Article

The (Mis)utilization of Cues During Deception Detection in 911 Homicide Calls

Homicide Studies
1–20

DOI: 10.1177/10887679221129850
journals.sagepub.com/home/hsx
SAGE

Patrick Markey[1], Erika Feeney[1], Brooke Berry[1], Alec Martin[1], and Erica Slotter[1]

Abstract

The current study explored cue utilization during 911 homicide calls to better understand deception detection in this high-risk situation. A sample of 93 participants judged the deceptiveness of a random subset of 110 homicide calls placed to 911. A separate group of raters coded 86 different cues expressed by each 911 caller. Results indicated that judges were unable to detect deception accurately. While judges appeared to utilize some cues correctly when assessing callers' deception, subsequent analyses found that judges likely had difficulty detecting deception because they failed to utilize emotionally related cues correctly.

Keywords

judgment, interpersonal interaction, 911, social behavior, perception

Introduction

On a seemingly ordinary Sunday night on August 20, 1989, a 911 call came into the Beverly Hills Emergency Communications Center around midnight. Uncontrollable sobs could be heard from the caller, and then the chilling words, "Somebody killed my parents." The following two and a half minutes were filled with the caller's emotional pleas, unanswered questions, and unintelligible comments. The emotionally charged cues expressed by the caller would ultimately be scrutinized by law enforcement, the media, and a jury to help determine whether the caller was guilty of the crime they

[1]Villanova University, PA, USA

Corresponding Author:
Patrick Markey, Villanova University, 800 Lancaster Avenue, Villanova, PA 19085-1603, USA.
Email: patrick.markey@villanova.edu

called to report (cf., Thornton, 1995). Eventually, it would be discovered that the caller, Lyle Menendez murdered his parents with the help of his brother Eric Menendez.

The public and legal interest concerning the meaning of the cues exhibited by Lyle Menendez during the 911 call reporting the murder of his parents illustrates the need for a better understanding of whether individuals can accurately judge the deception of 911 homicide callers, and what specific cues might aid or hinder this process. The goal of the current research was to examine these issues. We aimed to (a) establish whether judges are accurate in determining when a caller reporting a homicide to 911 is being deceptive, (b) examine which cues these judges use to make their determinations, and (c) investigate whether the cues that judges use are, in fact, cues that actually indicate deception on the part of the caller.

Theoretical Background

Whether someone calling 911 to report a homicide is being deceptive is not directly observable to a judge. Because it is unlikely there are perfect indicators of deception in this context, judges need to combine information from various imperfect cues to accurately estimate a caller's deception (Brunswik, 1956). Egon Brunswik's Lens Model helps explain this process by examining how the environment links reality to perceptual cues and judgments (see Figure 1). A distinguishing feature of this model is that the target (e.g., a 911 caller) should be studied in parallel to the judge. The model comprises four elements: a target; cues emitted by the target; cues that the judge uses; and the judgment made.

According to this model, cues can serve as a kind of lens through which judges indirectly perceive the deception or honesty of a target—someone calling 911 to report a homicide in the case of the current research (Hursch et al., 1964). As seen on the left side of the lens, *cue validity* pertains to how strongly cues are characterized by their link to callers' actual deception and can be represented by the correlation between a given cue and the deception of 911 callers. For example, deceptive callers might behave self-defensively and extremely emotional but not necessarily display long pauses when talking. On the right side of the model, *cue utilization* characterizes how strongly cues relate to judges' assessment of deception and can be represented by the correlation between a given cue and judges' rating of 911 callers' deception. For instance, judges' assessments of callers' deception might be associated with callers' behaving self-defensively and taking long pauses when talking but not with the caller appearing overly emotional.

Errors in judgment will likely occur when judges hold incorrect explicit or implicit beliefs about which cues are related to deception or fail to consider valid cues. For example, as seen in Figure 1, judges might generally believe that deceptive individuals often take long pauses when talking to another person. However, most empirical studies find that response latency is not a valid cue of deception (cf. DePaulo et al., 2003). Similarly, in our example, judges might fail to utilize the valid cue of extreme emotions when judging deception. In either case, accuracy will be relatively poor when

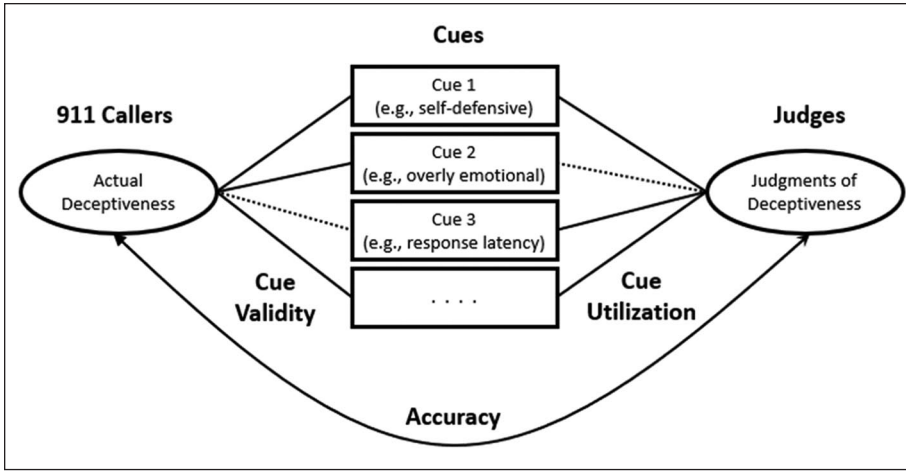

Figure 1. Brunswik's (1956) lens model.

judges' utilization of cues fails to match the validity of the cues (e.g., Gosling et al., 2002; Marcus et al., 2006).

Literature Review

Numerous studies have examined various cues that judges utilize to assess deception. Often these studies are conducted by first coding a multitude of cues displayed by deceptive or honest individuals during a video or audio interaction. Later, judges assess the deceptiveness or honesty of these individuals. These judgments are then correlated with the various cues to determine which cues are related to deception judgments (cf., Bond et al., 1985; Frank & Ekman, 2004). This methodology does not assume that judges are aware of the cues they utilize when deciding an individual's deceptiveness. Given the rapid and implicit nature of judgment and deception detection, it is likely that such explicit beliefs concerning deception cues have little impact on the actual judging of deception (Hartwig & Bond, 2011). Instead, this methodology measures the strength and direction of the relation between a given cue and judges' perceptions of deceptiveness.

One large meta-analysis examining cue utilization during deception judgments found that individuals tend to be judged deceptive if they appear incompetent, uncertain, and indifferent. In contrast, individuals are judged as honest when they are cooperative, involved, and provide statements that seem plausible and spontaneous (Hartwig & Bond, 2011). However, these findings are based primarily on research examining fairly mundane deception paradigms where individuals were asked to make either honest or deceptive claims in laboratory settings. For example, past studies asked individuals to lie about their feelings concerning controversial topics

(e.g., capital punishment), deceive judges as to the content of a photograph, mislead others about their academic accomplishments, or even lie about the suit of a playing card (e.g., Cheng & Broadhurst, 2005; Heinrich & Borkenau, 1998; Riggio & Friedman, 1983). While such research helps provide insight into cue utilization during everyday interactions, no research has studied cue utilization within the emotionally charged context of 911 homicide calls. Therefore, it is unknown if the cues utilized to judge deception during routine interactions in a laboratory will generalize to real-world situations with life-or-death outcomes.

Although cue utilization by judges in the perception of deception has not been studied in the context of 911 homicide calls, several studies have attempted to examine which cues are valid indicators of actual deception during 911 calls. Harpster (2006) and Harpster et al. (2009) carried out one of the earliest systematic attempts to examine cues related to caller deception during 911 homicide calls. Harpster et al. (2009) report the same data and results as those reported in Harpster (2006) and are discussed here as a single study. In this study, 20 dichotomous cues were related to the deception or honesty of 911 homicide callers. Deceptive callers tended to be uncooperative, blamed the victim for the situation, and gave conflicting facts to the 911 operator. On the other hand, honest callers were more likely to make demanding and urgent pleas while sounding emotional. Findings from this study resulted in the creation of the *Considering Offender Probability Statements (COPS)* scale, which was designed as a tool for law enforcement to use to help gain insight into the potential deceptiveness of 911 callers (Harpster & Adams, 2017). However, subsequent attempts at replicating the findings from Harpster and colleagues have consistently failed to reproduce most of their results (Cromer et al., 2019; M. L. Miller et al., 2021).

The findings from the Harpster (2006) and Harpster et al. (2009) likely failed to replicate due to the study's limitations. Only a single judge, the primary author, seemed to have coded the 911 calls. Because this judge also "personally contacted" each lead detective to obtain the 911 calls, it is unclear whether the judge was blind to whether the callers were deceptive (Harpster, 2006). The possibility of non-masked coding might explain why some of the effect sizes from the Harpster study seemed unreasonably large. For example, the single dichotomous cue "extraneous information" was correlated with deception $r = .81$ and accurately predicted deception 91% of the time. If this effect size were accurate, it would be the largest effect size found examining deceptive behaviors and among the largest effect sizes ever discovered in the social sciences (DePaulo et al., 2003; Richard et al., 2003).

In order to overcome the limitations of the Harpster study, Markey et al. (2022) recently examined a sample of 911 homicide calls to gain insight into which cues are valid indicators of deception. This study employed a group of judges blind to whether the caller was being deceptive to code 86 different cues. Results indicated that deceptive callers were self-dramatizing, moody, worried, and related narratives that lacked structure, clarity, and focus. Whereas honest callers were forthright, candid, focused, cooperative, and provided a coherent sequence of events. Unfortunately, while such findings help provide insight into which cues are valid

indicators of deception during 911 calls, no research has examined whether or not judges accurately utilize such cues when determining whether a caller is being deceptive.

