

## Harmless Error Analysis: How Do Judges Respond to Confession Errors?

D. Brian Wallace · Saul M. Kassin

© American Psychology-Law Society/Division 41 of the American Psychological Association 2011

**Abstract** In *Arizona v. Fulminante* (1991), the U.S. Supreme Court opened the door for appellate judges to conduct a harmless error analysis of erroneously admitted, coerced confessions. In this study, 132 judges from three states read a murder case summary, evaluated the defendant's guilt, assessed the voluntariness of his confession, and responded to implicit and explicit measures of harmless error. Results indicated that judges found a high-pressure confession to be coerced and hence improperly admitted into evidence. As in studies with mock jurors, however, the improper confession significantly increased their conviction rate in the absence of other evidence. On the harmless error measures, judges successfully overruled the confession when required to do so, indicating that they are capable of this analysis.

**Keywords** Confessions · Harmless error · Judges · Fulminante

### Harmless Error Analysis: Judges' Performance with Confession Errors

In *Chapman v. California* (1967), the U.S. Supreme Court ruled that a constitutional violation of a defendant's rights was no longer sufficient to automatically overturn a conviction. Articulating the doctrine of harmless error, the Court reasoned that in some cases, a violation would not have an impact on the outcome of the case: "We conclude

that there may be some constitutional errors which in the setting of a particular case are so unimportant and insignificant that they may... be deemed harmless, not requiring the automatic reversal of the conviction" (p. 22). This ruling thus empowered federal appellate judges to employ a two-step decision process: (1) To determine if a trial error had occurred, and (2) if there was an error, to evaluate whether it prejudiced the case or instead was "harmless" beyond a reasonable doubt (for a history of the harmless error rule, see Bilaisis, 1983). At the same time, the *Chapman* Court noted that some errors were so prejudicial and so "basic to a fair trial" that violations were not subject to a harmless error analysis. Notably, the Court excluded from the scope of harmless error all infringements on the right to counsel, an impartial judge, and freedom from coerced confessions.

Twenty-four years later, in *Arizona v. Fulminante* (1991), the Supreme Court broke new ground. In that case, the Court ruled that erroneously admitted confession evidence is available for harmless error analysis. Even if the confession was deemed coercive and its admission at trial erroneous, the conviction could be maintained if other trial evidence was compelling enough that the jury would still have found the defendant guilty beyond a reasonable doubt. Asserting that the erroneously admitted coerced confession does not constitute a "structural defect" akin to a lack of competent counsel or an impartial judge, the court compared it to other types of mere "trial error" (e.g., improper jury instructions) and noted that it can be "quantitatively assessed in the context of other evidence presented in order to determine whether its admission was harmless beyond a reasonable doubt" (*Arizona v. Fulminante*, p. 280).

Over the years, a number of legal scholars have criticized the *Fulminante* ruling on constitutional grounds (Ogletree, 1991), on the belief that it will encourage

---

D. B. Wallace (✉) · S. M. Kassin  
Department of Psychology, John Jay College of Criminal Justice,  
City University of New York Graduate Center, New York,  
NY 10019, USA  
e-mail: dwallace@jjay.cuny.edu

increasingly coercive methods of interrogation (Kamisar, 1995), and on the argument that appeals court judges are ill-equipped by intuition to estimate the strength of the prosecutor's case and the cumulative or "harmless" nature of the confession in dispute (Mueller & Kirkpatrick, 1995). As Hirsch (2007) put it, "Fulminante should be reconsidered because, in cases involving confessions, judges are ill-equipped to assess harmless error. Judges, no less than juries, are tainted by the awareness that the defendant confessed" (p. 3). The *Fulminante* decision did not directly address the question of whether judges were capable of performing a harmless error analysis. Indeed, to our knowledge no empirical studies have ever directly tested this proposition with regard to harmless errors of any type, including the erroneous coerced confession.

If the ability of jurors to disregard inadmissible confession evidence is any indication, then some degree of pessimism is warranted when it comes to appellate court judges. In one study, Kassin and Sukel (1997) presented mock jurors with one of three versions of a murder trial transcript. In a low-pressure version, the defendant had confessed to police immediately upon questioning. In a high-pressure version, the defendant was in pain and interrogated aggressively by a detective who waved his gun in a menacing manner. In a control version, there was no confession in evidence. Presented with the high-pressure confession, participants reasonably judged the statement to be involuntary and reported that it did not influence their decisions. Nevertheless, this confession significantly boosted their conviction rate. This increase was exhibited even by those participants who were specifically admonished to disregard confessions they found to be coerced. Although people can recognize the coercive nature of certain interrogation tactics, they do not perceive a risk of false confessions (Leo & Liu, 2009).

In the only studies designed to assess the extent to which judges disregard information banned from evidence, Witrich, Guthrie, and Rachlinski (2005) tested 265 federal and state judges. In a series of experiments, judges were presented with vignettes describing a criminal or civil case and asked to make a substantive decision. Across scenarios, one of two versions was presented: A control version in which a decision could be based on all the facts contained in the scenario and a suppression version in which explicitly inadmissible information was added (i.e., a communication protected by attorney client privilege; settlement demand made in a pretrial offer; an alleged victim's sexual history; a defendant's prior criminal record; statements made by a defendant as part of a failed plea agreement; the outcome of an illegal search; a confession obtained in violation of *Miranda*). Overall, these studies showed that wittingly or unwittingly, judges, like juries, often do not disregard inadmissible information.

