Failing to Express Emotion on 911 Calls Triggers Suspicion Through Violating Expectations and Moral Typecasting

Jessica M. Salerno[1,2], Samantha R. Bean[1], Nicholas D. Duran[1], Alia N. Wulff[1],
Isabelle Reeder[1], and Saul M. Kassin[3]
[1] School of Social and Behavioral Sciences, Arizona State University
[2] Department of Psychology, Cornell University
[3] Department of Psychology, John Jay College of Criminal Justice, City University of New York

Coming to suspect that someone has engaged in wrongdoing based on their unexpected behavior is a common phenomenon—yet, little is known about what triggers initial suspicion. We investigated how violating expectations for high emotionality during a traumatic event can trigger suspicion that one has engaged in immoral—or even criminal—activity through moral typecasting. Five studies demonstrate this theory in a criminal context with dire real-world consequences: 911 callers reporting violent crimes generating suspicion by exhibiting unexpected behavior, which could trigger confirmation bias in investigations leading to wrongful convictions. Using both real and tightly controlled simulated 911 calls, we demonstrated that failing to express the expected level of emotion on a 911 call reporting a violent crime led laypeople and police to morally typecast the caller as more of a moral agent capable of perpetrating immoral acts and less of a moral patient capable of being the victim of immoral acts—ultimately increasing suspicion that they were involved in the crime and support for treating them as a suspect. We advance moral psychological theory by demonstrating that failing to express expected levels of emotion about a moral violation can shape moral inferences about someone's capacity to commit versus be the victim of moral wrongs, thereby generating suspicion that they might have engaged in wrongdoing. We demonstrated this theory in criminal settings to explain how one tragedy can become two: altruistic witnesses calling 911 to plead for help on behalf of another person becoming suspects of the crime they reported because they failed to exhibit the expected emotional demeanor.

Statement of Limitations

We replicated our findings across both tightly controlled, simulated calls and a diverse set of real 911 calls, modeling the content of the 911 calls as a random category. This enabled us to conceptually generalize the current results to the larger universe of unsampled calls, similar to how researchers generalize results beyond unsampled participants (Judd et al., 2012). However, our conclusions should be somewhat limited by methodological choices. Our emotion expression construct is limited to general emotionality, preventing us from drawing conclusions about specific discrete emotions. Our findings are also limited to a context where high emotion is expected—investigating whether the findings would generalize to violating expectations for low emotion on moral typecasting is important. Finally, it is well-established that people of color face increased suspicion and likelihood of wrongful conviction, which we were not able to address because the race of 911 callers in our studies is unknown. Further, our samples were predominantly White—an issue that mirrors the underrepresentation of people of color on the police force, judicial bench, and juries. A very important next step, however, would be to test whether our findings are moderated by caller or participant race.

Many of us have had the experience of beginning to suspect someone of wrongdoing and concealment based on a hunch formed by observing their demeanor. Perhaps we start to suspect that the anxious student taking an exam is cheating, or the partner who seems distant is having an affair. Once someone falls under the lens of suspicion, whether guilty or not, we tend to seek out, attend to, and consider confirmatory evidence more than evidence that contradicts our suspicion (Kunda, 1990; Nickerson, 1998). Yet, very little is known about what sparks these initial suspicions that someone has engaged in immoral—or even criminal—wrongdoing.

A likely trigger of suspicion is when someone violates our expectations for behavior in the given context. The distrust people have toward others who violate their expectations for normative or appropriate behavior (Levine et al., 2000) might trigger suspicion that they are hiding something. The emotion that people do—or do not—express is often a salient cue to infer that someone is hiding wrongdoing (Bean et al., 2024; Heath, 2009). We focused on the impact of violating expectations for ostensibly "appropriate" levels of emotion—predicting that in contexts where high emotionality is expected, low levels of emotion would trigger suspicion.

Further, we tested a novel theory that failing to express the expected level of emotion could shape moral inferences that others draw about the expressor. If a moral violation occurs, whether someone expresses the emotion that we expect for the context might determine whether we morally typecast them as capable of being a moral perpetrator or a moral victim of the violation. In a context where high emotionality would be expected, we predicted that failing to express emotion might lead people to morally typecast the caller (Gray & Wegner, 2009) as more capable of being a moral perpetrator and less capable of being a moral victim.

We investigated this phenomenon in the high-stakes real-world domain of criminal investigations—a context with devastating consequences for an innocent person's life and liberty. If a police officer becomes suspicious of a witness to a crime, this spark of suspicion is likely to set a series of forensic confirmation biases into motion that can lead to wrongful convictions (Kassin et al., 2013; Scherr et al., 2020). A wealth of anecdotal evidence about wrongful convictions suggests that detectives often become suspicious of innocent people when they violate their expectations for appropriate levels of emotion when reporting a violent crime (Heath, 2009).

In five studies, we investigated whether laypeople and police officers perceive a person calling 911 to report a violent crime as suspicious and worthy of further investigation when they express an unexpectedly low level of emotion. We also tested a psychological explanation for *why* low levels of emotion on a 911 call reporting a violent crime triggers suspicion: because it violates expectations, leading listeners to morally typecast the caller as more capable of being a perpetrator and less capable of being a victim.

The Social Function of Emotion Expression

Social functional theories of emotion describe how people draw inferences about others from the emotions they express (Hareli & Hess, 2010; Keltner & Haidt, 1999; Van Kleef et al., 2011) and the extent to which they seem appropriate for the context (Van Kleef et al., 2011). Perceptions of emotion appropriateness can also be based on a *lack* of emotion (Warner & Shields, 2009). For example, anecdotally, many wrongfully convicted people seemed suspicious to detectives who found their level of emotionality inappropriately low (Heath, 2009)—a misguided emotion expectation that can trigger a guilt-presumptive investigation. We propose a theoretically informed explanation for *why* failing to express the appropriate level of emotion might trigger suspicion: The violation of expectations for the context will shape moral inferences about the caller as someone who has greater capacity to be a perpetrator rather than a victim.

Violation of Expectations

First, we predicted that people would become suspicious of low levels of emotion because it violates expectations for the context of calling 911 to report a violent crime. People commonly assume that individuals "should" react to trauma, such as witnessing a violent crime, with high levels of affect (Heath, 2009). Yet, people do not always express emotions in predictable ways. Expectations notwithstanding, emotions do not leave reliable "fingerprints" (Barrett, 2017). Individuals differ in behavioral responses to tragedy, with some people appearing "cool and collected" rather than grief-stricken (Tyhurst, 1951; Wortman & Boerner, 2012; Wortman & Silver, 1989, 2001). Despite common beliefs that victims of violence will show extreme distress when reporting their trauma, victims often report their experiences with flat emotion (for review, see Salerno, 2021). As a result, crime victims who fail to express high levels of emotion are often perceived to be less credible (e.g., Ask & Landström, 2010; Landström et al., 2015; for a review, see Salerno, 2021). We predicted that if 911 callers do not sound sufficiently

emotional when reporting a violent crime, they would violate the listeners' expectations. We also varied two factors that might shape people's emotion expectations: the caller's gender and relationship with the victim.

Gender

We predicted that people would expect greater levels of emotion from female callers than male callers as a result of well-documented gender stereotypes that women are more emotional and express sadness and fear more than do men (Fabes & Martin, 1991; Salerno et al., 2019). Consistent with this stereotype, we predicted that (a) for female callers, expressing low emotion would violate stereotype-driven expectations for women more than moderate or high emotion, which would trigger greater suspicion, but that (b) for male callers, expressing *high* emotion would violate stereotype-driven expectations more than moderate emotion, which would trigger greater suspicion.

Relationship With the Victim

We predicted that people would expect more emotion from those reporting that a loved one was harmed relative to a stranger (Heath & Grannemann, 2014, 2015). Consistent with these emotion expectations, we predicted that (a) for callers reporting that a loved one was hurt, those expressing low emotion would violate these relationship-based expectations more than moderate or high emotion, which would trigger greater suspicion, but that (b) for callers reporting a stranger was hurt, expressing *high* emotion would violate these relationship-based expectations more than moderate emotion, which would trigger more suspicion.

Moral Typecasting

Our major theoretical goal was to propose a novel psychological explanation for why violating emotion expectations would trigger suspicion. Specifically, we hypothesized that failing to express the expected level of emotion would trigger suspicion by shaping the moral inferences that people make about the caller. The theory of dyadic morality suggests that when an immoral act occurs, people naturally engage in "moral typecasting"—the act of classifying people involved as either a wrongdoing "moral agent" or a suffering "moral patient" (Gray et al., 2012; Gray & Wegner, 2009). Moral agency refers to the capacity to do right and wrong, such as committing an immoral act (i.e., a perpetrator of a violent crime). Moral patiency refers to the capacity to be the *target* of right and wrong, such as a vulnerable victim of an immoral act (i.e., a victim of a violent crime). Even when someone witnesses a violation that is victimless, the moral dyad is psychologically incomplete, and they will complete the dyad by perceiving someone or something as a victim. For example, if someone perceives consensual homosexual acts as immoral, they might complete this victimless dyad by arguing that the act will harm children or society in general (Gray et al., 2014).

If someone calls 911 to report that someone was hurt as a result of a violent crime, police will certainly classify the victim as a moral patient and search for a perpetrator to complete that moral dyad. When 911 callers are present at the scene of a crime, police officers will inevitably consider whether they could be the perpetrator—even in the absence of evidence to that effect. Police officers are literally put in the position of professionally morally typecasting the caller: Are they also a victim of this harm, or might a hunch begin to form that they are capable of having perpetrated the crime they have reported?

It is possible that, in this context, violation of expectations and moral typecasting might be related. People might have the default assumption that people calling 911 for help are victims (even if just emotionally) and expect them to be upset. Instead, hearing someone they expected to be a vulnerable victim be calm when reporting a violent crime might violate expectations, which in turn might trigger the hunch that perhaps they are capable of being the perpetrator and less capable of being the victim, both of which might generate suspicion. That is, the degree to which callers' emotion level violates expectations might be positively associated with the degree to which observers think they are capable of being a perpetrator and negatively associated with the degree to which observers think they are capable of being a victim—which both might be related to greater suspicion. This suggests a serial model: Low (vs. at least moderate) emotion would violate expectations more, which would be related to perceptions of greater moral agency and less moral patiency, which in turn would be related to greater suspicion.

However, moral typecasting theory could also suggest a parallel model that does not necessitate the violation of expectations, wherein caller emotion has a direct impact on moral typecasting. The caller's emotionality might directly influence the degree to which people see them as a likely victim (buffering them against suspicion) or a likely perpetrator (triggering suspicion). Why? Portraying oneself as a victim communicates moral patiency and reduces blame (Gray & Wegner 2011a). If the default assumption is that people calling 911 for help are victims, hearing the caller upset when reporting a violent crime might confirm that the caller is vulnerable and therefore more capable of victimhood (i.e., moral patiency) and less capable of hurting someone (i.e., moral agency)—potentially buffering them against suspicion. Conversely, hearing a caller who is calm when reporting a violent crime might make them seem *less capable* of vulnerability and victimhood and instead more capable of hurting someone. Much of the research supporting moral typecasting theory manipulates moral agency or moral patiency and tests the effect on measures of perceived capacity for feeling (Gray & Wegner 2011b; Shepherd et al., 2019). We test the reverse by manipulating emotion expression in response to a traumatic moral violation to see if it influences perceptions of moral agency and patiency.

Thus, we tested a serial model predicting that hearing low (vs. at least moderate) emotionality on a 911 call would increase violation of expectations and, in turn, moral typecasting of the caller as less capable of being a vulnerable moral patient and more capable of being a perpetrating moral agent—ultimately triggering greater suspicion that they were involved in the crime that they are reporting. We also test an alternative parallel model in which emotion level directly impacts moral typecasting and, in turn, suspicion—above and beyond what is explained by violation of expectations (i.e., controlling for violation of expectations).

High-Stakes Domain: Crime Witnesses' Emotion Expression and Wrongful Convictions

Wrongful convictions are growing at an alarming rate. Since its founding in 2012, the National Registry of Exonerations has archived over 3,500 exonerations in the United States alone (*National Registry of Exonerations*, n.d.) as of May 2024. Although the

psychological process of confirmation bias in criminal investigations leading to false convictions is well-established, we know very little about how and why an innocent person *first* becomes a suspect—this original misclassification error is the "first and the most consequential error police will make" and the "least studied and thus least well-understood" (Leo & Drizin, 2010, p. 13; but see Lowrey-Kinberg et al., 2019).

In numerous wrongful convictions, witnesses have been implicated for appearing to express an "inappropriate" level of emotion (Heath, 2009). To cite just a few (of many) examples: 14-year-old Michael Crowe was falsely accused of his sister's murder after detectives said he reacted to her death with too little emotion (Kassin & Gudjonsson, 2004), 17-year-old Marty Tankleff was induced to falsely confess to killing his parents by a detective who said he was "too calm" (Firstman & Salpeter, 2008), and 41-year-old Gary Gauger was also induced to falsely confess to killing his parents when detectives determined that his voice on the 911 call was too flat (Shapiro, 1998). High levels of emotionality may also be seen as inappropriate—as when police targeted 16-year-old Jeffrey Deskovic because he seemed "overly distraught" after the murder of a high school classmate (Santos, 2006). In Italy, Amanda Knox was targeted for the murder of her roommate, in part, because she was said to be overly emotional at one moment and not emotional enough at another (Kassin, 2022b). In light of these and similar anecdotes, we sought to provide experimental evidence for the proposition that people might become suspicious of an innocent person who violates their expectations for what is an appropriate level of emotion during a highly traumatic event.

Research Overview

Five experiments tested the impact of a crime witness expressing emotion at the very beginning of a criminal investigation before anyone could be suspected: when they call to report a violent crime in real (Studies 1, 4) and simulated (Studies 2–3, Supplemental Study S1) 911 calls. We tested a theoretical explanation for *why* low emotion might trigger suspicion in this context: violation of expectations and moral typecasting of the caller as more capable of being a perpetrator and less capable of being a victim (Studies 2–3, Supplemental Study S1). We also tested the impact of factors likely to shape emotion expectations: the caller's gender (Studies 1–4, S1) and relationship to the victim (Studies 1, 4, S1). Further testing the boundaries of these processes, we sampled laypeople (Studies 1–2, 4, S1) and police officers (Study 3). In addition to advancing basic theory on the psychological impact of emotion expression, we respond to a paucity of research on moral typecasting in real-world settings (Shepherd et al., 2019) by examining the problem in a highly consequential venue involving altruistic citizens who called 911 for help only to become a target of police suspicion.

We hypothesized that expressing low emotion in a 911 call would be more suspicious than at least moderate emotion indirectly through (a) violating expectations and (b) morally typecasting the caller more as a moral agent and less as a moral patient. We also hypothesized interactions, such that the impact of emotion expression on suspicion would depend on moderators that might shape expectations for emotionality, including the caller's gender and relationship with the victim.