Judgment Accuracy and the Current Study

It is difficult to detect deception in others accurately. Even specially trained law enforcement officers have trouble recognizing when deception occurs (DePaulo & Pfeifer, 1986; Köhnken, 1987; Vrij, 1993). In one large-scale analysis of 206 papers, the average accuracy for detecting deception in various situations was 54% (with 50% being chance; Bond & DePaulo, 2006). According to this review, humans are only slightly better than chance at classifying honest and deceptive statements. Two major explanations tend to be proposed for why deception detection is prone to errors: the *weak objective cue hypothesis* and the *wrong subjective cue hypothesis* (Hartwig & Bond, 2011). *The weak objective cue hypothesis* proposes that accurate deception judgments are difficult because there are minimal behavioral differences between deceptive and honest individuals. Whereas *the wrong subjective cue hypothesis* is the notion that deception detection is challenging because individuals often hold incorrect explicit and implicit beliefs about which cues are related to deception and employ these erroneous beliefs in their decision making (e.g., Akehurst et al., 1996; DePaulo et al., 2003; Global Deception Research Team, 2006). In other words, and as anticipated by Brunswik's Lens model, accurate deception detection might be problematic because (1) valid cues are only weakly related to deception and/or (2) there is a disconnect between valid cues and how judges utilize cues.

While no research has examined judges' ability to detect deception accurately during 911 homicide, there are reasons to suspect that accurate judgments will be more likely to occur during high-risk interpersonal interactions, such as 911 homicide calls, than during more routine interactions often studied by past researchers (Markey et al., 2022). First, cues related to deception may become more evident when people are lying about serious transgressions (e.g., committing a homicide) than mundane issues (G. R. Miller & Stiff, 1993). Second, high-risk interactions are typically accompanied by strong emotions which are difficult to fake (Porter & ten Brinke, 2010). Third, 911 calls are usually placed soon after a crime, giving deceptive individuals little time to create and rehearse a false narrative (Harpster et al., 2009). Finally, a person who committed a homicide is unlikely to know how an innocent person would typically behave when calling 911. Taken together, it is likely that the high-stakes situation deceivers face while placing a 911 homicide call will result in considerable leakage of valid cues related to deception. Nevertheless, it is unclear whether judges can successfully utilize these cues to assess callers' deception accurately.

The current study's overarching goal was to investigate cue utilization during 911 homicide calls to help understand deception detection during this high-stakes situation. Using a sample of 911 homicide calls collected previously by Markey et al. (2022), participants judged the deceptiveness of 911 callers. A separate set of raters,

blind to the deception of the callers, coded the degree to which 86 various cues were expressed during each 911 call. Five research questions (RQ) were examined using these data.

RQ1: Are judges able to accurately rate the deceptiveness of 911 homicide callers?
RQ2: Do judges agree with each other in their deceptiveness ratings of 911 homicide callers?
RQ3: What cues do judges utilize when rating callers' deception?
RQ4: What cues are valid indicators of callers' deception?
RQ5: Are the cues utilized by judges to evaluate deception similar to the cues that were valid indicators of deception?

Method

Data Sources and Judges

911 calls. Consistent with prior 911 research (Cromer et al., 2019; M. L. Miller et al., 2021), calls were obtained from publicly available open-source data, such as news sources, police department releases, and various archives. Calls placed to 911 were eligible for the study using similar criteria employed by past researchers (cf., Burns & Moffitt, 2014; Cromer et al., 2019). (1) The call involved the killing of another person. (2) Emergency services were notified. (3) The caller was aware of and able to communicate the general nature of the emergency. (4) At least two news sources could verify prosecution, admission of guilt, or another outcome ultimately resulting from the call. (5) The call was less than 7 minutes in length. (6) The caller did not confess to wrongdoing. Callers claiming extenuating circumstances (e.g., self-defense, accident, etc.) that led to the death of another person were included.

Based on a priori power, a sample size of 110 (55 deceptive and 55 honest) for the non-directional predictions concerning cue utilization and validity was necessary to achieve 90% power for a moderate effect size ($r_{pb} = .30$). The majority of 911 calls used in the current study were taken from the data set previously examined by Markey et al. (2022) investigating cue validity. This previous research did not examine cue utilization, the main focus of the current research, and examined cue validity using a 25% smaller and less powerful sample than the current study.

Judges. Ninety-three undergraduate students were recruited to serve as judges for course credit. Of the judges in the current study, 75% were female, and 25% were male. The majority of the participants were White (74%), and the remaining were Asian or Pacific Islander (11%), Hispanic (10%), and Black (5%). The mean age of the participants was 19.18 years ($SD = 2.07$), and the median age was 19. Participation in the study, including the informed consent procedure, was conducted according to institutional review board stipulations and with the institutional review board's approval.

Procedure

Judges' assessments of 911 callers' deception. Judges were informed that researchers were examining how accurately people can detect deception during 911 homicide calls. Each judge was then asked to listen to a random sample of 911 homicide calls. After listening to each call, judges were instructed to indicate whether they thought the caller was deceptive or honest during the call. Depending on call length, each judge rated between four to five calls (*M* rating per judge = 4.78), and each 911 call was rated by three to five judges (*M* ratings per call = 4.05), resulting in 445 ratings of deception or honesty.

Determination of 911 callers' deception. At least two external sources (usually media reports detailing the crime's legal proceedings) were used to determine callers' deception or honesty. To be coded as "deceptive," the caller was required to have been found guilty in a court of law. In cases when an indictment was not possible, such as the death of the caller, expert opinions were employed (e.g., opinion of the medical examiner, police investigators, grand jury findings, etc.). The remaining callers not deemed guilty were coded as "honest." This coding method is consistent with past 911 research (Cromer et al., 2019; M. L. Miller et al., 2021; Markey et al., 2022).

Coding of cues during 911 calls. The 911 Q sort was designed to code audio recordings of 911 homicide calls (Markey et al., 2022). Cues in the 911 Q sort were created to capture audio cues at a psychologically meaningful level, requiring as little subjective interpretation from the coders as possible. First, a set of relevant cues was generated from past 911 studies (e.g., Harpster et al., 2009), from research examining deception cues in everyday interactions (e.g., DePaulo et al., 2003), and from the Riverside Behavioral Q-Sort (an assessment designed to examine various cues during dyadic interactions; Funder et al., 2000). Next, a group of researchers eliminated overly redundant cues and reworded cues to be relevant in the context of 911 calls. The final set of 86 cues ranged from items directly relevant to 911 calls (e.g., "Caller makes the dispatcher confused," "Caller quickly asks for help for the victim") to cues that assessed a caller's general behavior (e.g., "Caller acts in a reckless manner," "Caller is talkative").

The 911Q was employed using a Q-sort methodology (Ozer, 1993). As discussed in Markey et al. (2022), three raters, blind to the deception of the caller, independently coded the audio of all 911 calls after listening to each call (see the Supplemental File for a copy of the instructions associated with coding procedures). Raters were graduate research assistants whose training consisted of reviewing the 86,911Q cues. Before coding the data set, all raters practiced Q-sorting five 911 calls (not included in the final analysis), and discrepancies and concerns among the coders were discussed until resolved. Raters used a modified version of the online software HtmlQ (2019) to sort the 86,911Q cues into nine categories (1 = extremely uncharacteristic; 9 = extremely characteristic) with the assigned distribution, respectively, 3, 6, 10, 15, 18, 15, 10, 6, and 3. While time-consuming, Q-sorts are less susceptible to biases like extremity

bias, midpoint responding, acquaintance bias, and halo effects than other more traditional assessments (Ozer, 1993; Serfass & Sherman, 2013). The cue ratings used for the current study were coded previously by Markey et al. (2022), and the median cue reliability was .62, which is similar to the reliability of *single* items in other behavioral Q-sorts (cf., Dunkel et al., 2015; Sherman et al., 2013).

Results

RQ1: Are judges able to accurately rate the deceptiveness of 911 homicide callers?
Judges accurately assessed a caller's dishonesty 53% of the time. Because ratings were nested within judges, judge accuracy was tested for significance using hierarchical binary logistic regression where judges' ratings of deception were predicted from actual deception. As expected from the low accuracy percentage, this analysis found no significant relationship, Wald's χ^2 (1) = 2.13, $p = .145$, between judgment ratings of deception and actual deception of 911 callers.

RQ2: Do judges agree with each other in their deceptiveness ratings of 911 homicide callers?
Judges agreed with each other whether or not a caller was being deceptive 76% of the time. Cohen's kappa (k) was calculated according to the Fleiss-Cuzick extension (Fleiss & Cuzick, 1979), which allows the estimation of agreement when different observations are rated by different judges and the number of judges per observation varies. Results indicated that judges obtained a significant and moderate (Landis & Koch, 1977) overall level of judge agreement, $k = .45$, $p < .001$. Given the moderate level of judge agreement, each call was dichotomously coded as being judged "deceptive" or "honest" when the majority of judges rated the call in that manner. Two calls were dropped from the following analyses because judges were evenly split in their assessments of the caller's deceptiveness.

RQ3: What cues do judges utilize when rating callers' deception?
Drawing on the logic of Brunswik's (1956) lens model, a series of analyses were conducted to examine the extent to which cues expressed by 911 callers were related to judges' assessments of deception. Specifically, point-biserial correlations were computed between judges' deception rating (coded 1 = caller perceived as deceptive and 0 = caller perceived as honest) and each of the 86,911Q cues. Therefore, positive cue utilization correlations indicate the cues callers tended to exhibit when judges viewed them as deceptive and negative correlations denote the cues callers expressed when judges viewed them as honest. Table 1 displays the resulting pattern of significant cue utilization correlations (see Supplemental Table for results of all 86,911 Q-sort cues).

The previous set of 86 analyses found 29 (33%) significant effects linking 911Q cues to deception judgments. However, given the number of non-independent analyses conducted, some of these effects might be significant simply due to chance. Sherman and Funder's (2009) randomization method was therefore employed to determine the

Table 1. Cue Utilization: Significant Correlations Between 911Q Cues and Judges' Ratings of Deception During 911 Homicide Calls.