It appears that the U.S. Supreme Court and psychological research disagree over the question of whether appellate court judges can reasonably be expected to perform a harmless error analysis in cases containing an erroneously admitted coerced confession. This study was thus conducted with two goals in mind. First, we sought to test judges' decision making in cases that involve police-induced confessions. Research has shown that jurors are quasi-rational in their use of confession evidence: They perceive confessions as coerced when elicited through high-pressure tactics; yet, they use that evidence anyway as a basis for conviction. We tested judges within the same paradigm. Our second objective was to determine the extent to which judges are capable of making appropriate harmless error judgments in cases involving disputed confessions. Underlying the harmless error doctrine is the proposition that appellate courts render two judgments: (1) That the admission at trial of a coerced confession was erroneous and, if so, (2) That this error was harmless in its effect on the jury, not prejudicial, only when other evidence presented at trial formed a sufficient basis for conviction. We tested these two issues experimentally by varying the strength of non-confession evidence, as well as the presence of a confession and the pressure used to elicit it.

## Method

### Participants and Design

One hundred thirty-two judges from three states (MA = 39, PA = 32, MO = 60) were recruited by e-mail or at a judicial training conference. They were not compensated for their participation. Massachusetts and Missouri judges were recruited through e-mail and participated online via the web. Solicitation permission and e-mail list access were provided by administrative contacts in the two states. Solicitation e-mails described a questionnaire in which participants would evaluate "trial evidence" (the subject of confessions was not specifically cited). Pennsylvania judges completed the study in person at a conference preceding a lecture by one of the authors (again, instructions did not cite a focus on confessions).

Participants were randomly assigned to one of the six cells produced by a 2 (evidence strength: strong vs. weak)  $\times$  3 (confession: high-pressure, low-pressure, or no confession control) between-subjects factorial design.

### Procedure

Participants were presented with one of six versions of a fictional criminal case summary that we adapted by

drawing on facts taken from actual cases containing confession evidence. Entitled “State vs. Jason Lee Hill,” the case concerned the murder of a young woman. The defendant, Jason Hill, was charged with killing the victim, his next-door neighbor, late at night. All versions of the summary were approximately two pages long and included a description of the facts presented, the state’s allegations, and defense arguments. After reading the scenario, participants were asked to complete a questionnaire.

### Case Summary

In the baseline summary of State v. Jason Lee Hill, participants read about a murder case in which Teresa Brown was found bludgeoned to death in her home. Her jewelry and credit cards were missing. There was no evidence of rape, hence no expectation that DNA evidence would be found. The defendant, Jason Hill, lived next door to the victim. Neighbors reported witnessing several arguments between the defendant and victim during the months preceding the murder. The defendant became a suspect after one neighbor witnessed a man matching the defendant’s description leaving the victim’s house at midnight—at approximately the time that she was determined to have been killed.

**Confession Manipulation.** In the two confession-present conditions, the defendant, who was interrogated, provided a written confession to police that he later recanted. Modeled after a typical narrative confession (see Appleby, Hasel, Shlosberg, & Kassin, 2009; Garrett, 2010), this statement contained an admission of guilt, details about the crime that were generally consistent with the evidence, a motive statement in which the defendant explained that he was upset about the breakup of his relationship with the victim, and an apology to the victim’s family.

In the *high-pressure confession* condition, the defendant was questioned for 15 h before succumbing to produce the confession. The defendant claimed that his interrogators had coerced him by screaming, threatening the death penalty, and waving a gun, all while refusing to accept his claims of innocence. Importantly, it was noted that a videotape of the interrogation confirmed the defendant’s account. In the *low-pressure confession* condition, the defendant was questioned for only 30 min before producing a confession. The defendant claimed generally that he was coerced into confessing, but he described no specific behaviors or events and it was noted that the videotape did not confirm the defendant’s claims of coercive behavior on the part of the interrogators. In the *no-confession* condition, participants were told only that the defendant was

questioned by police, during which time he denied any involvement.

**Evidence Strength Manipulation.** To rule a confession error harmless, appellate courts must determine if there was sufficient evidence for the jury to have convicted the defendant even in the absence of the erroneous confession. Hence, two levels of evidence strength were created. In the *strong evidence* condition, participants were told that hair found on the victim was sent to a laboratory for testing and that the results were consistent with the defendant’s hair. They were also told that police searched the defendant’s home and found the victim’s jewelry there. In the *weak evidence* condition, participants were told that the hair found on the victim was tested against the defendant’s hair but that the results were inconclusive. In addition, they were told that the defendant’s home was searched but that the victim’s missing items were not found.

### Dependent Measures

Participants were asked first to evaluate the overall strength of evidence against the defendant on a 0–10 point Likert scale and to render a dichotomous guilty or not-guilty verdict. Those in the confession conditions were then asked to indicate their judgment of whether the defendant’s confession was voluntary or involuntary.