Transparency and Openness

Across all studies, we report how we determined our sample size, all data exclusions, all manipulations and measures. All data, analysis code and research materials are accessible on the Open Science Framework (OSF): https://osf.io/yat52/?view_only=8154999c3be5483ba061a9018b39756f. Study 1 was not preregistered. Study 2's design, hypotheses, and analyses were preregistered (available on the OSF at https://osf.io/pquzm/?view_only=5c71a52caf7e405fa07b9f6116848981). Study 3 was not preregistered but followed the Study 2 preregistration for hypotheses and analysis. Study 4's design and analysis were also planned before analysis on the OSF at https://osf.io/w38tm/?view_only=66b677a1736146258c6aed431fcf0a8a. Deviations from the preregistration are noted where relevant in the study-specific method and results sections. In each study, we collected additional exploratory measures of other behaviors (beyond emotionality) that might trigger suspicion about 911 callers, which are reported in another article (Bean et al., 2024) and detailed in the Supplemental Materials.

Study 1

Participants listened to a brief, real 911 call that included (a) a male or female caller (b) expressing low, moderate, or high levels of emotion while reporting that (c) their parents or a stranger had been shot. They reported how suspicious the caller was and their level of support for treating the caller as a suspect.

Method

Participants and Design

Participants were 1,119 Amazon Mechanical Turk online workers. We excluded 176 ($n = 16\%$) for failing a caller gender manipulation check. The remaining 943 participants were 47% female; $M_{age} = 36$, $SD_{age} = 11$; and 77% White, 10% Black, 6.5% Hispanic/Latinx, 5% Asian, 1% other. These participants were randomly assigned to one of 12 cells in a 2 (caller gender: male, female) × 3 (emotion expression: low, moderate, high) × 2 (relationship: parents, stranger) between-subjects design. Our sample size was determined by following the recommendations of more than 50 participants per cell that were current at the time (Simmons et al., 2013).

Materials

Stimuli. We found six 911 recordings online in which someone reported that there had been a shooting but were ambiguous regarding the caller's relationship to the victim(s). All audio clips were between 13 and 21 s long and are available on our OSF project page at https://osf.io/yat52/?view_only=8154999c3be5483ba061a9018b39756f. Of the six calls, three included female callers (expressing low, moderate, or high emotion), and three included male callers (expressing low, moderate, or high emotion).

A pilot study with 59 undergraduates confirmed that the emotion manipulation successfully impacted how much emotion participants reported the caller as having expressed, $F(2, 53) = 48.56$, $p < .001$, $\eta_p^2 = .65$. Participants were asked "To what degree did the caller express emotion?" and responded on a 5-point scale (1 = *not at all*, 5 = *an extreme amount*). Bonferroni post hoc tests confirmed that participants who listened to one of the calls classified as "low

emotion" rated the caller as having expressed significantly less emotion ($M = 1.43$, $SD = 0.51$) than those who listened to "moderate" emotion calls ($M = 3.18$, $SD = 0.96$), $p < .0001$, $d = 2.35$. In turn, they rated the moderate emotion callers as having expressed significantly less emotion than those who listened to a "high emotion" call ($M = 3.86$, $SD = 0.77$), $p = .041$, $d = .78$. Further, the caller's gender did not have a significant main effect, $F(1, 53) = .97$, $p = .97$, $d = .26$, or an interactive effect, $F(2, 53) = 0.77$, $p = .47$, $\eta_p^2 = .03$, on perceived emotionality of the caller. Thus, the effect sizes were large, and the magnitude of the emotion variable effect did not differ for male and female callers.

Relationship Manipulation. At the outset, participants read that they would "listen to a short clip from an actual 911 call in which the caller reports that [their parents/someone they didn't know] had been shot." Immediately after this description of the task, participants responded to a set of questions that they had to get correct before advancing. This included the following: "Please report below what the relationship was between the 911 caller and the victim," (the caller's coworker, parents, a stranger, or this information was not given). They were also asked to report on what they were being asked to do (the correct answer being *listen to 911 audio and answer questions*) and what happened to the victim (the correct answer being *they were shot*). Once participants selected the correct response for all attention checks, they were able to move on to being randomly assigned to one of the 911 calls and the survey.

Measures

All measures are available in Appendix A.

Participants responded to a five-item scale indicating the degree to which they perceived the caller to be suspicious, acting sincere (Reverse Scored [RS]), acting normal (RS), hiding something, and playing a role in the crime on a 7-point response scale ($\alpha = .90$).

Participants responded to a two-item scale indicating their support for police expending resources to investigate the caller as a suspect. These items are meant to gauge participants' willingness to translate their suspicion into concrete consequences. Items were "The police should treat the caller as a suspect" and "The police should use their resources to investigate the caller's involvement in the crime further" and rated on a 7-point scale ranging from *strongly disagree* to *strongly agree* (Spearman–Brown coefficient = .87).

At the end of the survey, participants completed manipulation checks. They were again asked to report the caller's relationship to the victim. Participants chose from six response choices: parents, brother, coworker, neighbor, the victim was someone the caller didn't know, or I was not given this information. Participants reported the degree to which the caller expressed emotion on a 5-point scale from *not at all* to *an extreme amount*. Participants reported the gender of the caller from the recording they heard. The response choices were male, female, or unsure.

Procedure

Participants were told that the study was investigating perceptions of people who call 911 and that they would be randomly assigned to listen to one of a large set of real 911 calls. We told participants that some of the callers might have been involved in the crime they were reporting. We also told them that the calls were brief, so they may not feel that they had all the information they needed but that their

gut impressions were still informative. Participants were randomly assigned to read that the caller was reporting that their parents or a stranger had been shot. They had to correctly answer three questions before listening to the call to ensure that they encoded the caller's relationship to the victim and two filler questions from the task description. Next, they were presented with the 911 audio file with one of the six calls and were reminded to turn up their volume. They were told to listen to the call three times—and that they would not be able to advance to the next page until at least 60 s had elapsed. The audio file remained available to them in a separate tab to revisit throughout the survey. We also reposted the audio file they were assigned to hear within each block of measures. They completed suspicion measures, judgments of whether the caller should be treated as a suspect, manipulation checks, and demographic information. Participants were compensated $1.50 in exchange for their participation in the study.

Results

On average, participants rated the callers as approximately around the midpoint on the Suspicion Scale ($M = 3.13$, $SD = 1.41$) and above the midpoint on support for treating them as a suspect ($M = 4.15$, $SD = 1.67$). These dependent measures were strongly correlated but not redundant, $r = .70$, $p < .001$. For each dependent measure, we conducted a 2 (caller gender: male, female) × 3 (caller emotion: low, moderate, high) × 2 (relationship to caller: stranger, parents) between-subjects analyses of variance (ANOVAs; See Table 1 for all ANOVA results).

Emotion Manipulation Check

Our emotion manipulation was again successful. There was a significant main effect of emotion condition on perceived emotion (Table 1). Bonferroni post hoc tests revealed that participants rated low-emotion callers as expressing significantly less emotion ($M = 1.68$, $SD = 0.80$) than moderate-emotion callers ($M = 3.59$, $SD = 1.06$), $M_{\text{diff}} = -1.91$, $p < .0001$, $d = 2.03$; in turn, they rated moderate-emotion callers as having expressed significantly less emotion than high-emotion callers ($M = 3.99$, $SD = 0.99$), $M_{\text{diff}} = -0.40$, $p < .0001$, $d = .39$. In contrast to the pilot study, the strength of the manipulation depended on the caller's gender. Although the emotion manipulation was significant for both male and female callers (Table 1), the effect was somewhat stronger for female callers (female: $\eta_p^2 = .48$; male = $\eta_p^2 = .25$)—a limitation we address in follow-up studies. Examination of the means (Figure 1) suggests that the difference between low and moderate emotion was perceived to be large for both male and female callers—but the difference between moderate and high emotion was smaller for female callers than for male callers.

Suspicion Scale

Consistent with hypotheses, there was a significant main effect of emotion expression (Table 1). Bonferroni post hoc tests confirmed our hypothesis that participants found low emotion ($M = 4.07$, $SD = 1.28$) significantly more suspicious than moderate emotion ($M = 2.51$, $SD = 1.12$), $M_{\text{diff}} = 1.56$, $p < .0001$, $d = 1.31$, and high emotion ($M = 2.71$, $SD = 1.24$), $M_{\text{diff}} = 1.36$, $p < .0001$, $d = 1.08$.

Table 1

ANOVA Table for the Effects of Caller Gender, Caller Emotion Expression, and Victim Relationship to the Caller on Dependent Variables, Study 1

Dependent variable	Independent variable	df	F	p	Effect size
Perceived emotion (MC)	Caller gender	1	21.62	<.001	$d = .33$
	Relationship to the caller	1	9.27	.002	$d = .07$
	Caller emotion	2	548.53	<.001	$\eta_p^2 = .54$
	Caller Gender × Relationship	1	0.05	.82	$\eta_p^2 < .001$
	Caller Gender × Caller Emotion	1	33.38	<.001	$\eta_p^2 = .07$
	Male caller: Emotion simple effect	1	158.17	<.001	$\eta_p^2 = .25$
	Female caller: Emotion simple effect	1	435.60	<.001	$\eta_p^2 = .48$
	Caller Emotion × Relationship	1	2.92	.054	$\eta_p^2 < .01$
	Gender × Emotion × Relationship	1	0.07	.93	$\eta_p^2 < .001$
	Error	931			
Suspicion Scale	Caller gender	1	3.15	.08	$d = .10$
	Relationship to the caller	1	7.00	<.001	$d = .31$
	Caller emotion	2	176.27	<.001	$\eta_p^2 = .28$
	Caller Gender × Relationship	1	0.02	.90	$\eta_p^2 < .001$
	Caller Gender × Caller Emotion	1	27.15	<.001	$\eta_p^2 = .06$
	Male caller: Emotion simple effect	1	60.67	<.001	$\eta_p^2 = .12$
	Female caller: Emotion simple effect	1	151.46	<.001	$\eta_p^2 = .25$
	Caller Emotion × Relationship	1	3.88	.02	$\eta_p^2 = .01$
	Stranger: Emotion simple effect	1	68.60	<.001	$\eta_p^2 = .13$
	Parents: Emotion simple effect	1	109.34	<.001	$\eta_p^2 = .19$
	Gender × Emotion × Relationship	1	0.15	.86	$\eta_p^2 < .001$
	Error	930			
Treat as suspect	Caller gender	1	0.92	.34	$d = .03$
	Relationship to the caller	1	22.80	<.001	$d = .28$
	Caller emotion	2	57.17	<.001	$\eta_p^2 = .11$
	Caller Gender × Relationship	1	0.77	.38	$\eta_p^2 = .001$
	Caller Gender × Caller Emotion	1	16.68	<.001	$\eta_p^2 = .04$
	Male caller: Emotion simple effect	1	29.28	<.001	$\eta_p^2 = .06$
	Female caller: Emotion simple effect	1	47.95	<.001	$\eta_p^2 = .09$
	Caller Emotion × Relationship	1	1.16	.32	$\eta_p^2 = .002$
	Gender × Emotion × Relationship	1	0.07	.94	$\eta_p^2 < .001$
	Error	930			

Note. Following Lakens (2013) recommendations, we calculated Cohen's *d* values for all two-group comparisons and partial eta² (η_p^2) values for all effects that included more than two groups. ANOVA = analysis of variance; MC = Manipulation Check.

Moderate and high emotion did not differ, $M_{diff} = -0.20$, $p = .13$, $d = .17$.

There was also a significant main effect of relationship, such that people were more suspicious of callers reporting that their parents had been shot ($M = 3.35$, $SD = 1.45$) relative to when the caller reported that it was a stranger had been shot ($M = 2.93$, $SD = 1.33$). The main effect of caller gender was not significant.

The main effects were qualified, however, by the hypothesized caller Emotion × Gender interaction (Table 1, Figure 2a). Specifically, the simple effect of emotion was significant for both male and female callers (Table 1), but the patterns were different. As hypothesized, women who violated gender stereotypes and expressed low emotion were significantly more suspicious than women who expressed moderate, $M_{diff} = 1.81$, $d = 1.47$, $p < .001$, 95% confidence interval (CI) [1.56, 2.05], and high levels of emotion, $M_{diff} = 1.97$, $d = 1.63$, $p < .001$, 95% CI [1.72, 2.22]. Women who expressed moderate and high emotion did not differ, $M_{diff} = 0.17$, $d = .15$, $p = .19$, 95% CI [−.08, .41]. Similarly, men who expressed low emotion were significantly more suspicious than men who expressed moderate emotion, $M_{diff} = 1.38$, $d = 1.16$, $p < .001$, 95% CI [1.14, 1.63]. However, as predicted, men who violated gender stereotypes by expressing a *high* level of emotion were significantly *more* suspicious than those who expressed moderate emotion, $M_{diff} = 0.83$, $d = .74$, $p < .001$, 95% CI [.54, 1.13]. In summary, people were more

suspicious of women who violated gender stereotypes by expressing low emotion and men who expressed too little *and* who violated gender stereotypes by expressing too much emotion, relative to a moderate level.

Figure 1

Mean (Standard Error) of Perceived Emotion Expression as a Function of the Emotion Expression Manipulation and Caller Gender, Study 1

□ Low Emotion   ■ Moderate Emotion   ■ High Emotion
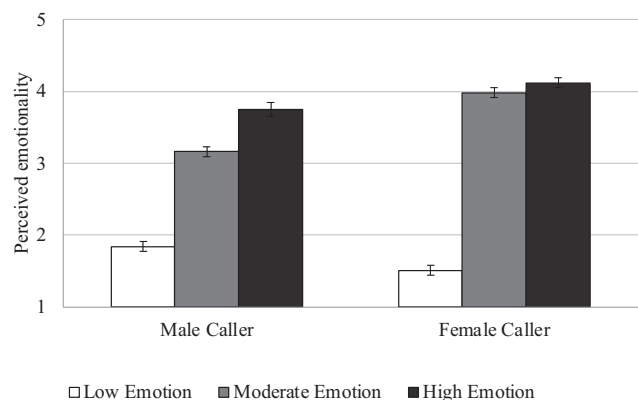
Figure 2

Mean (Standard Error) of Suspicion as a Function of Caller Emotion Expression and (a) Caller Gender and (b) Relationship to the Victim, Study 1
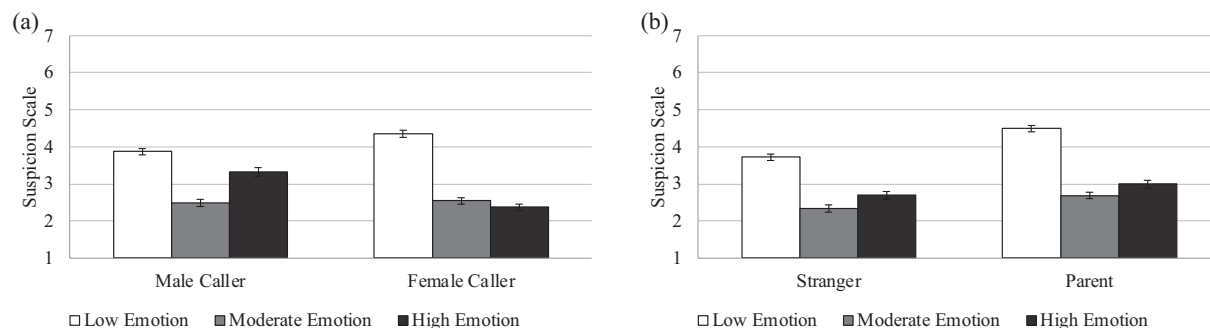
The main effects were also qualified by the hypothesized Caller Emotion × Relationship With the Victim interaction, (Table 1, Figure 2b). As hypothesized, when calling about parents being shot, 911 callers who violated expectations by expressing low emotion were significantly more suspicious than those expressing moderate, $M_{diff} = 1.80$, $d = 1.56$, $p < .001$, 95% CI [1.55, 2.05], or high emotion, $M_{diff} = 1.50$, $d = 1.32$, $p < .001$, 95% CI [1.22, 1.77]. Contrary to hypotheses, high emotion was significantly more suspicious than moderate emotion, $M_{diff} = 0.31$, $d = .12$, $p = .03$, 95% CI [.04, .58]. When the 911 caller reported that a *stranger* had been shot, the pattern was the same. Those who expressed low emotion were again significantly more suspicious than those who expressed a moderate level of emotion, $M_{diff} = 1.39$, $d = 1.18$, $p < .001$, 95% CI [1.15, 1.63], or high emotion, $M_{diff} = 1.03$, $d = .94$, $p < .001$, 95% CI [.76, 1.29]. Expressing high emotion was again significantly *more* suspicious than moderate emotion, $M_{diff} = 0.36$, $d = .49$, $p = .01$, 95% CI [.09, .63]. Thus, the interaction is driven by the pattern of the effect of emotion being the same, but it is stronger when parents were shot because suspicion was particularly high in the condition that paired parents being shot with low levels of emotion expression (Figure 2b).