Cue	r_{pb}	95% CI [LL, UL]	*p*-Value
Positive correlations (cues utilized for deception judgments)			
61. Caller is detached from the situation	.35	[0.17, 0.50]	<.001
81. Caller is emotionally bland; has flattened affect.	.34	[0.16, 0.50]	<.001
33. Caller makes attempts to convince dispatcher of innocence	.33	[0.15, 0.49]	<.001
65. Caller is self-defensive	.32	[0.14, 0.48]	.001
48. Caller is awkward.	.31	[0.13, 0.47]	.001
34. Caller gives conflicting facts/information or contradicts themselves	.28	[0.10, 0.44]	.003
20. Caller expresses worry that they might be blamed	.26	[0.08, 0.43]	.006
28. Caller takes long pauses	.26	[0.07, 0.43]	.007
26. Caller uses equivocation (i.e., caller is vague)	.26	[0.07, 0.43]	.007
70. Caller is relaxed	.26	[0.07, 0.42]	.008
23. Caller minimizes own involvement in event	.24	[0.05, 0.41]	.013
21. Caller expresses ingratiating remarks to gain favor	.23	[0.04, 0.40]	.017
27. Caller is evasive	.23	[0.04, 0.40]	.018
30. Caller's answers are overly brief	.21	[0.02, 0.38]	.032
54. Caller seems to be purposely trying to undermine the victim getting help	.20	[0.01, 0.38]	.035
36. Caller claims not to know something but later indicates he/she does know	.20	[0.01, 0.38]	.037
Negative correlations (cues utilized for honesty judgments)			
38. Caller sounds emotional	−.32	[−0.48, −0.14]	.001
2. Caller's message is plausible and believable	−.32	[−0.48, −0.14]	.001
60. Caller speaks in a loud voice	−.30	[−0.46, −0.12]	.002
63. Caller is working hard and is focused on the event	−.30	[−0.46, −0.11]	.002
52. Caller express sympathy (to anyone)	−.29	[−0.45, −0.10]	.003
67. Caller behaves and acts quickly	−.28	[−0.44, −0.09]	.004
66. Caller seeks reassurance from the dispatcher	−.25	[−0.42, −0.07]	.009
40. Caller apologizes to the victim	−.24	[−0.41, −0.05]	.013
37. Caller uses isolated "please" (i.e., the word "please" is said alone)	−.23	[−0.40, −0.04]	.018
6. Caller repeats words or phrases with little or no intervening pauses	−.23	[−0.40, −0.04]	.019
19. Caller consistently and forcefully demands help for the victim	−.22	[−0.40, −0.04]	.020
74. Caller is forthright and candid	−.20	[−0.37, −0.01]	.042
85. Caller is overwhelmed by the situation	−.19	[−0.37, −0.01]	.044

Note. $n = 108$.

Table 2. Randomization Tests of 911Q Cues Correlates to Judges' Deception Ratings (Cue Utilization) and Actual Deception (Cue Validity).

	Number significant ($p < .05$)	Mean absolute r_{pb}
Cue utilization		
Observed value	29	.15
Expected by chance	4.24	.08
SE	3.88	.01
p-Value	<.001	<.001
Cue validity		
Observed value	50	.20
Expected by chance	4.35	.08
SE	3.94	.01
p-Value	<.001	<.001

probability of obtaining 29 significant results under a random model of no association between deception judgments and the 911Q. This was done by: (1) randomly redistributing deceptive vs. honest judgments and (2) computing the number of significant ($p < .05$) correlations between judgments of deceptiveness and the 911Q cues in this new sample. This procedure was repeated 10,000 times. The resulting values were used to form a sampling distribution indicating the number of expected significant effects under the null of no relation between deception judgments and the 911Q. This analysis found that the probability of the current study finding 29 significant ($p < .05$) correlations linking judgments of deception to the 911Q by simple chance was $p < .001$ (see Table 2).

Randomization tests were again used to estimate the significance of the mean effect size found linking deception judgments to the entire set of 911Q cues (Sherman & Funder, 2009). The absolute correlation value between judgments of deceptiveness and the entire set of 86,911Q cues was computed first (average $r_{pb} = .15$). Like the previous analysis, 10,000 random samples were used to estimate the probability of obtaining this observed effect size under a random model of no association between deception judgments and the 911Q. This analysis indicated that the chance of the current study finding this observed effect size linking deception judgments to the 911Q was $p < .001$ (see Table 2).

RQ4: What cues are valid indicators of callers' deception?

Analyses were next conducted to examine the extent to which cues expressed by 911 callers were related to callers' actual deception. Point-biserial correlations were computed between callers' deception (coded 1 = caller was deceptive and 0 = caller was honest) and each of the 86,911Q cues. Therefore, positive cue validity correlations indicate the cues deceptive callers tended to exhibit, and negative correlations denote the cues honest callers expressed. Table 3 displays the resulting pattern of significant cue validity correlations (see Supplemental Table for results of all 86,911 Q-sort cues).

Table 3. Cue Validity: Significant Correlations Between 911Q Cues and Callers' Actual Deception During 911 Homicide Calls.

Cue	r_{pb}	95% CI [LL, UL]	*p*-Value
Positive correlations (valid cues of deception)			
10. Caller acts in a reckless manner.	.47	[0.31, 0.61]	<.001
83. Caller is self-dramatizing. Theatrical; exaggerates emotions.	.44	[0.28, 0.58]	<.001
72. Caller's behavior is unpredictable	.42	[0.26, 0.57]	<.001
24. Caller uses many non sequiturs	.42	[0.25, 0.56]	<.001
27. Caller is evasive	.36	[0.19, 0.52]	<.001
85. Caller is overwhelmed by the situation	.36	[0.18, 0.51]	<.001
32. Caller makes lots of unintelligible comments	.34	[0.17, 0.50]	<.001
75. Caller's mood fluctuates frequently	.33	[0.15, 0.49]	<.001
25. Caller makes the dispatcher confused	.32	[0.14, 0.48]	.001
49. Caller talks at rather than with the operator	.32	[0.13, 0.48]	.001
29. Caller self-interrupts	.32	[0.13, 0.48]	.001
48. Caller is awkward.	.31	[0.12, 0.47]	.001
6. Caller repeats words or phrases with little or no intervening pauses	.30	[0.12, 0.46]	.002
38. Caller sounds emotional	.29	[0.11, 0.46]	.002
78. Caller interprets simple situations in unnecessarily complicated ways	.28	[0.09, 0.44]	.004
65. Caller is self-defensive	.27	[0.08, 0.43]	.005
54. Caller seems to be purposely trying to undermine the victim getting help	.27	[0.08, 0.43]	.005
77. Caller sounds depressed.	.26	[0.08, 0.43]	.006
58. Caller expresses self-pity or feelings of victimization	.26	[0.07, 0.42]	.007
5. Caller is uncertain and insecure	.25	[0.06, 0.42]	.010
9. Caller is nervous (e.g., voice is tense, anxious, etc.)	.24	[0.06, 0.41]	.011
20. Caller expresses worry that they might be blamed	.24	[0.06, 0.41]	.012
37. Caller uses isolated "please" (i.e., the word "please" is said alone)	.24	[0.05, 0.41]	.014
34. Caller gives conflicting facts/information or contradicts themselves	.23	[0.04, 0.40]	.018
33. Caller makes attempts to convince dispatcher of innocence	.20	[0.01, 0.38]	.035
22. Caller blames the victim for the situation	.20	[0.01, 0.37]	.039
30. Caller's answers are overly brief	.20	[0.01, 0.37]	.043

(continued)

Table 3. (continued)

Cue	r_{pb}	95% CI [LL, UL]	*p*-Value
Negative correlations (valid cues of honesty)			
4. Caller responds in ways that are direct, relevant, and clear	−.49	[−0.62, −0.33]	<.001
76. Caller focuses on the important issues and provides relevant details	−.47	[−0.60, −0.30]	<.001
7. Caller is cooperative and helpful	−.44	[−0.58, −0.27]	<.001
63. Caller is working hard and is focused on the event	−.44	[−0.58, −0.27]	<.001
74. Caller is forthright and candid	−.44	[−0.58, −0.27]	<.001
11. Caller corrects any errors they made in the call	−.42	[−0.56, −0.25]	<.001
3. Caller's message is consistent and provides a coherent sequence of events	−.41	[−0.55, −0.23]	<.001
64. Caller appears to have a high degree of intellectual capacity	−.37	[−0.52, −0.20]	<.001
2. Caller's message is plausible and believable	−.36	[−0.52, −0.19]	<.001
80. Caller offers or gives advice to dispatcher	−.33	[−0.49, −0.15]	<.001
71. Caller is moralistic	−.33	[−0.49, −0.15]	<.001
82. Caller is verbally fluent.	−.33	[−0.49, −0.15]	.001
86. Caller expresses distrust	−.31	[−0.47, −0.13]	.001
40. Caller apologizes to the victim	−.28	[−0.44, −0.09]	.004
12. Caller admits to not knowing information	−.27	[−0.43, −0.08]	.005
67. Caller behaves and acts quickly	−.26	[−0.43, −0.08]	.006
73. Caller is assertive	−.26	[−0.43, −0.07]	.007
15. Caller accepts the death of the victim	−.25	[−0.42, −0.07]	.008
70. Caller is relaxed	−.25	[−0.42, −0.06]	.010
81. Caller is emotionally bland; has flattened affect.	−.25	[−0.41, −0.06]	.011
57. Caller blames others (for anything)	−.24	[−0.41, −0.06]	.012
41. Caller asks permission to complete tasks given by the dispatcher	−.24	[−0.41, −0.05]	.012
50. Caller expresses agreement unusually often	−.19	[−0.37, 0.01]	.047

Note. $n = 108$.

The set of 86 analyses examining cue validity produced 50 (58%) significant effects linking cues to actual deception during 911 homicide calls. As with the cue utilization, randomization tests were used to examine the probability of obtaining this number of significant effects, and the probability of obtaining the absolute effect size yielded in this analysis (average $r_{pb} = .21$) due to chance. As seen in Table 2, the probability of this set of analyses finding 50 significant ($p < .05$) correlations linking actual deception to the 86 cues was $p < .001$. Additionally, the chance of this set of analyses finding an average observed effect size of $r_{pb} = .21$ was $p < .001$.

RQ5: Are the cues utilized by judges to evaluate deception similar to the cues that were valid indicators of deception?

Consistent with Brunswik's lens model, accurate judgments of callers' deception are expected when judges utilize valid cues and disregard invalid cues. Therefore, the pattern of cue utilization correlations was compared to the pattern of cue validity correlations. All effects were first transformed using Fisher's *r*-to-*z* formula, and then these transformed values were correlated across all 86 cues. The resulting vector correlation reflects the extent to which judges are sensitive to the validity of cues and will be highest when judges utilize valid cues rather than invalid cues. Overall, there was a small effect (vector $r = .22$), suggesting some overall congruence between the cue utilization and cue validity patterns. However, this effect is considerably smaller than vector correlations reported in a meta-analysis examining deception during more mundane interactions (vector $r = .71$; Hartwig & Bond, 2011) and suggests that judges did not utilize some cues correctly when assessing caller deception.

