcrp.12148

Nahari, G., Vrij, A., & Fisher, R. P. (2012). Does the truth come out in the writing? SCAN as a lie detection tool. *Law and Human Behavior*, *36*(1), 68–76. https://doi.org/10.1037/h0093965

O'Donnell, D., Burd, T., & Huffman, M. (2024, August 9). *Evaluation of the 911 COPS scale as a deception detection method for 911 calls*. https://osf.io/3gkpd

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

Olsson, J. (2004). *Forensic linguistics: An introduction to language, crime and the law*. Continuum International Publishing Group.

Pizarro, J. M., Terrill, W., & LoFaso, C. A. (2020). The impact of investigation strategies and tactics on homicide clearance. *Homicide Studies*, *24*(1), 3–24. https://doi.org/10.1177/1088767918816741

Proutskova, P., & Gruszczynski, M. (2020). *kripp.boot: Bootstrap Krippendorff's alpha intercoder reliability statistic* (R package Version 1.0.0) [Computer software]. https://github.com/MikeGruz/kripp.boot

Puranik, D., Paul, P., & Krishnamurthy, R. (2015). Application of equivocal death psychological autopsy for investigation: A case study. *International Journal of Indian Psychology*, *2*(2), 31–36. https://doi.org/10.25215/0202.063

R Core Team. (2023). *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org

Saxena, G., & Saini, V. (2017). Psychological autopsy: A way to revealing the enigma of equivocal death. *International Journal of Forensic Sciences*, *2*(2), Article 16000123. https://doi.org/10.23880/ijfsc-16000123

Schlesinger, L. B., Gardenier, A., Jarvis, J., & Sheehan-Cook, J. (2014). Crime scene staging in homicide. *Journal of Police and Criminal Psychology*, *29*(1), 44–51. https://doi.org/10.1007/s11896-012-9114-6

Schneidman, E. S. (1981). The psychological autopsy. *Suicide and Life-Threatening Behavior*, *11*(4), 325–340. https://doi.org/10.1111/j.1943-278X.1981.tb01009.x

Schroeder, D. A., & White, M. D. (2009). Exploring the use of DNA evidence in homicide investigations: Implications for detective work and case clearance. *Police Quarterly*, *12*(3), 319–342. https://doi.org/10.1177/1098611109339894

Sporer, S. L. (1997). The less travelled road to truth: Verbal cues in deception detection in accounts of fabricated and self-experienced events. *Applied Cognitive Psychology*, *11*(5), 373–397. https://doi.org/10.1002/(SICI)1099-0720(199710)11:5<373::AID-ACP461>3.0.CO;2-0

Steller, M., & Köhnken, G. (1989). Criteria-based content analysis. In J. Doris (Ed.), *The suggestibility of children's recollections* (pp. 217–-245). American Psychological Association. https://doi.org/10.1037/t27704-000

Stroebe, M., Schut, H., & Boerner, K. (2017). Cautioning health-care professionals. *OMEGA Journal of Death and Dying*, *74*(4), 455–473. https://doi.org/10.1177/0030222817691870

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. https://doi.org/10.1017/CBO9780511490071.010

Strömwall, L. A., Hartwig, M., & Granhag, P. A. (2006). To act truthfully: Nonverbal behaviour and strategies during a police interrogation. *Psychology, Crime & Law*, *12*(2), 207–219. https://doi.org/10.1080/10683160512331331328

Visentin, S., Massaro, L., Viel, G., Cecchetto, G., & Montisci, M. (2019). Suicide identification during on-site inspection. Proposal and application of an interpretative method for death scene investigation. *Forensic Science International*, *297*, 148–155. https://doi.org/10.1016/j.forsciint.2019.01.029

Volbert, R., & Steller, M. (2014). Is this testimony truthful, fabricated, or based on false memory? Credibility assessment 25 years after Steller and Köhnken (1989). *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*. John Wiley & Sons.

Vrij, A., Edward, K., Roberts, K. P., & Bull, R. (2000). Detecting deceit via analysis of verbal and nonverbal behavior. *Journal of Nonverbal Behavior*, *24*(4), 239–263. https://doi.org/10.1023/A:1006610329284

Vrij, A., Fisher, R., & Leal, S. (2023). How researchers can make verbal lie detection more attractive for practitioners. *Psychiatry, Psychology and Law*, *30*(3), 383–396. https://doi.org/10.1080/13218719.2022.2035842

Vrij, A., Granhag, P. A., Ashkenazi, T., Ganis, G., Leal, S., & Fisher, R. P. (2022). Verbal lie detection: Its past, present and future. *Brain Sciences*, *12*(12), Article 1644. https://doi.org/10.3390/brainsci12121644

Vrij, A., & Leal, S. (2020). Proportion of complications in interpreter-absent and interpreter-present interviews. *Psychiatry, Psychology and Law*, *27*(1), 155–164. https://doi.org/10.1080/13218719.2019.1705197

Vrij, A., Leal, S., Fisher, R., Mann, S., Deeb, H., Jo, E., Campos, C., & Hamzeh, S. (2019). The efficacy of using countermeasures in a Model Statement interview. *The European Journal of Psychology Applied to Legal Context*, *12*(1), 23–34. https://doi.org/10.5093/ejpalc2020a3

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., Leal, S., Mann, S., Fisher, R., Dalton, G., Jo, E., Shaboltas, A., Khaleeva, M., Granskaya, J., & Houston, K. (2018). Using unexpected questions to elicit information and cues to deceit in interpreter-based interviews. *Applied Cognitive Psychology*, *32*(1), 94–104. https://doi.org/10.1002/acp.3382

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*(6), 741–755. https://doi.org/10.1002/acp.1200

(*Appendix follows*)

Appendix

COPS Scale Indicators

Guilty indicators	Innocent indicators
Forced choice indicators	
No Immediate Plea for Help for Victim	Immediate Plea for Help for Victim
No Immediate Assessment of Victim	Immediate Assessment of Victim
Focus on Caller	Focus on Victim
Acceptance of Victim's Death	No Acceptance of Death
No Aid Provided to Victim	Aid Provided to Victim
Extraneous Information	Relevant Information
No Immediate Sensory Details	Immediate Sensory Details
Inappropriate Order	Prioritized Order
Blood/Brains Comments	Bleeding Comments
Inappropriate Delivery	Expected Delivery
Lack of Urgency	Urgency
No Fear for Caller's Safety	Fear for Caller's Safety
No Proximity to Victim	Proximity to Victim
Initial Delays	Initial Sounds or Comments
Failure to Address/Coach Victim	Address/Coach Victim
Dispatcher Confusion	Dispatcher Comprehension
Inappropriate Delivery	Relevant Information
Inappropriate Order	Urgency
Standalone/no forced choice indicators	
Defendant Mentality	Aggressive Demands
Ingratiating Remarks	Free Flow of Information
Mental Miscues	
Minimizes	
Diversion	
Equivocation	
Evasion	
Hangs Up	
Only Answers What's Asked	
Pauses	
Repetition	
Self-interruption	
Short Answers	
Unintelligible Comments	
Attempts to Convince	
Awkward Phrases	
Conflicting Facts	
"Huh?" Factor	
"I Don't Know"	
Insult/Blame Victim	
Isolated "Please"	
Lack of Contractions	
No Modulation	
Unexplained Knowledge	

Received April 14, 2024
Revision received July 29, 2024
Accepted July 31, 2024 ■