Treat Caller as a Suspect

We replicated our emotion effect and Emotion × Gender interaction when we analyzed our alternative outcome (i.e., police should treat caller as a suspect scale). We did not, however, replicate the Emotion × Relationship interaction. We report the overall ANOVA results in Table 1, but we provide the detailed follow-up analyses in Supplemental Materials.

Discussion

We found support for our hypothesis that people found someone who expresses low (vs. moderate or high) levels of emotion on a 911 call to report a violent crime to be more suspicious and worthy of investigation. Further, people were particularly suspicious when emotion levels did not match expectations, such as when (a) low emotion was expressed in general but especially when high emotion was expected, such as in the case of female callers or when the caller was reporting that their parents had been shot, or when

(b) high emotion was expressed when lower emotion would be expected, such as in the case of male callers.

Study 2

We attempted to replicate our findings after addressing methodological limitations of Study 1 by using a set of tightly controlled and scripted 911 calls and including direct measurement of violation of expectations and moral typecasting.

Method

We made several changes to the methodology to address limitations of Study 1, including a lack of stimulus sampling, potential confounds resulting from a different call for each experimental cell, and the emotion manipulation being stronger for the female callers. First, we hired professional voice actors to each record the same 911 call script three times—with low, moderate, and high emotion. This enabled us to manipulate the caller's emotion level while holding the script of the call constant. Second, we conducted stimulus sampling by using several male and several female voice actors to better isolate the gender effect. Third, we directly measured our proposed mediators: violation of expectations and moral typecasting. Because participants were always more suspicious of callers who reported that their parents had been shot compared to a stranger, we dropped the relationship manipulation (but return to it in Supplemental Study S1).

We predicted that we would replicate our Study 1 findings. We also tested a serial mediation model in which low (vs. at least moderate) emotion would violate expectations more, which would be related to perceptions of greater moral agency and less moral patiency, which in turn would both predict greater suspicion. We also tested an alternative parallel mediator model in which (a) the emotional victim effect would generalize to a context where the target's perpetrator/victim status is ambiguous, such that expressing low (vs. moderate or high) emotion on a 911 call would result in greater violation of expectations, which would be associated with greater suspicion (i.e., a significant indirect effect), and (b) expressing low (vs. moderate or high) emotion on a 911 call would directly increase perceptions of moral agency and decrease perceptions of moral patiency, which would be associated with greater suspicion—indirect effects that we expected would be

significant above and beyond what is explained by violation of expectations.[1]

Participants and Design

We recruited 1,553 Mturk online workers. We excluded a total of 929 (60%) participants for failing a manipulation check at the end of the study regarding the gender of the caller they heard ($n = 410$, 26%) and/or a relationship attention check ($n = 467$, 30%). Those who said that they thought the call was fake were also excluded ($n = 50$, 3%). These high exclusion rates were, in large part, due to a programming error.[2] We report analyses that exclude these participants who got the gender and/or relationship checks at the end of the study incorrect. However, we also verified that we find the same pattern of results when we rerun the analysis while including participants who failed manipulation checks and/or were affected by the programming error. That is, the pattern of results did not differ when we used our preregistered exclusion criteria and excluded participants who failed our manipulation/attention checks versus when we analyzed the full sample.

The remaining 624 participants were 45% women; $M_{age} = 36$, $SD_{age} = 11$; and 71.5% White, 15.4% Black, 6% Hispanic/Latinx, 5% Asian, and 2% other. These participants were assigned to one of six cells of a 2 (caller gender: male, female) × 3 (emotion expression: low, moderate, high) between-subjects design.

Our sample was determined by a power analysis detailed in the preregistration for this study, which prescribed 976 participants. Due to the programming error described above, we were below this target—but, as noted, we found the same pattern of results in an analysis that included the participants who failed manipulation/ attention checks (likely due to the programming error) and was adequately powered.

Materials

We created a scripted dialogue between a 911 operator and a caller reporting a shooting. We listened to many real 911 calls and replicated specific phrases to create a script that was ambiguous as to the caller's involvement in the reported crime. The scripted calls were longer than the calls used in Study 1. See Appendix B for the call script.

In our scripted dialogue, the caller reports to 911 that their spouse had been shot outside their house and needed medical aid—leaving it ambiguous as to whether the spouse was alive or dead. To create the calls, we hired 14 professional voice actors to perform the same 911 call script (seven women, seven men). We also created a single female operator recording to combine with the voice actors' recordings. The 911 caller and operator recordings were combined using Adobe's Audition software. To enhance authenticity, we added a dial tone to the beginning of each call and a telephone noise filter to the entire clip of each call. All audio clips were between 73 and 101 s in length.

Pilot Testing. We pilot tested these calls with 332 participants from Amazon's Mechanical Turk and asked participants to listen to one of the 14 moderate emotion calls and report their open-ended initial impression. We eliminated calls in which 25% or more of participants spontaneously mentioned that they thought the call sounded fake. This resulted in five female actors and four male actors.

Gender Manipulation. We used a stimulus sampling approach for our gender manipulation, such that we randomly assigned each participant to hear one of the four male callers or one of the five female callers. The female callers reported that their husband had been shot, and male callers reported that their wife had been shot.

Emotion Manipulation. Each of the nine voice actors performed the 911 call script three times, with low, moderate, and high expressions of emotion (totaling 27 recordings). To provide guidance for how to modulate emotions, we first hired a single actor and coached this person through multiple feedback cycles on how to express the three distinct emotion levels. These recordings were given to the other voice actors as a reference.

Measures

We included the questions from Study 1 and added measures of the following proposed mediators. All measures are included in Appendix A.

Participants responded to a Violation of Expectations Scale comprising four items presented in a random order that assessed their agreement on whether the caller violated their expectations of how someone should behave in the caller's situation (e.g., "I was genuinely surprised about how the caller acted").[3] Each item was rated on a 6-point scale from *strongly disagree* to *strongly agree* ($\alpha = .83$).

Participants then completed measures of moral patiency and moral agency presented in a random order. Participants responded to five items indicating the degree to which they saw the caller as capable of being a moral agent and having greater potential to commit immoral acts (e.g., "The caller seems like someone who could be capable of doing immoral things"). Each item was rated on a 6-point scale from *strongly disagree* to *strongly agree*. We conducted a confirmatory factor analysis. The reliability was improved by deleting two items, which resulted in a three-item measure for the Moral Agency Scale ($\alpha = .75$).

Participants also responded to five items indicating the degree to which they saw the caller as capable of being a moral patient and

---

[1] Note that these two theoretical models were a deviation from our preregistration. Originally, we preregistered (a) measuring violation of expectations but did not specify hypotheses about that variable and (b) exploratory analyses to assess whether moral agency and moral patiency were simultaneous mediators of the effect of emotion expression on suspicion.

[2] We had an unfortunately high exclusion rate in this study (57%) due to a programming error that was found after data collection began and was corrected only midway through the data collection process. Before hearing the 911 call, participants completed an attention check in which they needed to correctly identify the relationship between the caller and the victim as well as answer two filler questions. If participants "failed" this attention check, they could submit a new answer, but they could not proceed until they "passed" this check. Unfortunately, there was a logic error in one condition that erroneously required participants to give an incorrect answer about the caller's relationship to the victim to proceed to the survey. This meant that they were forced to give an incorrect answer initially to participate in the study, which likely produced confusion later on about how to answer the manipulation/ attention checks once they got to the end of the study—leading to a large number failing the gender and/or relationship question at the end. We report analyses excluding all of these people who failed the manipulation/attention checks at the end of the study (consistent with our preregistration), but also replicated the pattern of results when we included the full sample.

[3] Note that we preregistered a description of this scale comprising five items, which was an error. We had only four items, which we used in this study (and in Studies 2–3, Supplemental Study S1), which are reported in Appendix A.

having greater potential of being a vulnerable victim of immoral acts (e.g., "The caller seems very vulnerable"). Each item was rated on a 6-point scale from *strongly disagree* to *strongly agree*. The reliability was improved by deleting two of the items, which resulted in a three-item measure for Moral Patiency ($\alpha = .76$).

Participants then completed the same outcome measures as Study 1: the Suspicion Scale ($\alpha = .91$) and the Treat as Suspect Scale (Spearman–Brown coefficient = .88). Participants completed the same emotion and gender manipulation checks from Study 1 to ensure that they correctly identified the caller as a man or as a woman.

Procedure

We again recruited participants from Amazon's Mechanical Turk. Prior to listening, participants read a written description about the nature of the call (caller reported a shooting of his or her spouse) and were allowed to proceed once they answered multiple-choice attention checks confirming their correct understanding. Participants were then randomly assigned to hear a 911 call with the exact same script that was (a) recorded by a woman or man and (b) delivered with either low, moderate, or high emotion. Participants were asked to listen to their assigned call three times before they could advance to the next phase of the study; they were also able to relisten to the call at any point while completing the measures. Next, participants completed the same measures from Study 1 in the following order: suspicion, support for treating the caller as a suspect, manipulation checks, and demographic information. Unique to Study 2, the participants also completed measures for violation of expectations (before suspicion) and moral typecasting (before manipulation checks).

Results

Descriptive information and correlations among measures are reported in Table 2.

For each dependent measure, we conducted a 2 (caller gender: male, female) $\times$ 3 (caller emotion: low, moderate, high) between-subjects ANOVA (See Table 3 for all ANOVA results).

Emotion Manipulation Check

Our new emotion manipulation with simulated calls was successful. There was a main effect of caller emotion condition on perceived emotion (Table 3). Bonferroni post hoc tests revealed that participants rated low-emotion callers as having expressed significantly less emotion ($M = 2.59$, $SD = 1.08$) than moderate-emotion callers ($M =$

Table 2
Descriptive Statistics and Pearson Correlation Matrix for Mediators and Dependent Variables, Study 2

Measure	M (SD)	1	2	3	4	5
1. Suspicion	3.44 (1.57)	—				
2. Treat as suspect	4.84 (1.54)	.69*	—			
3. Violation of expectations	3.21 (1.36)	.72*	.48*	—		
4. Moral patiency	3.76 (1.20)	−.53*	−.32*	−.56*	—	
5. Moral agency	3.30 (1.17)	.67*	.58*	.54*	−.21*	—

* p < .001.

3.58, $SD = 1.03$), $p < .0001$, $d = .94$; in turn, they rated moderate-emotion callers as having expressed significantly less emotion than high-emotion callers ($M = 4.09$, $SD = 0.94$), $p < .0001$, $d = .51$. Unlike in Study 1, the caller's gender had no main or interactive effects on perceived emotion.

Suspicion Scale

There was a significant main effect of emotion expression (Table 3, Figure 3). Bonferroni post hoc tests confirmed our hypothesis and replicated Study 1: Participants found low emotion to be significantly more suspicious than moderate emotion, $p < .0001$, $d = .57$, and high emotion, $p < .0001$, $d = .77$. Moderate and high emotion again did not differ, $p = .22$, $d = .17$. In contrast to Study 1, however, there were no significant gender effects.

Mediators

We replicated the main effect of emotion expression on each of our mediators. Low emotion led to increased violation of expectations, increased moral agency, and decreased moral patiency relative to moderate and high emotion. Further, moderate emotion increased violation of expectations relative to high emotion. For detailed ANOVA results, see Supplemental Materials.

Serial Indirect Effect Models. Because violation of expectations was correlated with both increased moral agency and decreased moral patiency (Table 2), we tested a post hoc (non-pre-registered) model in which violation of expectations and moral typecasting are related. More specifically, we tested whether, when someone calls 911 to report a violent crime with low emotion, they violate expectations for victimhood, which would be related to perceiving the person to be less capable of being a moral patient and more capable of being a moral agent, which in turn would both ultimately predict greater suspicion.

To provide preliminary support for this theory, we used Hayes' PROCESS Model 6 to test two serial models: the indirect effect of expressing low (vs. moderate or high) emotion on suspicion through (a) increased violated expectations and, in turn, increased moral agency and (b) increased violated expectations and, in turn, decreased moral patiency. The serial indirect effect through moral agency, $M_{\text{indirecteffect}} = 0.29$, 95% CI [.22, .38], and through moral patiency, $M_{\text{indirecteffect}} = 0.13$, 95% CI [.08, .20], were significant (Figure 4). Thus, in this context, callers expressing an unexpectedly low level of emotion shaped moral inferences that triggered suspicion. Although our data support this model, we interpret it with caution given that it is a post hoc hypothesis, and chronological causality cannot be established with these correlational data.

Simultaneous Indirect Effect Model. Next, we conducted a parallel model testing whether expressing low emotion generates suspicion indirectly through (a) violating the listeners' expectations for 911 callers reporting a violent crime and (b) triggering moral typecasting of the caller as less capable of being a moral patient and more capable of being a moral agent, relative to those who express at least moderate emotion. More specifically, we tested a simultaneous mediator model with low (vs. moderate or high) levels of emotion, predicting three simultaneous mediators (expectation violation, moral agency, and moral patiency), with caller gender as a potential moderator and suspicion as our dependent variable.