To gain insight into which cues judges might not have utilized correctly, each cue's cue utilization correlation was compared with its cue validity correlation. For example, the cue "Caller sounded emotional" tended to be displayed by callers whom judges perceived as honest, $r\ (106) = -.32$, $p = .001$. However, this cue tended to be displayed by deceptive callers, $r\ (106) = .29$, $p = .002$. The difference between these effect sizes (r_{diff} = Cue validity r_{pb} − Cue utilization r_{pb}) was positive and significant ($r_{diff} = .61$, $z = 4.69$, $p < .001$), suggesting that judges tended to utilize this cue incorrectly and overattributed it to honesty. In contrast, the cue "Caller is emotionally bland" tended to be displayed by callers whom judges perceived to be deceptive, $r\ (106) = .34$, $p < .001$. Though, this cue tended to be displayed by honest callers, $r\ (106) = -.25$, $p = .011$. The difference between these effect sizes ($r_{diff} = -.59$, $z = -4.50$, $p < .001$) was negative and significant, suggesting that judges utilized this cue incorrectly and overattributed it to deception.

Of the 86 cues examined, 30 (35%) produced significant cue errors. Table 4 displays cues that were utilized incorrectly by the judges (see Supplemental Table for results of all 86,911 Q-sort cues). In this table, positive cue error values are cues judges overattributed to callers' honesty, and negative cue error values are cues judges overattributed to callers' deception.

Discussion

The current study found that judges could not accurately detect the deception of callers reporting homicides to 911. This poor accuracy detection rate (53%) was almost identical to the accuracy rate found in other situations (54%; Bond & DePaulo, 2006). However, while judges' deception detection was inadequate, judges tended to agree with each other in their assessments of 911 callers' deception. Such agreement suggests that judges are not assigning ratings of deception randomly but instead share a common view of how a prototypically deceptive 911 caller behaves. According to Brunwik's lens model, one likely explanation for why judges performed poorly at detecting deception is that this prototypical view of a deceptive 911 caller does not entirely match how deceptive 911 callers genuinely behave.

The cues linked to actual deception were consistent with past theories of deception. For example, various models of deception predict that the cognitive load and uncertainty inherent in crafting false narratives will cause deceivers to tell unhelpful narratives which are less forthcoming and less compelling (DePaulo, 1992; Ekman, 1992; Zuckerman et al., 1981). This conjecture is consistent with the current study's cue validity coefficients (see Table 3) that found deceptive individuals expressed many unhelpful cues, including "Caller gives conflicting facts," "Caller seems to be purposely trying to undermine the victim getting help," and "Caller is self-defensive." Likewise, these models speculate that the disconnect between a deceiver's narrative and reality will cause the leakage of emotionally related cues (DePaulo, 1992; Ekman, 1992; Zuckerman et al., 1981). Consistent with this reasoning, deceptive callers expressed emotional cues, such as "Caller is self-dramatizing," "Caller is overwhelmed," "Caller sounds emotional," and "Caller's mood fluctuates frequently." These findings are similar to those reported earlier by Markey et al. (2022), which employed a smaller subsample of the calls examined in the current study. Overall, it appears that deceptive 911 callers tended to express cues related to being unhelpful and emotional, whereas honest callers were more likely to be helpful and relatively calm.

The importance of helpful and emotional cues was also evident in the cues judges utilized when assessing callers' deception. Similar to the findings related to cue validity, judges were likely to perceive a caller as deceptive when the caller expressed unhelpful cues (e.g., "Caller gives conflicting facts," "Caller seems to be purposely trying to undermine the victim getting help," "Caller is self-defensive," etc.; see Table 1). However, in contrast to the cue validity results, which found deceptive callers tended to be emotional, judges viewed a caller as deceptive when the caller expressed cues related to being relatively calm (e.g., "Caller is emotionally bland," "Caller is relaxed," "Caller is detached from the situation," etc.).

Placing these findings within the context of Brunswik's Lens Model (1956) helps clarify why judges performed poorly when detecting the deception of 911 callers. As discussed earlier, Brunsik's Lens Model suggests that detection deception can be difficult for two possible reasons, (1) valid cues are only weakly related to deception (i.e., the *weak objective cue hypothesis*) or (2) there is a disconnect between valid cues and

how judges utilize cues (i.e., the *wrong subjective cue hypothesis;* Hartwig & Bond, 2011). Given the number of significant cues moderately related to 911 callers' actual deception (see Table 3), there is little support for the weak objective cue hypothesis. However, the disconnect between the cues which were valid indicators of deception (see Table 3) and the cues judges utilized when determining deception (see Table 2) is consistent with the wrong subjective cue hypothesis. Specifically, many of the significant differences between cue utilization and cue validity occurred for emotionally related cues (see Table 4), with judges misattributing various highly emotional cues to assessments of callers' honesty (e.g., "Caller is self-dramatizing," "Caller sounds emotional," etc.) and relatively calm cues to assessments of callers' deception (e.g., "Caller is emotionally bland," "Caller is relaxed," etc.).

Because most judges will never experience calling 911 to report a homicide, the cues they utilize when guessing callers' deception come from their intuition or external sources (Funder, 1995, 1999). Unfortunately, judges likely have trouble understanding how the unique psychological state of a person calling 911 to report a homicide will impact the callers' behavior. This cognitive bias, called the *cold-to-hot empathy gap* (cf., Loewenstein, 1996), implies that judges have difficulty intuitively utilizing deceptive cues correctly because their mental state while listening to 911 calls is dramatically different from the mental state of people calling 911 to report a homicide.

Judges might also incorrectly utilize deception cues due to information from external sources. For instance, popular media often presents a distorted picture of how a criminal behaves, implying that homicide offenders are simply "people who are killers," who suffer from some variant of psychopathy, are unemotional, cold, and lack empathy (Fabianic, 1997). Judges might, therefore, misattribute emotional cues due to these common portrayals of criminals in the media. For example, a judge may incorrectly assume a caller committed a homicide if the caller is emotionally bland, detached from the situation, and relaxed.

In addition to being susceptible to the misleading cues taught by the media, results from the current study suggest that another external source might cause people, including law enforcement officers, to utilize deception cues incorrectly. The *Considering Offender Probability Statements (COPS) scale* was created for members of law enforcement to help gain insight into the deceptiveness of 911 callers (Harpster & Adams, 2017). However, as noted earlier, this scale is based on the findings of Harpster et al. (2009), which other researchers have not been able to replicate (Cromer et al., 2019; M. L. Miller et al., 2021). Results from the current study further indicate that while some of the cues utilized by Harpster et al. (2009) to determine deception are valid cues (e.g., the caller is uncooperative, the caller blames the victim for the situation), many are unrelated (e.g., the caller demands help for the victim, the caller says "huh" or "what"), or are even indicators of honesty (e.g., the caller does not sound emotional, the caller accepts the death of the victim). Such findings suggest that law enforcement officers utilizing these invalid deception cues will likely perform poorly at detecting the deception of homicide reporting callers.

Table 4. Cue Errors: The 911Q Cues With Significantly Different Cue Utilization Correlations and Cue Validity Correlations.

Cue	Cue utilization (r_{pb})	Cue validity (r_{pb})	Cue error (r_{diff})	*p* Value
Cues overattributed to honesty (cue utilization < cue validity)				
83. Caller is self-dramatizing. Theatrical; exaggerates emotions.	−.18	.44	**.62**	<.001
38. Caller sounds emotional	−.32	.29	**.61**	<.001
85. Caller is overwhelmed by the situation	−.19	.36	**.55**	<.001
6. Caller repeats words or phrases with little or no intervening pauses	−.23	.30	**.52**	<.001
37. Caller uses isolated "please" (i.e., the word "please" is said alone)	−.23	.24	**.46**	<.001
32. Caller makes lots of unintelligible comments	−.12	.34	**.46**	<.001
10. Caller acts in a reckless manner.	.01	.47	**.46**	<.001
75. Caller's mood fluctuates frequently	−.08	.33	**.41**	.002
60. Caller speaks in a loud voice	−.30	.07	**.37**	.004
77. Caller sounds depressed.	−.10	.26	**.36**	.006
29. Caller self-interrupts	−.02	.32	**.33**	.010
9. Caller is nervous (e.g., voice is tense, anxious, etc.)	−.09	.24	**.33**	.012
66. Caller seeks reassurance from the dispatcher	−.25	.06	**.31**	.017
49. Caller talks at rather than with the operator	.02	.32	**.30**	.021
58. Caller expresses self-pity or feelings of victimization	−.04	.26	**.29**	.024
72. Caller's behavior is unpredictable	.14	.42	**.28**	.023
24. Caller uses many non sequiturs	.15	.42	**.27**	.026
19. Caller consistently and forcefully demands help for the victim	−.22	.04	**.27**	.040
Cues overattributed to deception (cue utilization > cue validity)				
81. Caller is emotionally bland; has flattened affect.	.34	−.25	**−.59**	<.001
4. Caller responds in ways that are direct, relevant, and clear	.02	−.49	**−.51**	<.001
70. Caller is relaxed	.26	−.25	**−.50**	<.001
12. Caller admits to not knowing information	.14	−.27	**−.41**	.002
15. Caller accepts the death of the victim	.14	−.25	**−.39**	.003
26. Caller uses equivocation (i.e., caller is vague)	.26	−.13	**−.39**	.003
61. Caller is detached from the situation	.35	−.02	**−.36**	.005
82. Caller is verbally fluent.	.03	−.33	**−.36**	.005
11. Caller corrects any errors they made in the call	−.06	−.42	**−.36**	.004
76. Caller focuses on the important issues and provides relevant details	−.15	−.47	**−.32**	.008
71. Caller is moralistic	−.02	−.33	**−.31**	.014
69. Caller seems to be aware of the impression he or she is making	.18	−.11	**−.29**	.026
7. Caller is cooperative and helpful	−.15	−.44	**−.29**	.018
23. Caller minimizes own involvement in event	.24	−.03	**−.27**	.038
57. Caller blames others (for anything)	.03	−.24	**−.27**	.038

Note. $n = 108$, bold values are the differences between cue validity r_{pb} and cue utilization r_{pb}.