Table 3

ANOVA Table for the Effects of Caller Gender and Caller Emotion Expression on Dependent Variables, Study 2

Dependent variable	Independent variable	df	F	p	Effect size
Perceived emotion (MC)	Caller gender	1	0.08	.78	$d = .03$
	Caller emotion	2	116.29	<.001	$\eta_p^2 = .27$
	Caller Gender × Caller Emotion	1	2.09	.13	$\eta_p^2 = .01$
	Error	618			
Suspicion Scale	Caller gender	1	0.58	.50	$d = .06$
	Caller emotion	2	31.27	<.001	$\eta_p^2 = .09$
	Caller Gender × Caller Emotion	1	0.88	.41	$\eta_p^2 = .003$
	Error	618			
Treat as Suspect Scale	Caller gender	1	3.87	.05	$d = .16$
	Caller emotion	2	12.65	<.001	$\eta_p^2 = .04$
	Caller Gender × Caller Emotion	1	2.06	.13	$\eta_p^2 = .01$
	Error	618			

Note. Following Lakens (2013) recommendations, we calculated Cohen's *d* values for all two-group comparisons and partial eta² (η_p^2) values for all effects that included more than two groups. ANOVA = analysis of variance; MC = Manipulation Check.

As expected based on the lack of gender effects on suspicion, caller gender did not moderate the indirect effect through violation of expectations, index of moderated mediation (IMM) = .06, standard error (*SE*) = .09, 95% CI [−.11, .24]; moral agency IMM = −.05, *SE* = .11, 95% CI [−.26, .15]; or moral patiency, IMM = −.03, *SE* = .07, 95% CI [−.15, .10]. Thus, we dropped gender as a moderator and reran the mediation model (Figure 5).

The model revealed three significant unique indirect effects of callers expressing low (vs. moderate or high) emotion on suspicion through violation of expectations, $M_{\text{indirecteffect}} = 0.51$, 95% CI [.37, .65]; moral agency, $M_{\text{indirecteffect}} = 0.36$, 95% CI [.25, .48]; and moral patiency, $M_{\text{indirecteffect}} = 0.28$, 95% CI [.19, .37]. The path coefficients confirmed our hypothesis: Expressing low (vs. moderate or high) emotion increases violation of expectations and perceptions of moral agency and decreases perceptions of moral patiency. All three mediators are uniquely associated with suspicion, controlling for each other.

Treat as Suspect

Note that we also replicated the main effect of emotion on suspicion and the parallel simultaneous mediation pattern when we analyzed our alternative outcome measure: Treat as Suspect Scale. For detailed analyses, see Supplemental Materials.

Figure 3

Mean (Standard Error) of Suspicion as a Function of Caller Emotion Expression and Caller Gender, Study 2
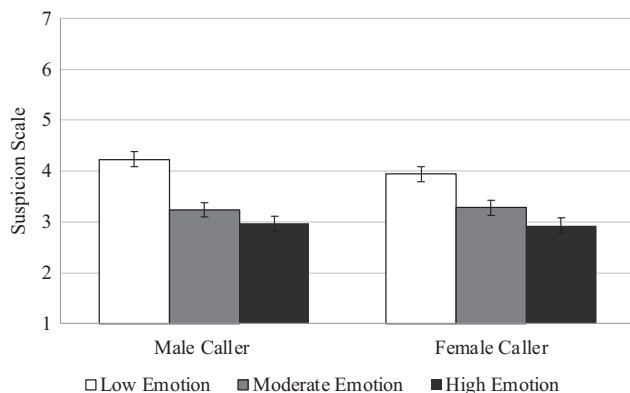
Discussion

We replicated our finding that people found 911 callers who expressed low (vs. moderate or high) emotion to be more suspicious and worthy of investigation. This effect replicated with a sample of calls that contained the same verbatim script, the only difference being the level of emotion expressed. We did not, however, replicate our finding that the effect of callers' emotionality depended on gender (an issue we return to in Study 4 and Supplemental Study S1).

We also identified a novel psychological explanation: Failing to express the expected level of emotion for this context triggered suspicion by shaping the moral inferences that people draw about the expressor. More specifically, 911 callers expressing low emotion violated expectations, which were related to perceiving the caller as more capable of being a moral agent and less capable of being a moral patient, which were both associated with greater suspicion—relative to callers who reported the exact same crime with the exact same words but with a more emotional tone. Our data also supported an alternative model, which revealed that when 911 callers expressed low emotion, this also *directly* caused people to perceive the caller as less capable of vulnerability and victimhood and instead more capable of immoral acts, which both triggered suspicion—above and beyond what was explained by the degree to which low emotion violated their expectations. Further, this model replicated the emotional victim effect in a new context where it is more ambiguous as to whether the target is a perpetrator or victim.

Study 3

We tested our hypotheses among a more consequential sample: police officers, who often are in a position to decide whether to investigate a witness as a suspect. We did not preregister this study, but followed the Study 2 preregistration with the same deviations noted in Study 2 (see Footnote 1).

Method

We recruited 311 police officers from the greater Phoenix area by sending an email solicitation via local police newsletters and listservs. We excluded a total of 13 participants (4%) for failing the gender manipulation check (*n* = 5, 1.6%), failing the relationship

Figure 4

Indirect Serial Effects of Emotion Expression on Suspicion Through Violation of Expectations, Moral Agency, and Moral Patiency, Study 2
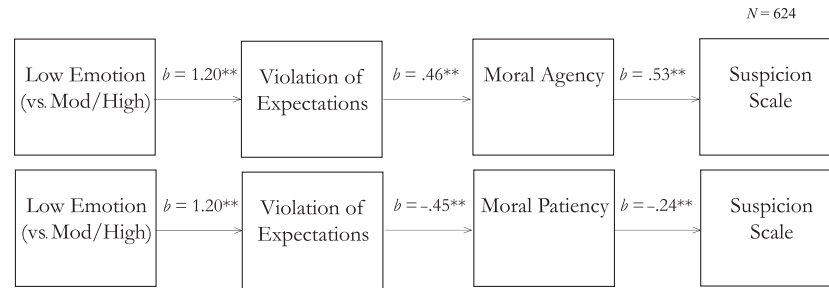
Note. Mod = moderate.
** $p < .001$.

attention check ($n = 3$, 1%), and/or revealing in an open-ended question that they thought the call was fake ($n = 5$, 1.6%). The remaining 298 officers were 21% women; $M_{age} = 40$, $SD_{age} = 10$; and 75% White, 2.4% Black, 15% Hispanic/Latinx, 1% Asian, and 5% other. Regarding job position, 49.5% were patrol officers, 27% were detectives, 14% were sworn supervisors, 2% were civilian employees, and 7.5% selected "other." Regarding years of experience, 5% had less than 1 year, 20% had 2–5 years, 11% had 6–10 years, 29% had 11–15 years, and 35% had 20+ years. Because police officers are a very difficult sample to recruit, our sample size was determined by recruiting as many police officers as possible within what local police departments were willing to do to help us recruit (i.e., they advertised the survey via newsletters and listservs and sent one reminder per notification).

The study followed the same design as Study 2, featuring a 2 (caller gender: male, female) × 3 (emotion expression: low, moderate, high) between-subjects design. Officers were exposed to the same materials, completed the same measures, and followed the same procedure as Study 2.

We again found that our scales had acceptable-to-good reliability: Suspicion Scale ($\alpha = .89$), Treat as Suspect Scale (Spearman–Brown

Figure 5

Indirect Effects of Emotion Expression on Suspicion Through Violation of Expectations, Moral Agency, and Moral Patiency, Study 2

Note. Mod = moderate; CI = confidence interval.
** $p < .001$.

coefficient = .62), Violation of Expectations Scale ($\alpha = .80$), Moral Agency Scale ($\alpha = .64$), and Moral Patiency Scale ($\alpha = .66$). Note that if we deleted the items that improved the reliability of the Moral Agency and Moral Patiency Scales in Study 2, reliability would have decreased in this sample, so we used all 10 of the original items from Study 2 (See Appendix A). The only other difference is that we added questions about the officers' job experience and two open-ended questions about what kind of factors in interviews make them suspicious of witnesses, which are reported in another article (Lawrence et al., 2024) and Supplemental Materials. To ensure a large enough sample, we increased the compensation to a $25 Amazon gift card.

Results

Descriptive information and correlations among measures are reported in Table 4.

Emotion Manipulation Check

Our emotion manipulation was again successful as was demonstrated by a significant main effect on perceived emotion (Table 5). Bonferroni post hoc tests revealed that participants rated low-emotion callers as having expressed significantly less emotion ($M = 1.98$, $SD = 0.73$) than moderate-emotion callers ($M = 2.84$, $SD = 0.93$), $M_{diff} = -0.86$, $p < .0001$, $d = 1.04$, who in turn were rated as having expressed significantly less emotion than high-emotion callers ($M = 3.73$, $SD = 0.91$), $M_{diff} = -0.89$, $p < .0001$, $d = .97$. Although officers perceived male callers to be less emotional ($M = 2.69$, $SD = 1.04$) than female callers ($M = 2.92$, $SD = 1.15$) overall, importantly (and consistent with Study 2) the strength of the manipulation again did not depend on the caller's gender (Table 5).

For each dependent measure, we conducted a 2 (caller gender: male, female) × 3 (caller emotion: low, moderate, high) between-subjects ANOVA (see Table 5 for ANOVA results).

Suspicion Scale

We also replicated the significant main effect of emotion expression on suspicion (Table 5). Bonferroni post hoc tests revealed a slightly different pattern from Study 2: Officers perceived low emotion
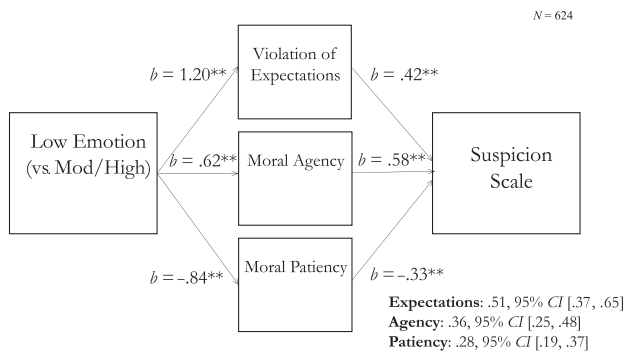
Table 4

Descriptive Statistics and Pearson Correlation Matrix for Mediators and Dependent Variables, Study 3

Measure	M (SD)	1	2	3	4	5
1. Suspicion	4.09 (1.16)	—				
2. Treat as suspect	5.43 (1.13)	.47*	—			
3. Violation of expectations	3.84 (1.17)	.65*	.31*	—		
4. Moral patiency	3.32 (0.85)	−.61*	−.36*	−.62*	—	
5. Moral agency	3.77 (0.83)	.68*	.53*	.55*	−.65*	—

* $p < .001$.

($M = 4.43$, $SD = 0.98$) as similarly suspicious as moderate emotion ($M = 4.09$, $SD = 1.18$), $M_{diff} = 0.35$, $p = .07$, $d = .32$. They found high emotion to be significantly less suspicious ($M = 3.70$, $SD = 1.21$) than low emotion, $M_{diff} = −0.73$, $p < .001$, $d = .67$, and moderate emotion, $M_{diff} = −0.38$, $p = .048$, $d = .32$. Thus, their threshold for how much emotion was needed to decrease suspicion was higher than for laypeople. Officers also exhibited a gender bias: Overall, they found male callers ($M = 4.34$, $SD = 1.13$) more suspicious than female callers ($M = 3.86$, $SD = 1.13$). Consistent with Study 2, caller gender and emotion did not interact in this study (Table 5; See Figure 6).

Mediators

We replicated the main effect of emotion expression on each of our mediators. Low emotion violated expectations more than moderate emotion, which in turn violated expectations more than high emotion. Low and moderate emotion increased perceptions of moral agency relative to high emotion. Finally, low emotion decreased perceptions of moral patiency relative to moderate emotion, which in turn decreased perceptions of moral patiency relative to high emotion. For detailed ANOVA results, see Supplemental Materials.

Serial Indirect Effect. We again tested the serial model from Study 2 in which the effect of expressing low or moderate (vs. high) emotion on a 911 call violates expectations for victimhood more,

which would be associated with perceiving the person to be less capable of being a moral patient and more capable of being a moral agent, which in turn would both ultimately predict increased suspicion. We again used Hayes' PROCESS Model 6 to test the same two serial models from Study 2: the indirect effect of expressing low or moderate (vs. high) emotion on suspicion through (a) increased violated expectations and, in turn, increased moral agency and (b) increased violated expectations and, in turn, decreased moral patiency. The serial indirect effect through moral agency, $M_{indirecteffect} = 0.22$, 95% CI [.15, 31], and through moral patiency, $M_{indirecteffect} = 0.18$, 95% CI [.10, .27], were again both significant (Figure 7).

Simultaneous Mediator Model. Next, we again tested the simultaneous mediator model from Study 2 with low or moderate (vs. high) levels of emotion, predicting three simultaneous mediators (expectation violation, moral agency, and moral patiency), with caller gender as a potential moderator and suspicion as our dependent variable.

Based on the lack of Emotion × Gender interaction effects, caller gender again did not moderate the indirect effect through violation of expectations, IMM = −.09, SE = .09, 95% CI [−.29, .10]; moral agency IMM = .03, SE = .12, 95% CI [−.21, .27]; or moral patiency, IMM = −.07, SE = .06, 95% CI [−.22, .01]. Thus, we again dropped gender as a moderator and reran the model (Figure 8).

The model replicated the three significant unique indirect effects of callers expressing low/moderate (vs. high) emotion on suspicion through violation of expectations, $M_{indirecteffect} = 0.32$, 95% CI [.19, .47]; moral agency, $M_{indirecteffect} = 0.22$, 95% CI [.10, .35]; and moral patiency, $M_{indirecteffect} = 0.14$, 95% CI [.02, .26]. The path coefficients again confirmed our hypothesis: Expressing low/moderate (vs. high) emotion increased violation of expectations and also *directly* increased perceptions of moral agency and decreased perceptions of moral patiency. All three mediators are uniquely associated with suspicion, controlling for each other. In summary, failing to express high levels of emotion increased suspicion indirectly through violating officers' expectations and leading officers to morally typecast the caller as more of a moral agent and less of a moral patient.

Table 5

ANOVA Table for the Effects of Caller Gender and Caller Emotion Expression on Dependent Variables, Study 3 (Police Officers)

Dependent variable	Independent variable	df	F	p	Effect size
Perceived emotion (MC)	Caller gender	1	7.29	.01	$d = .21$
	Caller emotion	2	101.50	<.001	$\eta_p^2 = .41$
	Caller Gender × Caller Emotion	1	0.06	.94	$\eta_p^2 < .001$
	Error	292			
Suspicion Scale	Caller gender	1	14.18	<.001	$d = .42$
	Caller emotion	2	10.93	<.001	$\eta_p^2 = .07$
	Caller Gender × Caller Emotion	1	0.32	.73	$\eta_p^2 = .002$
	Error	292			
Treat as Suspect Scale	Caller gender	1	5.98	.01	$d = .28$
	Caller emotion	2	0.89	.41	$\eta_p^2 = .01$
	Caller Gender × Caller Emotion	1	0.02	.98	$\eta_p^2 < .001$
	Error	292			

Note. Following Lakens (2013) recommendations, we calculated Cohen's d values for all two-group comparisons and partial eta^2 (η_p^2) values for all effects that included more than two groups. ANOVA = analysis of variance; MC = Manipulation Check.

Figure 6

Mean (Standard Error) of Suspicion as a Function of Caller Emotion Expression and Caller Gender, Study 3 (Police Officers)

□ Low Emotion ■ Moderate Emotion ■ High Emotion
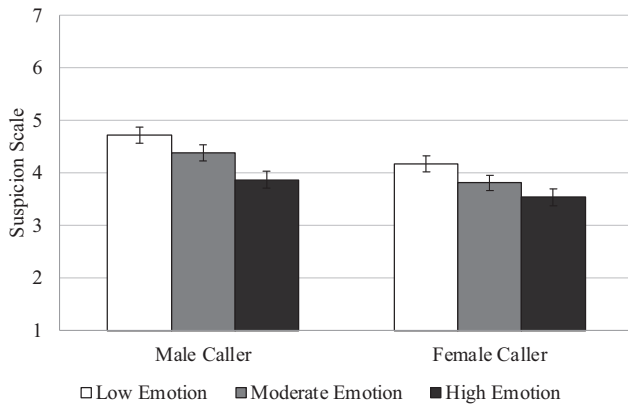
Treat as Suspect

We did not replicate all of our mediation findings on the Treat as Suspect Scale because police officers thought that all of the callers should be investigated as a suspect regardless of gender or emotion level—all of the Treat as Suspect Scale means were very high (for detailed analyses see Supplemental Materials). This makes sense in hindsight given that police routinely investigate the spouse in a crime like this.

Discussion

Police were highly suspicious of not only low but also moderate levels of emotion. Although laypeople required only moderate emotion to be less suspicious of the caller, police needed to see *high* emotion before becoming less suspicious. Although this threshold differed, the effect operated through the same psychological channels: Low or moderate (vs. high) levels of emotion made police more suspicious because it violated their expectations and led them to morally typecast the caller as more capable of committing immoral acts and less capable of being a victim of immoral acts, which both ultimately predicted increased suspicion. Police also exhibited an overall gender bias to find men more suspicious and

worthy of investigation than women—even though all callers described the same crime in the same words.

Supplemental Study S1: Further Investigating Caller Gender

We conducted an additional study to further investigate whether the impact of expressing emotion during 911 calls on suspicion differs for men and women (Supplemental Study S1). In Study 1, men calling 911 were more suspicious when expressing high emotion compared to moderate emotion regardless of when they reported their parents or a stranger was hurt—which was not the case for women calling 911. In contrast, this gender interaction was eliminated in Studies 2–3 when they all called about a spouse. We reasoned that perhaps people expect (and therefore are not suspicious of) greater emotion when a *spouse* is harmed—even for men. We tested whether emotion expression depended on gender in the context of a stranger relationship (like the significant interaction in Study 1), but *not* in the context of a *spouse* (like the null interaction in Studies 2–3). Ultimately, however, we did not find support for this hypothesis and instead replicated the null gender interaction from Studies 2–3. More specifically, low (vs. moderate/high) emotion was again more suspicious because it violated expectations and was associated with moral typecasting of the caller as having greater moral agency and reduced moral patiency. However, this effect of emotion expression again did not depend on gender—regardless of whether the caller reported a stranger or parents were shot (like Study 1 where gender moderated) or a spouse was shot (like Studies 2 and 3 where gender did not moderate). Thus, our proposed explanation that perhaps the gender moderation exists outside of the spousal context was not supported.

In Study 4, we addressed another potential explanation: Perhaps the gender moderation in the real 911 calls used in Study 1 was eliminated by using scripted 911 calls used in Studies 2–3 and Supplemental Study S1. Perhaps there are naturally occurring gender differences in how emotion is expressed that were not captured by our scripted calls.

Study 4

We conducted a conceptual replication with a diverse set of real 911 calls to test whether the robust experimental finding that lower

Figure 7

Indirect Serial Effects of Emotion Expression on Suspicion Through Violation of Expectations, Moral Agency, and Moral Patiency, Study 3 (Police Officers)

Note. Mod = moderate.
** *p* < .001.

Figure 8

Indirect Effects of Emotion Expression on Suspicion Through Violation of Expectations, Moral Agency, and Moral Patency, Study 3 (Police Officers)

N = 298

Expectations: .32, 95% CI [.19, .47]
Agency: .22, 95% CI [.10, .35]
Patiency: .14, 95% CI [.02, .26]
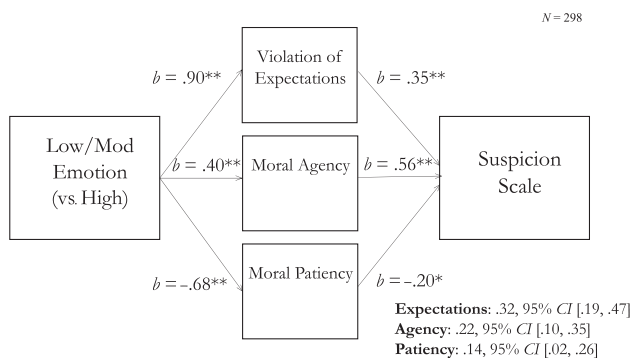
Note. Mod = moderate; CI = confidence interval.
* p < .01. ** p < .001.

emotionality is associated with greater suspicion generalizes to a set of real calls varying in circumstances surrounding the call. This also enables us to test for the gender moderation we found in Study 1 with naturalistic emotion expression.

Participants and Design

We recruited 250 participants from CloudResearch (43.2% women, 53.2% men, 3.6% nonbinary or other; $M_{age} = 38.71$, $SD_{age} = 10.94$; 69% White, 10% Black, 10% Hispanic/Latinx, 8% Asian, 3% other) to listen to eight 911 calls selected from a larger set of 88 real calls in which someone reported a violent incident that resulted in bodily harm or death. CloudResearch participants tend to pass attention checks and provide more meaningful answers than Mturk workers (Douglas et al., 2023). Indeed, although we did include two attention checks, no one failed either of them. We used a within-subjects design with participants' ratings of perceived emotionality of the caller, caller gender, and caller relationship to the victim varying within subjects. Participants were paid $12 for their participation.

We determined our sample size based on prior research that collected subjective ratings of a large database of video stimuli, which collected 20 ratings per video (Lloyd et al., 2018). To achieve a sample large enough to collect a minimum of 20 ratings per each of our eighty-eight 911 calls, we determined that we would need a minimum of 220 participants with each participant rating eight of the 88 calls (i.e., 220 participants × 8 Calls = 1,760 ratings; 1,760 ratings/88 calls = 20 ratings per call). We oversampled to account for participants excluded for failing attention checks, resulting in a final sample of 250 participants.

Stimuli

We collected 225 real 911 calls reporting violent crimes (homicide, aggravated assault, death investigations) in 2017–2018 from Phoenix and Tucson police departments. A subset of 88 calls were chosen based on the following criteria: (a) an easily identifiable male or female caller, (b) reporting of a violent crime, (c) a sufficient length of at least 100 words, (d) a lack of clarity as to who the perpetrator was, and (e) high audio quality. See Appendix C for brief descriptions of each call.

Call lengths averaged from 34 s to 9 min and 21 s ($M = 3:31$, $SD = 1:54$). We clustered the calls by call length into eight bins (11 calls per bin) so that we could randomly assign participants to listen to one call from each length category and ensure the duration of the full study was relatively similar across participants. Thus, participants were randomly assigned to eight of the 88 calls, somewhat balanced by call duration. Participants took 53.7 min on average to complete the study ($SD = 16.8$ min).

Moderators

A team of trained research assistants coded each usable call for moderators tested in Studies 1–3 (i.e., caller gender, relationship to the victim). Two independent coders coded the caller's gender and relationship with the victim for each call, and a third coder made the final decision in the rare instances in which the two coders disagreed. The two independent coders had strong agreement (caller gender agreement: 97%; relationship to the victim agreement: 100%).

The set of 88 calls included a similar number of female callers ($n = 41$; 47%) and male callers ($n = 47$; 53%). Coders also coded the caller's relationship to the victim into the following categories: 1 = calling about a stranger or unknown person ($n = 70$; 79.55%), 2 = calling about an acquaintance/friend ($n = 7$; 7.95%), 3 = calling about a family member ($n = 7$; 7.95%), and 4 = unsure/relationship not known ($n = 4$; 4.55%). The relationship was coded as a family member when the caller specifically used a family label to refer to the victim (e.g., "father," "mother," "cousin"). The relationship was coded as an acquaintance or friend when the caller used the label "friend," when they knew the victim's name or otherwise knew the victim personally. The relationship was coded as a stranger when the caller explicitly stated that they did not know the victim or when the incident occurred away from the caller's line of sight (e.g., hearing gunshots outside their house). The relationship was coded as unknown when the caller was reporting a specific victim but did not make it clear if or how they knew them.

Due to the small number of callers reporting friends or family as victims, we collapsed across the friends and family categories to create two categories: known victim (15.91%) and unknown victim (79.54%).

Measures

After listening to each call, participants responded to an open-ended question asking for their first impression of the caller followed by single-item questions capturing a set of behaviors theorized to predict suspicion—all of which, other than emotionality, are reported in the other article (Bean et al., 2024). Because participants listened to multiple calls, we asked a single-item representing each of the scales from Studies 1–3. Participants reported (a) the emotionality of the caller on a 5-point scale from *not at all* to *extremely*, (b) how suspiciously they thought the caller was acting on a 6-point scale from *not at all* to *extremely*, and (c) how much they thought the police should investigate the caller as a suspect on a 6-point scale from *not at all* to *extremely*.

Results

Descriptive information and correlations among measures are reported in Table 6.

Higher levels of caller emotionality were significantly associated with less suspicion. This constitutes a conceptual replication of Studies 1–3, suggesting that lower levels of emotionality triggers suspicion, now generalized across a diverse set of real calls reporting violent crimes—but in this study with naturalistic emotion expression.

Caller Gender

We first visualized the data using a scatterplot (Figure 9), which revealed a negative linear pattern, such that perceiving the caller as more emotional appears to be associated with less suspicion.

Given that participant responses were nested within calls for this study, we used a multilevel model that included perceived emotionality (grand-mean centered; Level 1 variable), caller gender (male as the reference group; Level 2 variable), and the interaction as predictors of suspicion. One of our primary goals of this study was to determine whether the curvilinear effect for male callers—but not for female callers—found in Study 1 would replicate with another set of real calls. Thus, we entered perceived emotionality as a polynomial variable with three potential relationships with suspicion: linear, quadratic, and cubic. Note that testing the linear and quadratic trends was part of our preregistration plan, but the inclusion of the cubic trend was exploratory. The random effects were participant at Level 1 and call at Level 2. The model was fitted with restricted maximum likelihood using the lme4 package (Bates et al., 2015) in R and the assessment of fixed effects was conducted using t tests in the lmerTest package (Kuznetsova et al., 2017). We ran the first model with caller gender (Table 7) as the moderator.

See Table 7 for the full model results. This model revealed a significant and negative linear relationship between emotionality and suspicion, such that lower emotionality significantly predicted increased suspicion, $\beta = -9.25$, $p < .001$, conceptually replicating Studies 1–3. There was no significant quadratic or cubic relationship between emotion and suspicion nor any associated interactions with caller gender. Consistent with Studies 1–2, there was no main effect of caller gender, $\beta = -0.20$, $p = .22$.

Thus, the gender moderation demonstrated in Study 1 did not replicate in a larger, more diverse set of 911 calls with naturalistic emotion expression. One possibility is that our set of 88 calls did not reach the level of high emotion where we saw reactions to male and female callers diverge in Study 1. However, when comparing the levels of perceived emotion expression across studies, we found that the average level of perceived emotion expression in the upper third of this study's calls was relatively similar to average levels of

Table 6

Descriptive Statistics and Pearson Correlation Matrix for Dependent Variables, Study 4

Measure	M (SD)	1	2	3
1. Suspicion	1.77 (1.31)	—		
2. Treat as suspect	2.07 (1.55)	.72**	—	
3. Perceived emotionality	2.69 (1.51)	−.05*	.03	—

*$p < .05$.   **$p < .001$.

Figure 9

Mean (Standard Error) of Suspicion Toward Real 911 Callers at Each Level of Perceived Emotionality, Study 4
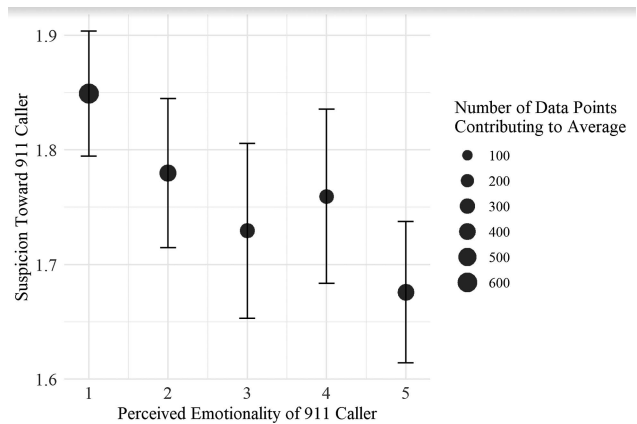
perceived emotion expression in Studies 1–3 (Table 8). Thus, the null gender interaction was not due to the real 911 calls in this study ($M = 4.20$) failing to reach similarly high levels of emotion evident in the real calls in Study 1 ($M = 3.99$).

Relationship to the Victim

To investigate the potential moderating role of the caller's relationship to the victim, we ran a similar multilevel model, replacing caller gender with relationship to the victim as the potential moderator. See Table 9 for the full model results. This model included perceived emotionality (grand-mean centered; Level 1 variable), relationship to the victim (stranger as the reference group; Level 2 variable), and the interaction as predictors of suspicion. We again entered perceived emotionality as a polynomial variable with three potential relationships with suspicion: linear, quadratic, and cubic. Note that because the relationship to the victim was unknown in four calls, this relationship model is run on a subset of observations that do not include those four calls (call $n = 84$, observation $n = 1,909$), as opposed to the gender model that was run on the complete set of observations (call $n = 88$, observation $n = 2,000$).

Consistent with Study 1 (and Supplemental Study S1), we found a significant effect of caller relationship, $\beta = 1.11$, $p < .001$. Participants were more suspicious of callers who reported that someone close to them was hurt than those who called about a stranger. This model again found the significant negative relationship between perceived emotionality and suspicion, but it was qualified by the hypothesized interaction between perceived emotionality and caller relationship, $\beta = -17.63$, $p < .001$. There were no significant quadratic or cubic relationships between emotion and suspicion nor any associated interactions with caller relationship to the victim.

We conducted a simple slopes analysis (that dropped the quadratic and cubic effects and associated interaction terms), which revealed that greater perceived emotionality significantly predicted less suspicion when the caller knew the victim and when they were a stranger. However, the interaction was driven by the fact that when the caller knew the victim (i.e., when more emotion would be expected), increased perceived emotionality predicted a larger

Table 7

Multilevel Model of Perceived Emotionality of the Caller and Caller Gender Predicting Suspicion Toward Real 911 Callers, Study 4

Factor	β	SE	df	t	p
Intercept	1.87	0.11	96	16.53	<.001
Level 1 variable					
Perceived emotionality (linear)	−9.25	2.38	1,683	−3.90	<.001
Perceived emotionality (quadratic)	−0.83	1.68	1,934	−0.49	.623
Perceived emotionality (cubic)	−2.68	1.54	1,829	−1.74	.082
Level 2 variable					
Caller gender	−0.20	0.16	86	−1.24	.220
Interaction term					
Perceived Emotionality × Caller Gender (linear)	−1.96	3.28	1,689	−0.60	.55
Perceived Emotionality × Caller Gender (quadratic)	4.49	2.41	1,886	1.86	.063
Perceived Emotionality × Caller Gender (cubic)	0.14	2.19	1,807	0.06	.950

Random effect	Variance	SD
Participant (Level 1)	0.21	0.46
Call (Level 2)	0.51	0.71
Residual	1.02	1.01

Note. SE = standard error.

decrease in suspicion, β = −0.37, p < .01, than when the caller did not know the victim, β = −0.12, p < .001.