Limitations and Conclusions

Results from the current study should be considered within the context of its limitations. The current study operationalized deception based on whether a person calling 911 to report a homicide, who did not confess to the offense during the call, was later convicted of the homicide. Even though this operationalization is consistent with past 911 research (cf., Cromer et al., 2019; M. L. Miller et al., 2021), given the imperfect nature of the criminal justice system, caution is warranted when a criminal conviction is used as a proxy assessment for deception.

Additionally, due to the homogeneous sample of judges used in the current study, it is unclear whether these findings will generalize to a more diverse set of judges, members of law enforcement, or 911 dispatchers. While past research has found that specially trained law enforcement officers perform similarly to lay judges when detecting deception (DePaulo & Pfeifer, 1986; Köhnken, 1987; Vrij, 1993), future research might consider examining deception detection accuracy and cue utilization among this critical group. Furthermore, researchers might consider the unique perspective of 911 dispatchers to help identify other potential deception cues worth investigating.

Because archival 911 calls were used, there was no transparent chain of custody for these calls, and it is unknown whether some calls were edited (e.g., names might have been deleted for privacy reasons). The generalizability of these results might also be limited as some states do not release 911 calls to the public, and all calls were from English speakers. Future researchers might consider obtaining calls directly from law enforcement or examine calls from a more diverse geographic area. Finally, the current study only focused on deception detection during 911 homicide calls. It is hoped that others will examine detection and cue utilization during other types of 911 calls, such as missing persons, aggravated assault, or arson.

Every day, emergency communication centers across the United States receive numerous 911 calls related to homicides. These calls might be from victims before death, innocent witnesses, or the perpetrator of the crime. Such calls have been used to provide insight into the guilt or innocence of suspects in courtrooms and by law enforcement during the early stages of criminal investigations. Unfortunately, the current study found that individuals cannot accurately detect the deception of 911 homicide callers because they often misinterpret the meaning of emotional cues. Therefore, it is strongly recommended that the use of 911 calls to establish a caller's guilt or innocence be done with extreme caution and only with cues that are known to be valid indicators of deception.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Patrick Markey https://orcid.org/0000-0002-2401-2327

Supplemental Material

Supplemental material for this article is available online.

References

Akehurst, L., Köhnken, G., Vrij, A., & Bull, R. (1996). Lay persons' and police officers' beliefs regarding deceptive behavior. *Applied Cognitive Psychology*, *10*, 461–471.

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

Bond, C. F., Jr., Kahler, K. N., & Paolicelli, L. M. (1985). The miscommunication of deception: An adaptive perspective. *Journal of Experimental Social Psychology*, *21*, 331–345

Brunswik, E. (1956). *Perception and the representative design of psychological experiments*. University of California Press.

Burns, M. B., & Moffitt, K. C. (2014). Automated deception detection of 911 call transcripts. *Security Informatics*, *3*, 1–9.

Cheng, K. H. W., & Broadhurst, R. (2005). The detection of deception: The effects of first and second language on lie detection ability. *Psychiatry, Psychology, and Law*, *12*, 107–118.

Cromer, J. D., 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.

DePaulo, B. M. (1992). Nonverbal behavior and self-presentation. *Psychological Bulletin*, *111*, 203–243.

DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological Bulletin*, *129*, 74–118.

DePaulo, B. M., & Pfeifer, R. L. (1986). On-the-job experience and skill at detecting deception. *Journal of Applied Social Psychology*, *16*, 249–267.

Dunkel, C. S., Summerville, L. A., Mathes, E. W., & Kesserling, S. N. (2015). Using the California Q-sort measure of life history strategy to predict sexual behavioral outcomes. *Archives of Sexual Behavior*, *44*, 1705–1711.

Ekman, P. (1992). *Telling lies: Clues to deceit in the marketplace, politics, and marriage*. W.W. Norton (Original work published 1985).

Fabianic, D. (1997). Television dramas and homicide causation. *Journal of Criminal Justice*, *25*, 195–203.

Fleiss, J. L., & Cuzick, J. (1979). The reliability of dichotomous judgments: Unequal numbers of judges per subject. *Applied Psychological Measurement*, *3*, 537–542.

Frank, M. G., & Ekman, P. (2004). Appearing truthful generalizes across different deception situations. *Journal of Personality and Social Psychology*, *86*, 486–495.

Funder, D. C. (1995). On the accuracy of personality judgment: A realistic approach. *Psychological Review*, *102*, 652–670.

Funder, D. C. (1999). *Personality judgment: A realistic approach to person perception*. Academic Press.

Funder, D. C., Furr, R. M., & Colvin, C. R. (2000). The Riverside Behavioral Q-sort: A tool for the description of social behavior. *Journal of Personality*, *68*, 451–489.

Global Deception Research Team. (2006). A world of lies. *Journal of Cross-Cultural Psychology*, *37*, 60–74.

Gosling, S. D., Ko, S. J., Mannarelli, T., & Morris, M. E. (2002). A room with a cue: Personality judgments based on offices and bedrooms. *Journal of Personality and Social Psychology*, *82*, 379–398.

Harpster, T. K. (2006). *The nature of 911 homicide calls: Using 911 homicide calls to identify indicators of innocence and guilt* [Master's thesis]. University of Cincinnati.

Harpster, T. K., & Adams, S. H. (2017). *Analyzing 911 homicide calls: Practical aspects and applications*. CRC Press.

Harpster, T. K., Adams, S. H., & Jarvis, J. P. (2009). Analyzing 911 homicide calls for indicators of guilt or innocence: An exploratory analysis. *Homicide Studies*, *13*, 69–93.

Hartwig, M., & Bond, C. F., Jr. (2011). Why do lie-catchers fail? A lens model meta-analysis of human lie judgments. *Psychological Bulletin*, *137*, 643–659.

Heinrich, C. U., & Borkenau, P. (1998). Deception and deception detection: The role of cross-modal inconsistency. *Journal of Personality*, *66*, 687–712.

HtmlQ. (2019). HtmlQ [computer software]. https://github.com/aproxima/htmlq

Hursch, C. J., Hammond, K. R., & Hursch, J. L. (1964). Some methodological considerations in multiple-cue probability studies. *Psychological Review*, *71*, 42–60.

Köhnken, G. (1987). Training police officers to detect deceptive eyewitness statements: Does it work? *Social Behaviour*, *2*, 1–17.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*, 159–174.

Loewenstein, G. (1996). Out of control: Visceral influences on behavior. *Organizational Behavior and Human Decision Processes*, *65*, 272–292.

Marcus, B., Machilek, F., & Schütz, A. (2006). Personality in cyberspace: Personal websites as media for personality expressions and impressions. *Journal of Personality and Social Psychology*, *90*(6), 1014–1031.

Markey, P. M., Feeney, E., Berry, B., Hopkins, L., & Creedon, I. (2022). Deception cues during high-risk situations: 911homicide calls. *Psychological Science*, *33*, 1040–1047.

Miller, G. R., & Stiff, J. B. (1993). *Deceptive communication*. SAGE.

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

Ozer, D. J. (1993). The Q-sort method and the study of personality development. In D. C. Funder, R. D. Parke, C. Tomlinson-Keasey, & K. Widaman (Eds.), *Studying lives through time: Personality and development* (pp. 147–168). American Psychological Association.

Porter, S., & ten Brinke, L. (2010). The truth about lies: What works in detecting high-stakes deception? *Legal and Criminological Psychology*, *15*, 57–75.

Richard, F. D., Bond, C. F., Jr., & Stokes-Zoota, J. J. (2003). One hundred years of social psychology quantitatively described. *Review of General Psychology*, *7*, 331–363.

Riggio, R. E., & Friedman, H. S. (1983). Individual differences and cues to deception. *Journal of Personality and Social Psychology*, *45*, 899–915.

Serfass, D. G., & Sherman, R. A. (2013). A methodological note on ordered Q-Sort ratings. *Journal of Research in Personality*, *47*, 853–858.

Sherman, R. A., Figueredo, A. J., & Funder, D. C. (2013). The behavioral correlates of overall and distinctive life history strategy. *Journal of Personality and Social Psychology*, *105*, 873–888.

Sherman, R. A., & Funder, D. C. (2009). Evaluating correlations in studies of personality and behavior: Beyond the number of significant findings to be expected by chance. *Journal of Research in Personality*, *43*, 1053–1063.

Thornton, H. (1995). *Hung jury: The diary of a Menendez Juror*. Temple University Press.

Vrij, A. (1993). Credibility judgments of detectives: The impact of nonverbal behavior, social skills, and physical characteristics on impression formation. *The Journal of Social Psychology*, *133*, 601–610.

Zuckerman, M., DePaulo, B. M., & Rosenthal, R. (1981). Verbal and nonverbal communication of deception. *Advances in Experimental Social Psychology*, *14*, 1–59.

Author Biographies

Patrick M. Markey is a professor in the department of Psychological and Brain Sciences and the director of the Interpersonal Research Group at Villanova University. His research focuses primarily on how behavioral tendencies develop and are expressed within a social environment.

Erika Feeney is a graduate of Villanova's Master of Psychology program, where she researched detecting deception in high-stress situations, specifically 911 homicide calls. She is currently working at Drexel University as a lab manager and research assistant focusing on substance use and suicide prevention research.

Brooke Berry is a former graduate student at Villanova University, with her research focusing on detecting deception in 911 missing person calls. She is currently working as a research associate at Veris Insights, a market research company.

Alec Martin is a graduate student at Villanova University researching how cues to deception are expressed during high-stakes situations.

Erica Slotter is an Associate Professor at Villanova University. She received her Ph.D. in social psychology from Northwestern University and studies the influence of the social world on identity and relationship processes.

Research Article

Psychological Science
1–8

DOI: 10.1177/09567976221077216
www.psychologicalscience.org/PS

Deception Cues During High-Risk Situations: 911 Homicide Calls

Patrick M. Markey, Erika Feeney, Brooke Berry, Lauren Hopkins, and Isabel Creedon
Department of Psychological and Brain Sciences, Villanova University

Abstract

During everyday interactions, cues tend to be weakly related to deception. However, there are theoretical reasons to suspect that such cues will be more prominent during high-risk interactions. The current study explored deception cues during one particular high-risk interaction—911 homicide calls placed by adults. In Sample 1, judges coded 911 homicide calls (n = 82) by Q-sorting 86 cues. Results indicated that deceptive callers tended to display emotional cues (e.g., self-dramatizing, moody, worried, emotional, nervous), appeared overwhelmed, and related narratives that lacked structure, clarity, and focus. Judges coded a separate sample of 911 calls (n = 64), and *deception scores* were computed using a template-matching approach based on the findings from Sample 1. Results indicated that deceptive 911 callers had higher deception scores than honest callers. The effect sizes yielded in this study highlight the relevance of deception cues during high-risk interactions and the usefulness of the person-centered Q-sort method.

Keywords

violent crime, judgment, interpersonal interaction, social behavior, open data, preregistered

Received 9/24/21; Revision accepted 12/9/21

It is difficult to accurately detect deception in other people (Bond & DePaulo, 2006). Even specially trained law-enforcement officers have trouble recognizing when deception occurs (DePaulo & Pfeifer, 1986; Köhnken, 1987; Vrij, 1993). Although it may be challenging to correctly judge when someone is lying, the act of deception often leaves a trail of cues. One large-scale meta-analysis examining deception in everyday interactions and university laboratories found that deceivers tended to appear nervous, tense, and uncooperative and told uncompelling narratives that lacked structure and logic (DePaulo et al., 2003). This trail of cues is consistent with some models of deception (cf. DePaulo, 1992; Ekman, 1985/1992; Zuckerman et al., 1981), but the effect sizes linking any single cue to deception during everyday interactions were comparatively low (range r_{pb} = .00–.31; median r_{pb} = .05).

There are theoretical and empirical reasons to suspect that the trail of cues related to deception will be more pronounced during high-risk interpersonal interactions, such as 911 homicide calls, than in everyday situations. First, cues related to deception may become more evident when people are lying about serious transgressions (e.g., committing a homicide) than mundane issues (G. R. Miller & Stiff, 1993). Second, high-risk interactions are typically accompanied by strong emotions that are difficult to fake (Porter & ten Brinke, 2010). Third, 911 calls are usually placed soon after a crime, giving deceptive individuals little time to create and rehearse a false narrative (Harpster et al., 2009). Finally, a person who has committed homicide is unlikely to know how an innocent person would typically behave when calling 911.

Harpster and colleagues (Harpster, 2006; Harpster et al., 2009) carried out one of the earliest systematic attempts to examine cues related to deception during 911 homicide calls. Harpster et al. (2009) reported the same data and results as those reported by Harpster (2006) and are therefore discussed here as a single

Corresponding Author:
Patrick M. Markey, Villanova University, Department of Psychological and Brain Sciences
Email: patrick.markey@villanova.edu

study. In the Harpster study, 20 dichotomous cues were related to the deception or honesty of 911 homicide callers. Deceptive callers tended to provide extraneous information, gave conflicting facts, and were resistant to answering questions from the 911 operator (see Table 1). On the other hand, honest callers were more likely to make demanding and urgent pleas while displaying high levels of voice modulation (i.e., emotionally charged speech).

The results from the Harpster study need to be considered in light of their limitations. Only a single judge, the primary author, seems to have coded the 911 calls. Because this judge also "personally contacted" each lead detective to obtain the 911 calls, it is unclear whether the judge was unaware of the guilt or innocence of the callers (Harpster, 2006). The possibility of nonmasked coding might explain why some of the effect sizes in this study seemed unreasonably large. For example, the single dichotomous cue "extraneous information" was correlated with deception (r = .81) and accurately predicted deception 91% of the time. If this effect size is accurate, it would be the largest effect size found examining deceptive behaviors and among the largest effect sizes ever discovered in the social sciences (DePaulo et al., 2003; Richard et al., 2003).

Recently, Cromer et al. (2019) and M. L. Miller et al. (2021) attempted to replicate a set of the findings from the Harpster study using multiple judges unaware of the deception of the 911 callers. Unfortunately, only two of the 22 (9%) replication analyses conducted by Cromer et al. and M. L. Miller et al. produced results similar to those in the Harpster study (see Table 1). However, it is essential to note that the studies by Cromer et al. and M. L. Miller et al. suffered from low power. Assuming a moderate effect size (r = .30) with a nondirectional test, the power estimates for these studies, given their unequal sample sizes, ranged between .48 and .62. Therefore, it is unclear whether these null results were due to a lack of a relationship between cues and deception or to inadequate power.

Past research examining 911 homicide calls has suffered from nonmasked-coding issues, low power, and inconsistent results. None of these studies investigated whether the results found in their exploratory analyses could predict deception using an independent sample. Additionally, because most studies on deception have tended to produce low to moderate effect sizes (DePaulo et al., 2003), it seems unlikely that any single cue would be practical for accurately detecting deception. Instead, exploring how a trail of cues combine within deceptive individuals will likely provide a more holistic perspective than examining cues in isolation (Ozer, 1993).

Statement of Relevance

Every day, emergency communication centers across the United States receive numerous 911 calls related to homicides. These calls might be from victims before death, innocent witnesses, or the perpetrator of the crime. Can the cues that callers display during these 911 calls be used to determine which callers are deceptive and guilty of homicide and which are innocent? We found that deceptive callers displayed a pattern of overly emotional cues, acted overwhelmed, and told narratives that lacked clarity. We also found that this unique pattern of deceptive cues can be used to help establish the guilt or innocence of 911 homicide callers. These findings suggest that law-enforcement officers and other people can use the pattern of cues displayed during 911 homicide calls to help identify people and areas of interest.

The Current Study

The current research used masked judges, examined a large sample of 911 homicide calls, employed a person-centered Q-sort methodology, and subsequently examined whether findings from the initial analysis could predict deception in an independent sample. Using the Q-sort methodology, a group of judges Q-sorted 86 cues expressed during 911 calls. These ratings were then used to generate a template of cues that distinguished deceptive 911 callers from honest 911 callers. In a second sample, this template was compared with the cue ratings of 911 callers to assess each caller's similarity to the prototypical cues of a deceptive 911 caller. In this manner, the probability of a caller being deceptive is viewed as a monotonically increasing function of how well this caller's cues matched the template of a prototypically deceptive individual (Bem & Funder, 1978; Reise & Oliver, 1994).

This study's research hypotheses, methodology, and analytic plans were preregistered on OSF before the data were coded (https://osf.io/pvsm3/). The data and coding instructions for the study have been made publicly available on OSF as well (https://osf.io/v4dx7/). The preregistration for this study discussed four planned analyses, both exploratory and confirmatory.

Exploratory analyses

In Sample 1, 86 cues were correlated to the callers' deception or honesty to create a deception template of

Table 1. Findings From Three Studies Examining the Link Between Various Cues and Deception During 911 Homicide Calls

Cue	Harpster et al. (2009)	Cromer et al. (2019)	M. L. Miller et al. (2021)
Extraneous information	Deception	Deception	Unrelated
Conflicting facts	Deception	Deception	Unrelated
Resistance to answer	Deception	Unrelated	Unrelated
Acceptance of death with relation	Deception		Unrelated
Inappropriate politeness	Deception	Unrelated	Unrelated
Repetition	Deception		Unrelated
Acceptance of death	Deception	Unrelated	Unrelated
Possession of problem	Deception	Unrelated	
Thinking pause	Deception	Unrelated	Unrelated
Plea for caller only	Deception		
Insulting the victim	Deception		Unrelated
Minimizing just	Unrelated	Unrelated	Unrelated
Minimizing just early	Unrelated		
Verbal reaction	Honesty		
Plea for help	Honesty		
Self-correction	Honesty		
Plea for victim only	Honesty		Unrelated
Location of plea	Honesty		
Voice modulation	Honesty		Deception
Urgency of plea	Honesty		Deception
Demanding plea	Honesty		Unrelated

Note: Cues were coded as being positively related to deception ("deception"), negatively related to deception ("honesty"), or unrelated to deception ("unrelated").

a prototypical deceptive caller. Our preregistration stated that a randomization test would be utilized to assess whether the number of significant correlations yielded when creating the deception template was greater than expected by chance (Sherman & Funder, 2009). Hypothesis 1 was that the number of significant correlations found in the initial analysis would be significantly larger than the number of significant correlations obtained in the randomization test.

We further preregistered our intention to assess the overall effect size of the deception-template cues in Sample 1, using a randomization test to evaluate the difference in the mean absolute effect size found when creating the deception template and the mean absolute effect size expected compared with chance (Sherman & Funder, 2009). Hypothesis 2 was that the mean absolute effect size yielded in the initial analysis would be significantly larger than the mean absolute effect size obtained in the randomization test.

Confirmatory analysis

In Sample 2, our preregistered intention was to examine a separate set of 911 homicide calls using the deception template and compute deception scores for each caller by examining the similarity of each caller's 86 cue ratings to the deception template derived in Sample 1 of the prototypical deceptive caller. Hypothesis 3 was that deceptive callers would yield greater deception scores than honest callers.

Method

Data and sources

Calls placed to 911 were deemed eligible for the study on the basis of criteria similar to those used by Cromer et al. (2019): (a) The call involved the killing of another person; (b) emergency services were notified; (c) the caller was aware of and able to communicate the general nature of the emergency; (d) at least two news sources could verify prosecution, admission of guilt, or another outcome resulting from the call; and (e) the caller did not confess to wrongdoing. Callers claiming extenuating circumstances (e.g., self-defense, accident) that led to the death of another person were also included.

Following prior 911 research (Cromer et al., 2019; M. L. Miller et al., 2021), we obtained audio calls from publicly available open-source data, such as news sources, police department releases, and various archives. An

a priori power analysis determined that a sample size of 82 (41 deceptive callers and 41 honest callers) for Sample 1's exploratory analyses and a sample size of 64 (32 deceptive callers and 32 honest callers) for Sample 2's confirmatory analysis would be necessary to achieve 80% power to detect a moderate effect (r_{pb} = .30).

Determination of deception

At least two external sources (usually media reports of the crime) were used to determine callers' deception or honesty. To be coded as "deceptive," the caller was required to have been found guilty in a court of law. When an indictment was not possible, such as the death of the caller, expert opinions were employed (e.g., medical examiner, police investigators, grand jury finding). Because none of the callers used in the study confessed to wrongdoing during the call, callers coded as "deceptive" were both those who lied by commission (i.e., the active use of false statements) and those who lied by omission (i.e., the passive omission of relevant information) during the 911 call. This coding is consistent with past research, which has defined *deception* as the use of statements or acts of omission that intentionally mislead (cf. Gaspar & Schweitzer, 2013; Tenbrunsel & Messick, 2004). The remaining callers not deemed guilty were coded as "honest." This coding method is consistent with past 911 research (Cromer et al., 2019; M. L. Miller et al., 2021).