In summary, we did find support for the hypothesis that the relationship between a caller's emotionality and suspicion depends on their relationship to the victim. Consistent with Study 1, we found that emotion expression had a stronger effect on suspicion when the caller was reporting that a known victim was hurt than when they were reporting that a stranger was hurt. The pattern was different from Study 1, however, in that the effect was curvilinear in Study 1 (i.e., moderate emotion was less suspicious than low or high emotion), but it was linear in this study (i.e., higher levels of emotion were less suspicious than lower levels). It is important to note that our call set may not have had enough calls in which someone was calling about someone they knew to draw strong conclusions about this hypothesis.

Treat as Suspect

We partially replicated the suspicion models' findings in the treat caller as a suspect models, such that (a) participants supported treating the caller as a suspect more when the caller was reporting that someone they knew was hurt compared to a stranger and (b) expressing relatively lower emotionality predicted more support for treating the caller as a suspect (although this emotionality effect manifested in the relationship model but not the gender model). We also replicated the interaction pattern with caller relationship, such that when callers knew the victim, the negative relationship between perceived emotionality and support for treating the caller as a suspect was stronger than when callers did not know the victim (though both effects were again significant).

Discussion

This study conceptually replicated our findings from tightly controlled, simulated calls in a diverse set of real 911 calls with naturalistic variation in emotion expression: Expressing lower emotion on real 911 calls is associated with greater suspicion that the caller is involved in the crime they are reporting—regardless of caller gender. We did, however, replicate the Study 1 finding that the strength of the relationship between emotion expression and decreased suspicion is stronger when the caller knows the victim relative to when they were calling about a stranger.

General Discussion

Emotions are a part of everyday life. As a result, it is common for people to believe that they can accurately read others' emotions and use that information to draw inferences about who they are (Van Kleef et al., 2011). We found that failing to express an expected level of emotion can lead others to draw *moral* inferences about someone's character and generate suspicion that they might have engaged in wrongdoing. If a witness to a tragic and emotional event fails to express a subjectively "appropriate" level of emotion, others infer that they are less capable of being a vulnerable victim of an immoral act and more capable of committing an immoral act, which triggers suspicion that they are concealing their involvement in the event they are reporting. In the language of moral typecasting theory (Gray & Wegner, 2009), holding all other things equal—including the exact details of the event and the script the witness used to report that event—failing to report the crime with an expected level of emotion leads others to morally typecast the witness more as a moral agent and less as a moral patient, which was associated with greater suspicion that they engaged in wrongdoing. In contrast, expressing the expected level of higher emotion can cause others to morally typecast the target more as a

Table 8

Mean and Standard Deviation of Levels of 911 Callers' Perceived Emotion Level Across Studies

Study	Caller emotion level		
	Low	Moderate	High
Study 1 (real calls, two selected per emotion condition)	M = 1.68, SD = 0.80[a]	M = 3.59, SD = 1.06[b]	M = 3.99, SD = 0.99[c]
Study 2 (simulated calls, manipulated emotion)	M = 2.59, SD = 1.08[a]	M = 3.58, SD = 1.03[b]	M = 4.09, SD = 0.94[c]
Study 3 (simulated calls, manipulated emotion)	M = 1.98, SD = 0.73[a]	M = 2.84, SD = 0.93[b]	M = 3.73, SD = 0.91[c]
Study 4 (real calls, sorted by average emotion ratings, then split into thirds)	M = 1.48, SD = 0.20[a]	M = 2.40, SD = 0.37[b]	M = 4.20, SD = 0.55[c]
Supplemental Study S1 (simulated calls, manipulated emotion)	M = 2.22, SD = 0.99[a]	M = 3.81, SD = 1.05[b]	M = 4.31, SD = 0.82[c]

Note. Perceived emotionality was rated on a response scale from 1 (*not at all*) to 5 (*extremely*). Differing subscripts denote significant difference at p < .001 within row.

Table 9

Multilevel Model of Perceived Emotionality of the Caller and Caller Relationship to the Victim Predicting Suspicion Toward Real 911 Callers, Study 4

Factor	β	SE	df	t	p
Intercept	1.59	0.07	111.22	22.66	<.001
Level 1 variable					
Perceived emotionality (linear)	−7.53	1.78	1055.32	−4.23	<.001
Perceived emotionality (quadratic)	1.71	1.32	1863.42	1.29	.196
Perceived emotionality (cubic)	−2.06	1.19	1781.31	−1.73	.084
Level 2 variable					
Caller relationship	1.11	0.16	93.17	6.72	<.001
Interaction term					
Perceived Emotionality × Caller Relationship (linear)	−17.63	4.46	885.89	−3.95	<.001
Perceived Emotionality × Caller Relationship (quadratic)	1.21	3.44	1816.81	0.35	.72
Perceived Emotionality × Caller Relationship (cubic)	−1.40	3.14	1810.61	−0.45	.655

Random effect	Variance	SD
Participant (Level 1)	0.22	0.47
Call (Level 2)	0.23	0.48
Residual	1.02	1.01

Note. SE = standard error.

moral patient and less as a moral agent—thereby buffering them against suspicion.

We demonstrated this theory in a context with life-or-death consequences: when witnessing a violent crime and calling 911 for help. Witnessing a violent crime is a traumatizing event. Although people and police officers might assume that they know what a "normal" emotional reaction would look like, individuals differ in the ways in which they react to trauma—from numbness and dissociation to hysteria and everything in between. Little do citizens who call 911 realize: How they behave when they make the plea for help in a moment of desperation can open them up to violating flawed assumptions about normal versus "suspicious" behavior. Even though 911 callers in simulated calls (Studies 2–3, S1) all reported the exact same incident and details with the same language, doing so in a less emotional tone made laypeople and police suspicious of the caller and more interested in investigating the caller as a suspect. When the caller expressed at least moderate emotion, laypeople were less suspicious; but the caller had to express a *high* level of emotion before police officers' suspicion dropped. We replicated this effect in a diverse set of real 911 calls with naturalistic emotion expression. With police officers' high confidence in their ability to read behavioral cues and detect deception despite being no better than chance (e.g., Masip et al., 2005; Meissner & Kassin, 2002), this suggests that one tragedy can become two: An innocent person failing to express enough emotion in their call for help can become a suspect and ultimately falsely convicted of the crime they reported.

Theoretical Explanation

Although this project makes several theoretical advances regarding the role of violating expectations and generating suspicion, we consider the major theoretical advance to be identifying a psychological explanation for *why* low emotion triggers suspicion in the context of a traumatic event: Failing to express the expected level of emotion for a given context will shape moral inferences that others draw about the expressor, thereby triggering suspicion. In the context of violent crimes, people appear to have a default assumption that people calling 911 for help are victims and expect them to be upset. Hearing someone they expect to be a vulnerable victim report the violent crime in a calm tone violates both laypeople's and police officers' expectations, which in turn was associated with them perceiving the caller to be less capable of vulnerability and victimhood and more capable of perpetrating immoral acts—both of which ultimately predicting increased suspicion.

This is highly problematic because the idea that it is abnormal to express low emotion in this context is contradicted by findings that emotions do not have recognizable and reliable fingerprints (Barrett, 2017). People vary wildly in their behavioral responses to tragedy—from appearing calm to hysterical and grief-stricken (e.g., Tyhurst, 1951; Wortman & Silver, 1989, 2001), with many victims of violence reporting their experiences with no emotion and numbness (for review, see Salerno, 2021). These expectations are strong: Contrary to hypotheses, the negative relationship between caller emotion expression and suspicion did not reliably depend on gender (except in Study 1)—despite robust gender stereotypes that women are more emotional than men (e.g., Fabes & Martin, 1991; Salerno et al., 2019). Police were, however, more suspicious of men than women overall—an effect we found among laypeople in only one study (Study S1; but not in Studies 1–2, 4). Both laypeople and police were consistently more suspicious of someone reporting that a loved one has been shot (parents, spouse) relative to a stranger in both tightly controlled, simulated calls and a diverse set of real calls. Further, we found that the caller's relationship to the victim shaped how suspicious people found low emotion: When the caller knew the victim (and therefore might be expected to express more emotion), relatively lower emotion predicted a greater increase in suspicion (Studies 1, 4, S1) and violated expectations more (Study S1) compared to when the caller did not know the victim.

Second, in line with social functional theories of emotion (Van Kleef et al., 2011) and the dyadic theory of morality (Gray & Wegner, 2009), an alternative mediation model revealed that emotion expression can also directly shape moral inferences about the caller and suspicion—above and beyond what was explained by laypeople's and police officers' expectations being violated. We discovered that expressions of emotion can cause people to perceive greater moral patiency and less moral agency in the expressor. Calling 911 to report a violent crime will always trigger an investigation. Police officers are literally put in the position of professionally morally typecasting witnesses: Are they also a victim of this terrible harm, or could they be the perpetrator? Low emotion in the context of a traumatic 911 call led to inferences that the caller was more capable of doing immoral things and less capable of having immoral things done to them. That is, low emotion caused listeners to morally typecast the caller as a moral agent and less as a moral patient. Thus, police might end up targeting innocent people who express low emotion because their

lack of emotion signals lesser capacity to be a victim and greater capacity to be a perpetrator of harm—independent from the actual case details and evidence strength. These indirect effects of emotion expression on suspicion through expectation violations and moral typecasting were robust to the caller's gender. It is important to note that although our design supports a serial model, it does not allow us to draw chronological causal conclusions about the serial model or to provide strong evidence for one mediation model over the other (Table 10).

We also believe that this project contributes theoretical advances to prior work on the emotional victim effect (e.g., Ask & Landström, 2010; Landström et al., 2015), which has demonstrated that victims who recount their experience with emotion are seen as more credible. We demonstrated the effect of emotionality beyond assessments of the credibility of the victim's account of what happened (i.e., did it happen the way they described?) to judgments of the target's moral character and suspicion that they might have actually perpetrated the incident they were reporting. Further, we extended the investigation from people already classified as victims to whether a bystander gets classified as a victim to begin with—or if they are reclassified as a potential perpetrator.

Limitations

We replicated our findings across both tightly controlled, simulated calls and a diverse set of real 911 calls, modeling the content of the 911 calls as a random category (Study 4). This enabled us to conceptually generalize the current results to the larger universe of unsampled calls, similar to how researchers generalize results beyond unsampled participants (Judd et al., 2012). However, the conclusions we can draw are somewhat limited by methodological choices (Table 10). We investigated the impact of the level of emotion on moral inferences and suspicion, but this construct was limited to general emotionality, preventing us from drawing conclusions about specific discrete emotions (e.g., anger, fear). We tested the impact of emotion expression on moral typecasting in the context of negative immoral acts but did not test if this generalized to *positive* moral acts as the theory of moral typecasting would suggest. Our findings are also limited to a context where high emotion is expected—investigating whether the findings would generalize to violating expectations for *low* emotion on moral typecasting is important. An even stricter test of moral typecasting theory would be to establish whether low emotion would continue to

Table 10
Limitations of the Research

Limitation
Although we used a set of diverse 911 calls (Study 4), they do not capture the vast diversity of circumstances and contexts in 911 calls. Continuing to investigate these effects in additional contexts and circumstances is important.
Our findings are limited to contexts in which high emotion is expected. It is important to test whether violating expectations for *low* emotion would also generate suspicion through violations of expectations and moral typecasting.
Although we did use stimulus sampling, our causal evidence that emotion expression affects violation of expectations, moral typecasting, and suspicion (Studies 2–3, Supplemental Study S1) is limited to one set of simulated calls that were modeled after one actor's style of emotion expression. It is therefore possible that we would have found different effects if the style of emotion was more diverse and natural. Our concern is somewhat assuaged by our replication with 88 real 911 calls comprising natural emotion expression, but more work with naturalistic emotion expression would be helpful.
We investigated perceptions of general emotionality, and as a result our findings we do not speak to how people would react to specific emotions (e.g., anger, fear, sadness) from men versus women. We know, for example, that people have more negative reactions to female anger compared to male anger (e.g., Brescoll & Uhlmann, 2008; Salerno et al., 2018, 2019; Salerno & Peter-Hagene, 2015). Thus, for example, people might have more negative reactions to female 911 callers if they are expressing anger.
Our samples were majority White. Further, our police officer sample was mostly White and male—a breakdown that mirrors the actual racial distribution of police officers in America (Police Officers, n.d.). It is important to test whether our findings generalize to other racial groups.
We did not investigate the race of the caller, which is very likely to impact suspicion. It is well-established that people are biased to find Black people guilty of the same crimes more often than White people (e.g., Mitchell et al., 2005), and Black people are disproportionately more likely to be wrongfully convicted relative to White people (Gross et al., 2017). It is important to manipulate or measure the perceived race of the caller in future research.
Our mediation models support our proposed theoretical pathway, but the models rely on correlational links between the mediators and our dependent variable. Thus, our models cannot establish causality or, in particular, the proposed chronology in the serial model. We do, however, provide an alternative parallel mediator model that establishes each mediator (moral agency, moral patiency, violation of expectations) uniquely explains the effect of emotion on suspicion when controlling for each other. Our design did not lend itself to being able to argue for one model over the other, but future research that does so directly would be helpful.
Future research should explore whether our findings generalize across other aspects of a caller that might trigger bias, such as accents or indicators of socioeconomic status.
Across studies, there was some slight fluctuation in the success of the emotion manipulation. For example, the emotion condition effect on perceived emotion (i.e., the manipulation check) was stronger for women in Study 1 and stronger for men in Supplemental Study S1. However, we replicated our pattern of results across experiments in which the emotion manipulation did not depend on gender in similar simulated calls (Studies 2–3) and a study with naturalistic emotion expression from men and women (Study 4). Further, perceiving similar levels of emotion differently based on gender expectations might be part of the phenomenon. For example, in Supplemental Study S1, participants also rated people reporting a stranger had been shot as more emotional than someone reporting a spouse was shot—despite people in the stranger versus spouse conditions listening to the exact same calls. Thus, these findings might not be a failure of the manipulation check but a gender or relationship effect on how participants interpret the same behaviors as emotional or not.
In Study 4, we investigated differences in suspicion due to the caller's relationship to the victim—comparing calls where the caller knew the victim versus when the caller did not know the victim. However, only 15% of our set of 88 real 911 calls were made by someone who knew the victim. Future research should investigate this question with not only more exemplars of callers reporting a loved one was hurt but also enough to be able to compare suspicion across different levels of closeness (e.g., spouses/partners vs. family vs. friends vs. acquaintances).

predict decreased moral patiency and increased moral agency when low emotion is expected for the context. Finally, it is well-established that people of color face increased suspicion and likelihood of wrongful conviction, which we were not able to address in our studies because the race of 911 callers in our studies is unknown. Further, our samples were predominantly White—an issue that mirrors the underrepresentation of people of color on the police force, judicial bench, and juries. A very important next step, however, would be to test whether our findings are moderated by caller or participant race.

Real-World Implications

Beyond the theoretical contributions of how emotion expression can affect moral inferences drawn about the expressor, these findings have dire real-world implications. We provide experimental evidence of a well-documented anecdotal phenomenon: When innocent witnesses to a violent crime fail to express appropriate emotion, they become vulnerable to being targeted as a suspect—triggering a snowballing confirmation bias effect leading to wrongful convictions (Scherr et al., 2020). Police investigators' (unfounded) confidence in their ability to judge behavior as "human lie detectors" is one of the main sources of misclassification of an innocent person as a suspect (Leo & Drizin, 2010). Indeed, the flawed assumption that low emotion is suspicious directly contradicts one study demonstrating that emotion on 911 calls can be associated with *greater* likelihood the caller was convicted for the incident they were reporting (Markey et al., 2022). This situation is even more urgent due to recent discoveries that police are being trained in a new frontier of junk science called "911 Call Analysis" (Murphy, 2022), in which a former officer provides a set of guilt/innocence indicators that he claims can enable police to identify 911 callers who committed murders they reported based only on their linguistic behavior on the call—including things like perceived emotionality and urgency.

Conclusion

The errors made in this context are not without consequence. Perceived emotionality is a nonevidentiary basis for suspicion that can lead police to insert innocent people into an eyewitness lineup, increasing the risk of misidentification (Wells et al., 2020), or haul them into an interrogation room, increasing the risk of a false confession (Moody et al., 2023). The net result: Innocent people being morally typecast as more capable of being a perpetrator than a victim and wrongfully convicted—whose only crime was to violate expectations for what a normal plea for help looks like.

References

Ask, K., & Landström, S. (2010). Why emotions matter: Expectancy violation and affective response mediate the emotional victim effect. *Law and Human Behavior*, 34(5), 392–401. https://doi.org/10.1007/s10979-009-9208-6

Barrett, L. F. (2017). Categories and their role in the science of emotion. *Psychological Inquiry*, 28(1), 20–26. https://doi.org/10.1080/1047840X.2017.1261581

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bean, S. R., Salerno, J. M., Duran, N., Kassin, S., Bussey, L., Olson, B., & Phalen, H. (2022, March). *What makes people suspicious? Identifying behavioral predictors of suspicion toward 911 callers* [Paper presentation]. Annual American Psychology-Law Society Conference, Denver, CO, United States.

Bean, S. R., Salerno, J. M., Duran, N. D., & Kassin, S. M. (2019, March). *Falling under the lens of suspicion: Expressing "just enough" emotion in 911 calls* [Paper presentation]. American Psychology-Law Society, Portland, OR, United States.

Bean, S. R., Wulff, A., Lawrence, M., Reeder, I., Duran, N., Kassin, S., & Salerno, J. M. (2024). *What sparks suspicion in 911 calls? Identifying behaviors that trigger suspicion* [Manuscript submitted for publication]. School of Social and Behavioral Sciences, Arizona State University.

Brescoll, V. L., & Uhlmann, E. L. (2008). Can an angry woman get ahead? Status conferral, gender, and expression of emotion in the workplace. *Psychological Science*, 19(3), 268–275. https://doi.org/10.1111/j.1467-9280.2008.02079.x

Douglas, B. D., Ewell, P. J., & Brauer, M. (2023). Data quality in online human-subjects research: Comparisons between MTurk, Prolific, CloudResearch, Qualtrics, and SONA. *PLOS ONE*, 18(3), Article e0279720. https://doi.org/10.1371/journal.pone.0279720

Fabes, R. A., & Martin, C. L. (1991). Gender and age stereotypes of emotionality. *Personality and Social Psychology Bulletin*, 17(5), 532–540. https://doi.org/10.1177/0146167291175008

Firstman, R., & Salpeter, J. (2008). *A criminal injustice: A true crime, a false confession, and the fight to free Marty Tankleff*. Ballantine Books.

Gray, K., Schein, C., & Ward, A. F. (2014). The myth of harmless wrongs in moral cognition: Automatic dyadic completion from sin to suffering. *Journal of Experimental Psychology: General*, 143(4), 1600–1615. https://doi.org/10.1037/a0036149

Gray, K., Waytz, A., & Young, L. (2012). The moral dyad: A fundamental template unifying moral judgment. *Psychological Inquiry*, 23(2), 206–215. https://doi.org/10.1080/1047840X.2012.686247

Gray, K., & Wegner, D. M. (2009). Moral typecasting: Divergent perceptions of moral agents and moral patients. *Journal of Personality and Social Psychology*, 96(3), 505–520. https://doi.org/10.1037/a0013748

Gray, K., & Wegner, D. M. (2011a). Dimensions of moral emotions. *Emotion Review*, 3(3), 258–260. https://doi.org/10.1177/1754073911402388

Gray, K., & Wegner, D. M. (2011b). To escape blame, don't be a hero—Be a victim. *Journal of Experimental Social Psychology*, 47(2), 516–519. https://doi.org/10.1016/j.jesp.2010.12.012

Gross, S. R., Possley, M., & Stephens, K. (2017). *Race and wrongful convictions in the United States*. The National Registry of Exonerations. https://www.law.umich.edu/special/exoneration/Pages/Race.aspx

Hareli, S., & Hess, U. (2010). What emotional reactions can tell us about the nature of others: An appraisal perspective on person perception. *Cognition and Emotion*, 24(1), 128–140. https://doi.org/10.1080/02699930802613828

Heath, W. P. (2009). Arresting and convicting the innocent: The potential role of an "inappropriate" emotional display in the accused. *Behavioral Sciences & the Law*, 27(3), 313–332. https://doi.org/10.1002/bsl.864

Heath, W. P., & Grannemann, B. D. (2014). How video image size interacts with evidence strength, defendant emotion, and the defendant-victim relationship to alter perceptions of the defendant. *Behavioral Sciences & the Law*, 32(4), 496–507. https://doi.org/10.1002/bsl.2120

Heath, W. P., & Grannemann, B. D. (2015). Expectations for defendant emotion. *Applied Psychology in Criminal Justice*, 11(2), 126–145.

Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology*, 103(1), 54–69. https://doi.org/10.1037/a0028347

Kassin, S. M. (2022a, March). *It's time to ask the second question* [Paper presentation]. Annual American Psychology-Law Society Conference, Denver, CO, United States.

Kassin, S. M. (2022b). It's time to bury three justice-corrupting myths once and for all. *Journal of Applied Research in Memory and Cognition*, *11*(2), 161–165. https://doi.org/10.1037/mac0000039

Kassin, S. M., Dror, I. E., & Kukucka, J. (2013). The forensic confirmation bias: Problems, perspectives, and proposed solutions. *Journal of Applied Research in Memory and Cognition*, *2*(1), 42–52. https://doi.org/10.1016/j.jarmac.2013.01.001

Kassin, S. M., & Gudjonsson, G. H. (2004). The psychology of confessions: A review of the literature and issues. *Psychological Science in the Public Interest*, *5*(2), 33–67. https://doi.org/10.1111/j.1529-1006.2004.00016.x

Keltner, D., & Haidt, J. (1999). Social functions of emotions at four levels of analysis. *Cognition and Emotion*, *13*(5), 505–521. https://doi.org/10.1080/026999399379168

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, *108*(3), 480–498. https://doi.org/10.1037/0033-2909.108.3.480

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for *t*-tests and ANOVAs. *Frontiers in Psychology*, *4*, Article 863. https://doi.org/10.3389/fpsyg.2013.00863

Landström, S., Ask, K., & Sommar, C. (2015). The emotional male victim: Effects of presentation mode on judged credibility. *Scandinavian Journal of Psychology*, *56*(1), 99–104. https://doi.org/10.1111/sjop.12176

Lawrence, M. L., Thomas, S., Kassin, S., & Salerno, J. M. (2024). *A thematic analysis of behaviors that makes officers suspicious of interviewees* [Manuscript in preparation]. School of Social and Behavioral Sciences, Arizona State University.

Leo, R. A., & Drizin, S. A. (2010). The three errors: Pathways to false confession and wrongful conviction. In D. Lassiter & M. Christian (Eds.), *Police interrogations and false confessions: Current research, practice, and policy recommendations* (pp. 9–30). American Psychological Association. https://papers.ssrn.com/abstract=1542901

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

Lloyd, E. P., Deska, J. C., Hugenberg, K., McConnell, A. R., Humphrey, B. T., & Kunstman, J. W. (2018). Miami University deception detection database. *Behavior Research Methods*, *51*, 429–439. https://doi.org/10.3758/s13428-018-1061-4

Lowrey-Kinberg, B., Senn, S. L., Dunn, K., Gould, J. B., & Hail-Jares, K. (2019). Origin of implication: How do innocent individuals enter the criminal justice system? *Crime & Delinquency*, *65*(14), 1949–1975. https://doi.org/10.1177/0011128718793618

Markey, P. M., Feeney, E., Berry, B., Hopkins, L., & Creedon, I. (2022). Deception cues during high-risk situations: 911 homicide calls. *Psychological Science*, *33*(7), 1040–1047. https://doi.org/10.1177/09567976221077216

Masip, J., Alonso, H., Garrido, E., & Antón, C. (2005). Generalized Communicative Suspicion (GCS) among police officers: Accounting for the investigator bias effect. *Journal of Applied Social Psychology*, *35*(5), 1046–1066. https://doi.org/10.1111/j.1559-1816.2005.tb02159.x

Meissner, C. A., & Kassin, S. M. (2002). "He's guilty!" Investigator bias in judgments of truth and deception. *Law and Human Behavior*, *26*(5), 469–480. https://doi.org/10.1023/A:1020278620751

Mitchell, T. L., Haw, R. M., Pfeifer, J. E., & Meissner, C. A. (2005). Racial bias in mock juror decision-making: A meta-analytic review of defendant treatment. *Law and Human Behavior*, *29*(6), 621–637. https://doi.org/10.1007/s10979-005-8122-9

Moody, S. A., Cabell, J. J., Livingston, T. N., & Yang, Y. (2023). Evidence-based suspicion and the prior probability of guilt in police interrogations.

Law and Human Behavior, *47*(2), 307–319. https://doi.org/10.1037/lhb0000513

Murphy, B. (2022). *They called 911 for help. Police and prosecutors used a new junk science to decide they were liars*. ProPublica. https://www.propublica.org/article/911-call-analysis-fbi-police-courts

National Registry of Exonerations. (n.d.). https://www.law.umich.edu/special/exoneration/Pages/about.aspx

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, *2*(2), 175–220. https://doi.org/10.1037/1089-2680.2.2.175

Police Officers. (n.d.). *Data USA*. https://datausa.io/profile/soc/police-officers

Reeder, I., Wulff, A., Bean, S., Ienaga, M., Sarraf, M., Stern, C., Duran, N., & Salerno, J. M., (2024, March). *Suspiciousness: Behaviors that lead to 911 callers becoming suspects* [Paper presentation]. American Psychology-Law Society Conference, Los Angeles, CA, United States.

Salerno, J. M. (2021). Emotion and legal fact finding. *Annual Review of Law and Social Science*, *17*, 181–203. https://doi.org/10.1146/annurev-lawsocsci-021721-072326

Salerno, J. M., Bean, S., Duran, N., & Kassin, S. (2023, June). *Emotion expression on 911 calls: Is the caller keeping their Involvement in the crime a secret?* [Paper presentation]. Annual Meeting of the European Association of Social Psychology, Krakow, Poland.

Salerno, J. M., Bean, S. R., Duran, N., & Kassin, S. (2022, March). *Lack of emotion in 911 calls increases police officers' suspicion due to violation of expectations and moral typecasting* [Paper presentation]. Annual American Psychology-Law Society Conference, Denver, CO, United States.

Salerno, J. M., & Peter-Hagene, L. C. (2015). One angry woman: Anger expression increases influence for men, but decreases influence for women, during group deliberation. *Law and Human Behavior*, *39*(6), 581–592. https://doi.org/10.1037/lhb0000147

Salerno, J. M., Peter-Hagene, L. C., & Jay, A. C. V. (2019). Women and African Americans are less influential when they express anger during group decision making. *Group Processes & Intergroup Relations*, *22*(1), 57–79. https://doi.org/10.1177/1368430217702967

Salerno, J. M., Phalen, H. J., Reyes, R. N., & Schweitzer, N. J. (2018). Closing with emotion: The differential impact of male versus female attorneys expressing anger in court. *Law and Human Behavior*, *42*(4), 385–401. https://doi.org/10.1037/lhb0000292

Santos, F. (2006). *DNA evidence frees a man imprisoned for half his life*. The New York Times. https://www.nytimes.com/2006/09/21/nyregion/21dna.html

Scherr, K. C., Redlich, A. D., & Kassin, S. M. (2020). Cumulative disadvantage: A psychological framework for understanding how innocence can lead to confession, wrongful conviction, and beyond. *Perspectives on Psychological Science*, *15*(2), 353–383. https://doi.org/10.1177/1745691619896608

Shapiro, J. (1998). The wrong men on death row. *U.S. News & World Report*, *125*(18), 22–26.

Shepherd, S., Kay, A. C., & Gray, K. (2019). Military veterans are morally typecast as agentic but unfeeling: Implications for veteran employment. *Organizational Behavior and Human Decision Processes*, *153*, 75–88. https://doi.org/10.1016/j.obhdp.2019.06.003

Simmons, J., Nelson, L., & Simonsohn, U. (2013, January). *Life after p-hacking* [Paper presentation]. Annual Meeting of the Society for Social and Personality Psychology, New Orleans, LA, United States. https://doi.org/10.2139/ssrn.2205186

Tyhurst, J. S. (1951). Individual reactions to community disaster: The natural history of psychiatric phenomena. *The American Journal of Psychiatry*, *107*(10), 764–769. https://doi.org/10.1176/ajp.107.10.764

Van Kleef, G. A., Van Doorn, E. A., Heerdink, M. W., & Koning, L. F. (2011). Emotion is for influence. *European Review of Social Psychology*, *22*(1), 114–163. https://doi.org/10.1080/10463283.2011.627192

Warner, L. R., & Shields, S. A. (2009). Judgements of others' emotional appropriateness are multidimensional. *Cognition and Emotion*, *23*(5), 876–888. https://doi.org/10.1080/02699930802212365

Wells, G. L., Kovera, M. B., Douglass, A. B., Brewer, N., Meissner, C. A., & Wixted, J. T. (2020). Policy and procedure recommendations for the collection and preservation of eyewitness identification evidence. *Law and Human Behavior*, *44*(1), 3–36. https://doi.org/10.1037/lhb0000359

Wortman, C. B., & Boerner, K. (2012). Beyond the myths of coping with loss: Prevailing assumptions versus scientific evidence. In H. S. Friedman (Ed.),

The Oxford handbook of health psychology (pp. 439–476). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195342819.013.0019

Wortman, C. B., & Silver, R. C. (1989). The myths of coping with loss. *Journal of Consulting and Clinical Psychology*, *57*(3), 349–357. https://doi.org/10.1037/0022-006X.57.3.349

Wortman, C. B., & Silver, R. C. (2001). The myths of coping with loss revisited. In M. S. Stroebe, R. O. Hansson, W. Stroebe, & H. Schut (Eds.), *Handbook of bereavement research: Consequences, coping, and care* (pp. 405–429). American Psychological Association. https://doi.org/10.1037/10436-017

Appendix A

Study Measures

Suspicion Scale (Studies 1–3, S1)

1. How suspicious was the caller acting?

   Not suspicious at all, a little suspicious, somewhat suspicious, moderately suspicious, really suspicious, extremely suspicious

2. How sincere do you think the caller is being? (RS)

   Extremely insincere, moderately insincere, somewhat insincere, somewhat sincere, moderately sincere, extremely sincere

3. How normal do you think the caller is acting? (RS)

   Extremely strange, moderately strange, somewhat strange, somewhat normal, moderately normal, extremely normal

4. How likely do you think it is that the caller is hiding something?

   Extremely unlikely, moderately unlikely, somewhat unlikely, neither unlikely nor likely, somewhat likely, moderately likely, extremely likely

5. How likely do you think it is that the caller played a role in the crime?

   Extremely unlikely, moderately unlikely, somewhat unlikely, neither unlikely nor likely, somewhat likely, moderately likely, extremely likely

Treat as Suspect Scale (Studies 1–3, S1)

- For each statement below, participants chose between: strongly disagree, disagree, somewhat disagree, neither agree nor disagree, somewhat agree, agree, strongly agree

Please indicate your level of agreement with the following statements.