Coding of 911 calls

The 911 Q sort was designed to code audio recordings of 911 homicide calls. Cues in the 911 Q sort were created to capture audio cues at a psychologically meaningful level, requiring as little subjective interpretation from the coders as possible. First, a set of relevant cues was generated from past 911 studies (e.g., Harpster et al., 2009), from research examining deception cues in everyday interactions (e.g., DePaulo et al., 2003), and from the Riverside Behavioral Q-Sort (an assessment designed to examine various cues during dyadic interactions; Funder et al., 2000). Next, a group of researchers eliminated overly redundant cues and reworded cues to be relevant in the context of 911 calls. The final set of 86 cues ranged from items directly relevant to 911 calls (e.g., "Caller makes the dispatcher confused," "Caller quickly asks for help for the victim") to cues that assessed a caller's general behavior (e.g., "Caller acts in a reckless manner," "Caller is talkative"). See https://osf.io/v4dx7/ for a complete list of the 911 Q-sort cues.

The 911 Q sort was conducted using a Q-sort methodology (Ozer, 1993). Three judges, unaware of the deception of the caller, listed to each 911 call and then independently coded the audio of all 146 calls (82 calls for Sample 1 and 64 calls for Sample 2). See https://osf.io/9hkaf/ for a copy of the online instructions associated with coding procedures. Judges were graduate research assistants whose training consisted of reviewing the eighty-six 911 Q-sort cues and then receiving directions on the Q-sorting procedures and instructions concerning the practical issues involving accessing the 911 audio clips and how to access the Q-sort program. Before coding the data set, all judges practiced Q-sorting five 911 calls (not included in the final analysis), and discrepancies and concerns among the coders were discussed until resolved. Judges used a modified version of the online software *HTMLQ* (Version 2.0; Killing, 2019) to sort the eighty-six 911 Q-sort cues into nine categories (1 = *extremely uncharacteristic*, 9 = *extremely characteristic*); cues were distributed as follows: 3, 6, 10, 15, 18, 15, 10, 6, and 3, respectively. By forcing judges to compare each cue with other cues, the Q-sort methodology produces a person-focused description (Ozer, 1993). Although time consuming, Q sorts are less susceptible to biases such as extremity bias, midpoint responding, acquaintance bias, and halo effects than other more traditional assessments (Ozer, 1993; Serfass & Sherman, 2013). The median cue reliability was .63, which is similar to the reliability of single items in other behavioral Q sorts (cf. Dunkel et al., 2015; Sherman et al., 2013). More importantly, as discussed later, judges' agreement was .86 when the entire 911 Q-sort deception template was used to compute 911 callers' overall deception scores.

Results

Exploratory analyses

Analyses were first conducted to create a template of 911 Q-sort cues that distinguished deceptive 911 callers from honest 911 callers. Point-biserial correlations were computed between callers' deception (coded 1 = *deceptive caller* and 0 = *honest caller*) and each of the eighty-six 911 Q-sort cues. Table 2 displays the cues that were significantly related to deception. The resulting pattern of correlations between the eighty-six 911 Q-sort cues and deception served as the template for the pattern of cues that differentiated a prototypical deceptive caller from an honest caller (results for all eighty-six 911 Q-sort cues are available at https://osf.io/v4dx7/). The replicability of this pattern of correlations was examined using Sherman and Wood's (2014) split-sample (SS)

Table 2. Significant Correlations in Sample 1 Between the 911 Q-Sort Items and Callers' Deception During 911 Homicide Calls (n = 82)

911 Q-sort item	r_{pb}
Positive correlations (indicating deception)	
10. Caller acts in a reckless manner.	.40**
83. Caller is self-dramatizing; histrionic. Theatrical; exaggerates emotions.	.38**
75. Caller's mood fluctuates frequently.	.38**
24. Caller uses many non sequiturs.	.36**
54. Caller seems to be purposely trying to sabotage or obstruct the victim getting help.	.35**
72. Caller's behavior is unpredictable.	.33**
65. Caller is self-defensive.	.33**
27. Caller is evasive.	.32**
20. Caller expresses worry that they might be blamed.	.32**
78. Caller interprets simple situations in unnecessarily complicated ways.	.32**
58. Caller expresses self-pity or feelings of victimization.	.29**
21. Caller expresses ingratiating remarks to gain favor.	.28*
22. Caller blames the victim for the situation.	.28*
29. Caller self-interrupts.	.26*
77. Caller sounds depressed.	.26*
85. Caller is overwhelmed by the situation.	.25*
48. Caller is awkward.	.25*
38. Caller sounds emotional.	.24*
33. Caller makes attempts to convince dispatcher of innocence.	.24*
25. Caller makes the dispatcher confused.	.24*
9. Caller is nervous.	.22*
49. Caller talks at rather than with the operator.	.22*
Negative correlations (indicating honesty)	
74. Caller is forthright and candid.	−.48**
76. Caller focuses on the important issues and provides relevant details.	−.46**
63. Caller is working hard and is focused on the event.	−.41**
7. Caller is cooperative and helpful.	−.39**
86. Caller expresses distrust (at anyone).	−.38**
4. Caller responds in ways that are direct, relevant, and clear.	−.38**
2. Caller's message is plausible and believable.	−.36**
11. Caller corrects any errors they made in the call.	−.36**
67. Caller behaves and acts quickly.	−.36**
3. Caller's message is consistent and provides a coherent sequence of events.	−.30**
12. Caller admits to not knowing information.	−.28*
57. Caller blames others (for anything).	−.27*
64. Caller appears to have a high degree of intellectual capacity.	−.26*
71. Caller is moralistic.	−.26*
15. Caller accepts the death of the victim.	−.24*
53. Caller acts irritated (toward anyone).	−.23*
40. Caller apologizes to the victim.	−.22*

*p < .05. **p < .01.

methodology, which utilized 1,000 random samples to estimate the template's reliability ($\hat{\rho}_{ss}$ = .80, SE = .06). This finding indicates that if the current study were replicated, the pattern of correlations in the replication study would be expected to have a correlation of .80 with the current study's deception template.

The previous set of 86 analyses found 39 significant effects linking 911 Q-sort cues to deception. However, given the number of nonindependent analyses conducted, some of these effects might be significant simply because of chance. Sherman and Funder's (2009) randomization method was therefore employed to

Table 3. Results From Randomization Tests of 911 Q-Sort Correlates of Deception in Sample 1

Statistic	Observed value	Value expected by chance	*SE*	*p*	95th percentile
Number of significant ($p < .05$) results	39	4.247	3.802	< .0001	12
Mean absolute r_{pb}	.193	.089	.016	< .0001	.119

determine the probability of obtaining 39 significant results under a random model of no association between the deception and the 911 Q-sort cues. This was done by (a) randomly redistributing deceptive and honest codes and (b) computing the number of significant ($p < .05$) correlations between deceptiveness and the 911 Q-sort cues in this new sample. This procedure was repeated 10,000 times. The resulting values were used to form a sampling distribution indicating the number of expected significant effects under the null hypothesis of no relation between deception and the 911 Q-sort cues. Consistent with Hypothesis 2, this analysis found that the probability of the current study finding 39 significant ($p < .05$) correlations linking deception to the 911 Q-sort cues by simple chance was $p < .0001$ (see Table 3).

Randomization tests were again used to estimate the significance of the mean effect size found linking deception to the entire set of 911 Q-sort cues (Sherman & Funder, 2009). The absolute correlation value between caller deceptiveness and the entire set of eighty-six 911 Q-sort cues was first computed (average $r_{pb} = .19$). As in the previous analysis, 10,000 random samples were used to estimate the probability of obtaining this observed effect size under a random model of no association between deception and the 911 Q-sort cues. Consistent with Hypothesis 3, this analysis indicated that the chance of the current study finding this observed effect size linking deception to the 911 Q-sort cues was $p < .0001$ (see Table 3). Taken together, the results from Hypotheses 2 and 3 suggest that the 911 Q-sort deception template displayed in Table 2 is real (i.e., beyond chance) and can likely be used to predict deception during 911 calls.

Confirmatory analysis

Deception scores were computed in Sample 2 using a template-matching approach (Bem & Funder, 1978; Reise & Oliver, 1994). This approach entails correlating (i.e., matching) the observed pattern of a caller's 911 Q-sort cues with the 911 Q-sort deception template derived from Sample 1. Specifically, each judge's sort of each 911 call was correlated with the matched 86 effect sizes of the 911 Q-sort deception template (see Table 2). To ease interpretation, we standardized the resulting correlations before conducting any analyses. Therefore, high deception scores indicate that the caller's 911 Q-sort pattern was similar to the pattern of deceptive individuals, and low deception scores indicate that the caller's 911 Q-sort pattern was similar to the pattern of honest callers. The three judges' deception scores were then aggregated for each 911 call; judges' reliability for deception scores was .86. Consistent with Hypothesis 3, results for Sample 2 showed that deceptive 911 callers were significantly more likely to have higher deception scores than honest callers, $t(62) = 5.29$, $p < .001$, $d = 1.32$, 95% confidence interval (CI) = [0.78, 1.86] (see Fig. 1).

Discussion

Every day, emergency communication centers across the United States receive numerous 911 calls related to homicides. These calls might be from victims before death, innocent witnesses, or the perpetrator of the crime. Such calls could be a crucial investigative tool by providing a unique insight into the guilt or innocence of a suspect. Consistent with this notion, results of the current study showed that cues expressed during 911 homicide calls were related to the deception of the caller.

Although we made no specific predictions concerning individual cues in the initial exploratory analysis, many of the study's findings were consistent with models of deception that suggest that the disconnect between a deceiver's narrative and reality causes the leakage of emotionally related cues (DePaulo, 1992; Ekman, 1985/1992; Zuckerman et al., 1981). For example, as seen in Table 2, deceptive callers were self-dramatizing, moody, reckless, worried, depressed, emotional, and nervous. Likewise, these models predict that the cognitive load and uncertainty inherent in crafting false narratives cause deceivers to tell less-than-compelling narratives. Again, this conjecture is consistent with the current study's findings that deceptive individuals appeared overwhelmed and related narratives that lacked structure, clarity, and focus.

Such findings have implications for understanding how high-risk situations might alter the importance of cues during deception and may serve as a helpful tool for detecting criminal deception. Unfortunately, past research has found that even specially trained law-enforcement officers tend to be poor at detecting

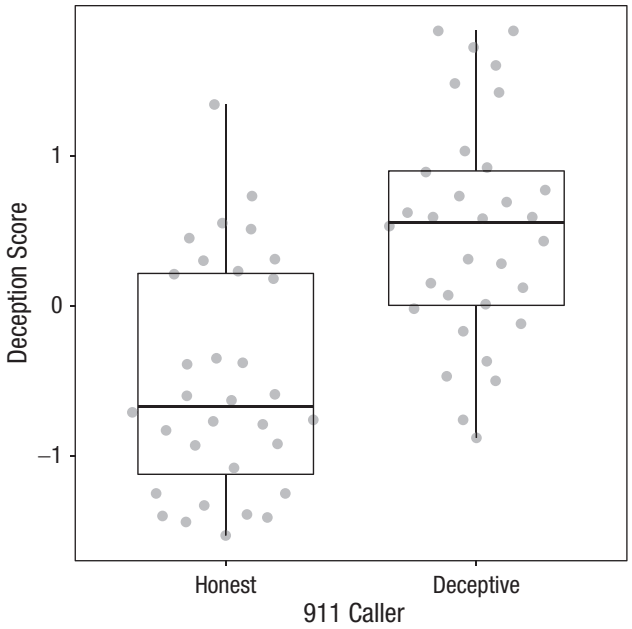

Fig. 1. Deception scores for honest and deceptive 911 callers in Sample 2. Each box represents the interquartile range (IQR), the whiskers represent the range of values within 1.5 times the IQR, and the horizontal line represents the median. Dots represent individual data.

deception (DePaulo & Pfeifer, 1986; Köhnken, 1987; Vrij, 1993). One possible explanation for this finding is that judges sometimes use invalid cues when determining whether a person is honest. For example, speech pauses, "ums," and "huhs" are often believed to be cues related to deception, but research (and the current study; see https://osf.io/v4dx7/) shows that such speech patterns are not valid cues to deception (Davis et al., 2005; DePaulo et al., 2003). However, this does not imply that judges are inaccurate at detecting cues, only that they have trouble understanding how the pattern of cues they observe is related to deception.

As in research examining clinical and personality judgments, the most accurate predictions regarding deception likely result from using judge-rated cues as input into a statistical model that accounts for multiple cues (e.g., template matching; see Wiggins, 1973, for a review). Consistent with this notion, our results showed that when 911 homicide calls were scored by applying the deception template to judge-rated cues, deceptive callers received substantially higher deception scores than honest callers. Furthermore, given past attempts at predicting deception during 911 calls using nonmasked judges (Cromer et al., 2019; M. L. Miller et al., 2021) or during everyday interactions (DePaulo et al., 2003), the effect size yielded from this analysis was larger than expected ($d = 1.32$) and highlights the usefulness of the person-centered Q-sort method.

Although the current study predicted a 911 caller's deception using the 911 Q sort, caution is warranted for anyone basing a caller's guilt or innocence solely on cues expressed during a 911 call. Such information might help law-enforcement officers identify people and areas of interest, but it would be a mistake to use it to make a definitive conclusion concerning criminal activity. Additionally, results from the current study should be considered within the context of its limitations. Because archival 911 calls were used, there was no transparent chain of custody for these calls. Therefore, it is unknown whether some calls were edited (e.g., names might have been deleted for privacy reasons). The generalizability of these results might be limited because some states do not release 911 calls to the public and all calls were from English speakers. Future researchers might consider obtaining calls directly from law enforcement or examining calls from a more diverse geographic area. Finally, the current study operationalized deception on the basis of whether a 911 homicide caller who did not confess to wrongdoing during the call was later convicted of the homicide. Although this operationalization is consistent with past 911 research (cf. Cromer et al., 2019; M. L. Miller et al., 2021), given the imperfect nature of the criminal justice system, caution is warranted when using criminal conviction as a proxy for deception.

The Q-sort methodology employed in the current study was time consuming, with judges taking approximately 25 min to listen and code each 911 call. It might, therefore, be of practical importance to reduce the number of 911 Q-sort cues to make this coding process more efficient. For example, an auxiliary analysis found that the deception scores computed using all eighty-six 911 Q-sort cues in Sample 2 were highly similar, $r(62) = .98$, $p < .001$, to the deception scores computed when only the 39 significant cues were used. Future researchers might also consider using automatic coding methodologies (e.g., voice-prosody analysis, natural-language processing) to compute deception scores more quickly. Finally, although most of the 911 Q-sort cues apply to a wide variety of crimes, the current study focused only on 911 homicide calls. It is hoped that others employ the person-centered Q-sort method presented here to examine the possibility of detecting deception in other high-risk criminal situations, such as missing-persons cases, aggravated assault, or arson.

Transparency

Action Editor: Kate Ratliff
Editor: Patricia J. Bauer
Author Contributions

P. M. Markey developed the study concept. L. Hopkins and I. Creedon contributed to the study design. Data collection was supervised by E. Feeney, B. Berry, and I. Creedon. P. M. Markey analyzed the data and drafted the manuscript. E. Feeney and B. Berry provided critical revisions to the manuscript. All the authors approved the final manuscript for submission.

Declaration of Conflicting Interests

The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

Open Practices

All data and coding instructions have been made publicly available via OSF and can be accessed at https://osf.io/v4dx7/. The design and analysis plans for the study were preregistered at https://osf.io/pvsm3/. This article has received the badges for Open Data and Preregistration. More information about the Open Practices badges can be found at http://www.psychologicalscience.org/publications/badges.

References

Bem, D. J., & Funder, D. C. (1978). Predicting more of the people more of the time: Assessing the personality of situations. *Psychological Review*, *85*, 485–501.

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

Cromer, J. D., 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.

Davis, M., Markus, K. A., Walters, S. B., Vorus, N., & Connors, B. (2005). Behavioral cues to deception vs. topic incriminating potential in criminal confessions. *Law and Human Behavior*, *29*, 683–704.

DePaulo, B. M. (1992). Nonverbal behavior and self-presentation. *Psychological Bulletin*, *111*, 203–243.

DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological Bulletin*, *129*, 74–118.

DePaulo, B. M., & Pfeifer, R. L. (1986). On-the-job experience and skill at detecting deception. *Journal of Applied Social Psychology*, *16*, 249–267.

Dunkel, C. S., Summerville, L. A., Mathes, E. W., & Kesserling, S. N. (2015). Using the California Q-sort measure of life history strategy to predict sexual behavioral outcomes. *Archives of Sexual Behavior*, *44*, 1705–1711.

Ekman, P. (1992). *Telling lies: Clues to deceit in the marketplace, politics, and marriage*. W. W. Norton. (Original work published 1985)

Funder, D. C., Furr, R. M., & Colvin, C. R. (2000). The Riverside Behavioral Q-Sort: A tool for the description of social behavior. *Journal of Personality*, *68*(3), 451–489.

Gaspar, J. P., & Schweitzer, M. E. (2013). The emotion deception model: A review of deception in negotiation and the role of emotion in deception. *Negotiation and Conflict Management Research*, *6*, 160–179.

Harpster, T., Adams, S. H., & Jarvis, J. P. (2009). Analyzing 911 homicide calls for indicators of guilt or innocence: An exploratory analysis. *Homicide Studies*, *13*(1), 69–93.

Harpster, T. K. (2006). *The nature of 911 homicide calls: Using 911 homicide calls to identify indicators of innocence and guilt* (Document No. ucin1163534092) [Master's thesis, University of Cincinnati]. http://rave.ohiolink.edu/etdc/view?acc_num=ucin1163534092

Killing, M. (2019). *HTMLQ* (Version 2.0) [Computer software]. GitHub. https://github.com/aproxima/htmlq

Köhnken, G. (1987). Training police officers to detect deceptive eyewitness statements: Does it work? *Social Behaviour*, *2*(1), 1–17.

Miller, G. R., & Stiff, J. B. (1993). *Deceptive communication*. SAGE.

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.

Ozer, D. J. (1993). The Q-sort method and the study of personality development. In D. C. Funder, R. D. Parke, C. Tomlinson-Keasey, & K. Widaman (Eds.), *Studying lives through time: Personality and development* (pp. 147–168). American Psychological Association.

Porter, S., & ten Brinke, L. (2010). The truth about lies: What works in detecting high-stakes deception? *Legal and Criminological Psychology*, *15*, 57–75.

Reise, S. P., & Oliver, C. J. (1994). Development of a California Q-set indicator of primary psychopathy. *Journal of Personality Assessment*, *62*, 130–144.

Richard, F. D., Bond, C. F., Jr., & Stokes-Zoota, J. J. (2003). One hundred years of social psychology quantitatively described. *Review of General Psychology*, *7*, 331–363.

Serfass, D. G., & Sherman, R. A. (2013). A methodological note on ordered Q-Sort ratings. *Journal of Research in Personality*, *47*, 853–858.

Sherman, R. A., Figueredo, A. J., & Funder, D. C. (2013). The behavioral correlates of overall and distinctive life history strategy. *Journal of Personality and Social Psychology*, *105*, 873–888.

Sherman, R. A., & Funder, D. C. (2009). Evaluating correlations in studies of personality and behavior: Beyond the number of significant findings to be expected by chance. *Journal of Research in Personality*, *43*, 1053–1063.

Sherman, R. A., & Wood, D. (2014). Estimating the expected replicability of a pattern of correlations and other measures of association. *Multivariate Behavioral Research*, *49*, 17–40.

Tenbrunsel, A. E., & Messick, D. M. (2004). Ethical fading: The role of self-deception in unethical behavior. *Social Justice Research*, *17*, 223–236.

Vrij, A. (1993). Credibility judgments of detectives: The impact of nonverbal behavior, social skills, and physical characteristics on impression formation. *The Journal of Social Psychology*, *133*, 601–610.

Wiggins, J. S. (1973). *Personality and prediction: Principles of personality assessment*. Addison-Wesley.

Zuckerman, M., DePaulo, B. M., & Rosenthal, R. (1981). Verbal and nonverbal communication of deception. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (Vol. 14, pp. 1–59). Academic Press.