1. The police should use their resources to investigate the caller's involvement in the crime further.

2. The police should treat the caller as a suspect.

Violation of Expectations Scale (Studies 2–3, S1; Items Presented in a Randomized Order)

- For each statement below, participants chose between: strongly disagree, disagree, slightly disagree, slightly agree, agree, strongly agree

Please indicate how much you agree with the following statements.

1. I was genuinely surprised about how the caller acted.

2. The caller behaved about how I expected. (RS)

3. The caller's behavior violated my expectations for what a 911 call would sound like.

4. If I was in the same situation I think I would behave similarly to the caller. (RS)

Moral Agency and Moral Patency (Studies 2–3 and S1; Items Presented in a Randomized Order)

- For each statement below, participants chose between: strongly disagree, disagree, slightly disagree, slightly agree, agree, strongly agree

- Items 1–5 comprised the Moral Agency Scale; we cut Items 2 and 4 in Study 2 to improve reliability (cutting these items would make reliability worse in Studies 3 and S1, so we used all five items in those studies)

- Items 6–10 comprised the Moral Patency Scale; we cut Items 8 and 10 in Study 2 to improve reliability (cutting these items would make reliability worse in Studies 3 and S1, so we used all five items in those studies)

Please indicate how much you agree with each statement.

1. The caller seems like someone who could be capable of doing immoral things.

2. It would be ridiculous to blame the caller at all for any part of what is happening. (RS)

3. The caller bears some responsibility for what is happening to the caller's spouse.

(Appendices continue)

4. The caller strikes me as a person incapable of hurting someone else. (RS)

5. The caller sounds like they are in control of the situation.

6. The caller seems very vulnerable.

7. I consider the caller to be as much of a victim of this event as the spouse.

8. The caller seems capable of handling a lot of fear. (RS)

9. The caller does not seem to be strong enough to handle the fear and pain that is resulting from this situation.

10. The caller doesn't seem to be capable of suffering that much. (RS)

Study 4 Measures

Suspicion

Do you think the caller was acting suspicious? Responded on a scale from 1 (*not at all*) to 6 (*extremely*)

Treat as Suspect

Do you think the police should use their resources to investigate the caller's involvement in the crime further? Responded on a scale from 1 (*not at all*) to 6 (*extremely*)

Emotionality

Caller seemed very upset. Responded with *not at all, a little, somewhat, moderately, extremely*

Manipulation/Attention Checks (Studies 1–3, S1)

Emotion

To what degree did the caller express emotion? *Not at all, a little, a moderate amount, a lot, an extreme amount.*

Gender

What was the gender of the caller? (*male, female, unsure*)

Relationship

At the beginning of the survey you were told the relationship between the caller and the victim. Which relationship were you told?

- Study 1 response options: Parents, brother, coworker, neighbor; The victim was someone the caller didn't know; I was not given this information.

- Studies 2–3, S1 response options: Parent, sibling, coworker, spouse; The victim was someone the caller didn't know; I was not given this information.

Appendix B

Call Script (Studies 2–3, S1)

911 Call Script

911: *911, what's your emergency?*

Caller: *I just heard shots outside my house and I think my [husband/wife] has been shot.*

911: *You said someone has been shot?*

Caller: *Yes my [husband/wife], [he/she] needs help please.*

911: *Okay, what's your address?*

Caller: *9718 East Cherry*

911: *Alright. Is your [husband/wife] breathing?*

Caller: *I think so but [he's/she's] gasping for air. Can you send help?*

911: *Yes, I'm sending medical units to your location now. Did you see where the shots came from?*

Caller: *No… no I didn't see anything I was inside the house till I heard them. My [husband/wife] was getting the mail outside…*

911: *Do you see anything right now?*

Caller: *Uhhh… no I don't see anything.*

911: *Okay, stay on the phone with me. I need you to stay in a safe spot until help arrives. Where are you right now?*

Caller: *I'm outside my house, my [husband/wife] is laying right here.*

911: *Can you go back inside the house?*

Caller: *Yeah, I'm near the door… I'm not sure if I should leave [him/her]. Are you almost here?*

911: *Help is almost to you, just stay on the phone with me. You said you heard a noise, did you see anything or anyone?*

Caller: *[He/she] needs help – can you send someone?*

911: *Did you see anyone?*

Caller: *[He's/She's] been shot – [he/she] can't breathe…*

911: *[Ma'am/Sir], please answer me: did you see anyone?*

Caller: *No… My [husband/wife] is gasping for air…*

911: *Help should be arriving right now [Ma'am/Sir].*

(*Appendices continue*)

Appendix C

Study 4 Call Details

Call	Reason for call	Relationship to victim	Length of call	Caller gender
1	Caller reported a man bleeding in the street	Stranger	6:03	Male
2	Adolescent caller reported a suicide in a neighbor's home	Acquaintance/friend	4:26	Male
3	Caller heard gunshots with at least one victim	Acquaintance/friend	3:48	Female
4	Caller heard gunshots and saw people on the ground	Stranger	3:32	Female
5	Security officer reported the alleged assault of an unconscious man in bathroom	Stranger	1:45	Male
6	Caller reported that his friend has been shot and provided descriptions of the assailants	Acquaintance/friend	3:25	Female
7	Store employee called 911 on behalf of a man who claimed his friend was shot outside the store	Stranger	4:59	Female
8	Caller reported that a man has been stabbed outside of his apartment, and someone is banging on his door	Stranger	5:03	Male
9	Father reported that his 11-year-old child deliberately shot himself in the head	Family	6:09	Male
10	Caller heard a drive-by shooting on a nearby street	Stranger	2:03	Female
11	Caller reported an alleged shooting of a bleeding man outside of his place of employment	Stranger	2:42	Male
12	Caller reported an alleged shooting outside her home involving her son though she is not sure who was shot	Family	3:36	Female
13	Caller reported a shooting of a man who is not breathing and describes the assailant	Stranger	1:54	Female
14	Caller reported a shooting outside of a store, reporting that both the victim and assailant fled the scene	Stranger	1:32	Male
15	Caller reported that he heard five gunshots outside of a store	Stranger	2:08	Male
16	Security officer reported hearing gunshots outside of a store	Stranger	1:21	Male
17	Caller reported that her mother was dead in her ex-boyfriend's home	Family	5:18	Female
18	Caller reported that her daughter's boyfriend stole her car and threatened to kill the caller	Family	3:37	Female
19	Caller reported that a women who appears to have been assaulted jumped out of a vehicle that then fled	Stranger	2:04	Male
20	Caller reported a drive-by shooting and described the vehicle	Stranger	1:24	Male
21	Caller reported that a man was shot outside of her place of business	Stranger	2:13	Female
22	Caller reported that a young man, and potentially two others, has been shot but abruptly ends the call	Unclear	0:48	Male
23	Caller reported that a woman has been shot and can see people running from the scene	Stranger	2:46	Female
24	Caller reported that three victims have been shot and described the injuries	Stranger	4:42	Female
25	Caller reported that several people have been shot in a hotel stairwell and can hear screaming	Stranger	1:58	Male
26	Witness called 911 again after a shooting to report that he now sees a vehicle fleeing the scene	Stranger	1:04	Male
27	Caller reported that a woman is dead, and police are on the scene, but the victim needs an ambulance	Stranger	2:01	Female
28	Caller reported that a man was shot at a motel	Stranger	1:02	Female
29	Caller reported that a man was shot and killed, describing details of a driver and his vehicle who fled the scene	Stranger	2:18	Male
30	Caller reported that people are dying and urgently requested an ambulance in addition to police officers	Stranger	0:41	Female
31	Silent witness reported that she knew who has the gun that was used in a shooting the prior week	Stranger	4:38	Female
32	Two individuals reported a fight that developed into a shooting at a hotel and that people are running	Stranger	7:06	Male
33	Caller reported that four victims have been shot	Stranger	3:32	Female
34	Caller reported hearing shots and screaming, adding that another witness saw that someone was on the ground	Stranger	2:31	Male
35	Employee reported hearing a shooting outside of a store, describing the victim and the car that fled the scene	Stranger	5:24	Female

(Appendices continue)

Appendix C (*continued*)

Call	Reason for call	Relationship to victim	Length of call	Caller gender
36	Caller reported a shooting and urges the dispatcher to send help	Unclear	2:45	Male
37	Caller reported that her parents heard shots fired and found a victim bleeding in an alley	Stranger	8:48	Female
38	Caller reported a man bleeding in an alley who appears to have been shot in the head	Stranger	6:47	Female
39	Caller reported a home invasion and that her roommate and an intruder were shot	Stranger	7:45	Male
40	Caller heard a gunshot and watched a victim fall to the ground	Stranger	5:55	Male
41	Caller reported a potential kidnapping after witnessing a struggle between a driver and bleeding passenger	Stranger	5:11	Male
42	Caller reported finding a deceased victim near her place of employment	Stranger	3:04	Female
43	Caller reported coming home to discover his family had been shot	Family	1:01	Male
44	Caller reported shots fired and a man lying on floor	Stranger	2:53	Female
45	Caller reported that a construction worker fell from a large height and is badly injured	Stranger	1:47	Male
46	Caller reported that a construction worker fell from a large height and is badly injured	Unclear	0:35	Male
47	Security guard reported that a man who was found at a construction site and may have overdosed was brought into the hospital	Stranger	2:05	Male
48	Caller heard gunshots in the parking lot	Stranger	3:03	Female
49	Caller reported that, while on the phone with her friend, she heard gunshots and her friend screaming	Acquaintance/friend	1:08	Female
50	Caller witnessed a shooting, and reported one man down and bleeding	Stranger	3:48	Male
51	Security officer reported that a customer was shot in a store	Stranger	3:01	Female
52	Caller heard gunshots and people yelling for help	Stranger	3:30	Female
53	Caller reported that someone has been shot in a neighborhood	Stranger	4:20	Female
54	Caller reported that a man was shot in the head outside of his store and appears to be deceased	Stranger	5:06	Male
55	Caller reported that a man came to her door and informed her that there is a deceased victim in the street	Stranger	6:28	Female
56	Caller reported that a theft victim hit the assailant with her vehicle in a park near the caller's home, leaving the assailant severely injured	Stranger	6:47	Female
57	Employee reported gunshots at a gas station and believed that a victim was shot in a car	Stranger	3:16	Male
58	Caller reported that their friend was shot, and the suspects fled the scene	Acquaintance/friend	4:51	Male
59	Employee heard gunshots and reported that a victim who was shot entered his store	Stranger	4:10	Male
60	Callers reported gunshots at their apartment complex with at least one known victim	Stranger	1:50	Male
61	Caller reported that a victim was shot in the mouth outside of her apartment complex	Stranger	1:20	Male
62	Caller reported that his cousin was shot	Family	1:19	Male
63	Caller reported that someone was shot but does not provide details regarding the incident	unclear	1:07	Male
64	Caller reported that a victim was shot and killed at a house party	Stranger	2:17	Female
65	Caller reported that a victim was shot in her apartment complex	Stranger	1:59	Female
66	Caller reported hearing gunshots in her apartment complex, noting that there is a drug dealer living in a building in the direction of the shots	Stranger	2:13	Female
67	Caller reported that she heard shots right outside her building, and she is hiding in the closet	Stranger	2:29	Female
68	Caller reported hearing shots and screaming outside her building, and a second witness saw someone took the weapon from the scene	Stranger	3:30	Female

(*Appendices continue*)

Appendix C (*continued*)

Call	Reason for call	Relationship to victim	Length of call	Caller gender
69	Caller reported several shots outside her apartment with a victim on the ground	Stranger	2:19	Female
70	Caller reported hearing shots fired outside his apartment	Stranger	2:23	Male
71	Caller reported an alleged suicide, according to a suicide note at the scene	Stranger	5:05	Female
72	Caller reported a stabbing, and the assailant fled on a bicycle	Stranger	3:35	Male
73	Caller reported finding a victim of an alleged stabbing	Stranger	1:53	Male
74	Caller reported that he found a deceased victim who appears to have been stabbed in a mobile home	Family	3:14	Male
75	Caller found a young male who appears to have been shot in the neck and is struggling to breathe	Stranger	3:33	Male
76	Caller reported that a bleeding victim has been shot and is claiming that someone is trying to kill him	Stranger	4:40	Male
77	Caller reported an armed robbery and shooting	Acquaintance/friend	2:08	Male
78	Caller reported that a man on the street is not moving and looks beaten up	Stranger	3:02	Male
79	Caller reported that a victim was shot in the head and killed in a yard	Stranger	7:23	Male
80	Caller reported two people lying on the road, both with apparent gunshot wounds to the head	Stranger	4:01	Male
81	Caller stated a young boy came to her house, claiming that his father shot his mother and brother	Stranger	8:14	Female
82	Caller reported that a man in his condominium has assaulted two people, noting that they already have a restraining order against the assailant	Acquaintance/friend	4:25	Female
83	Employee reported that two people have been shot in a club	Stranger	4:02	Male
84	Caller reported that an unconscious victim is bleeding profusely	Stranger	2:22	Male
85	Caller reported following a trail of blood that led to a deceased individual located outside their residence	Stranger	6:19	Female
86	Caller reported that a victim was shot in the head in her front yard	Stranger	5:23	Female
87	Caller reported discovering a deceased man in a park, with indications of either a gunshot wound to the head or severe physical assault	Stranger	5:19	Male
88	Caller reported a shooting at a church involving the injury of a child	Stranger	5:43	Female

Note. The reason for the call reflects the information based on only the 911 call rather than information that emerged later in the investigation.