When it comes to confessions, there are two components to a harmless error judgment: (1) Whether the confession was coerced as opposed to voluntary, and, therefore, that its admission at trial was erroneous; and (2) in light of the other evidence, whether the erroneous confession was harmless as opposed to prejudicial, having no impact on the jury’s verdict.

Harmless error judgments were assessed in two ways. First, we employed an *implicit* measure, a purely counterfactual approach in which participants in the confession conditions were asked to render a verdict as if no confession had been introduced into evidence. Judges were asked to “assume that there was no confession in evidence. Would you find the defendant guilty, or not guilty, if this were the case?” This question was designed to parallel the second component of the decision-making process in which judges are essentially asked to predict what the verdict would have been without the confession. The “implicit harmless error” variable was then constructed by comparing these counterfactual responses in the two confession conditions to the baseline guilt responses of judges in the no-confession condition. This allowed for a direct comparison between participants who never received a confession (and thus could not be influenced by it) and those who did but were told to discount it (and are

potentially subject to its lingering influence). Using a second approach, both components of the harmless error decision were assessed with a two-phased *explicit* measure. First, participants were asked to determine whether the ruling to admit the confession at trial constituted a trial error. Second, they were told to assume that the confession was admitted in error and asked to indicate whether they would deem that error to be harmless.

Verdicts, judgments of voluntariness, and implicit and explicit harmless error were measured with both a dichotomous response and a confidence rating. This allowed for the construction of more sensitive continuous variables (ranging from  $-9.5$  to  $+9.5$ ) by computing the product of the dichotomous judgment (coded as  $-1$  or  $+1$ ) and confidence rating (1–10), and applying a  $\pm 0.5$  correction to account for the absence of zero values.

After completing the primary dependent measures, all participants were asked to indicate the total number of years they had served as a judge and the number of years, if any, served on an appellate court. They were also asked to estimate the number of cases they had tried that involved a confession.

## Results

Judges in our sample had served on the bench for an average of 11.1 years ( $SD = 7.7$ ; median = 10), with tenures ranging from 0 to 42 years. Seventeen judges (14%) also reported having served on an appellate court. On average, our sample reported having tried 39.3 cases involving a confession ( $SD = 73$ ; median = 15); 49 said they had never presided over a trial involving a confession. One judge, who reported having presided over 500,000 confession cases, was excluded from the calculation of this mean (excluding him from later analyses of substantive measures did not change the pattern of any results). When entered into the analyses soon to be reported, none of the demographic variables (judicial experience, appellate

experience, estimated number of confession cases, state of residence) affected the results; nor did the research modality (on paper vs. internet).

## Perceptions of Evidence Strength

Ratings of the strength of the evidence against the defendant (on a 1–10 scale; see Table 1) were subjected to a  $2 \times 3$  factorial ANOVA, with evidence strength (strong vs. weak) and confession (high-pressure, low-pressure, none) as between-subjects factors. This analysis confirmed that the evidence strength manipulation was successful, with the evidence perceived as significantly stronger and more incriminating in the strong evidence ( $M = 8.17$ ,  $SD = 1.51$ , 95% CI [7.78, 8.55]) than weak evidence condition ( $M = 5.61$ ,  $SD = 2.40$ , 95% CI [5.00, 6.22]),  $d = 1.08$ ,  $F(1, 120) = 67.75$ ,  $p < .001$ ,  $\eta_p^2 = 0.361$ .

As one would expect, the presence of a confession likewise influenced perceptions of evidence strength,  $F(1, 120) = 26.08$ ,  $p < .001$ ,  $\eta_p^2 = 0.303$ . The means were arrayed in the expected order: cases that included a confession resulting from a low-pressure interrogation yielding the highest evidence strength ratings ( $M = 8.05$ ,  $SD = 1.66$ , 95% CI [7.52, 8.58]), followed by high-pressure confession cases ( $M = 7.43$ ,  $SD = 1.88$ , 95% CI [6.78, 8.08]), and no-confession cases ( $M = 5.67$ ,  $SD = 2.57$ , 95% CI [4.95, 6.39]). Tukey's HSD post hoc tests showed that ratings were significantly lower in the no-confession condition than in the two confession-present conditions, which did not differ from each other (no-confession vs. high-pressure confession  $d = 0.74$ ).

A significant evidence  $\times$  confession interaction,  $F(1, 120) = 11.11$ ,  $p < .001$ ,  $\eta_p^2 = 0.156$ , revealed a substantial effect of the confession in the weak evidence condition: The no-confession case ( $M = 3.69$ ,  $SD = 1.49$ ) was seen as having weaker evidence than the high-pressure confession case ( $M = 6.25$ ,  $SD = 2.02$ ),  $t(40) = 4.72$ ,  $p < .001$ ,  $d = 1.50$ ,  $M_{diff} = 2.56$ , 95% CI [1.46, 3.66]. The high-pressure confession case ( $M = 6.25$ ,  $SD = 2.02$ ) was also

**Table 1** Perceptions of evidence strength, verdicts, and verdict-confidence scores

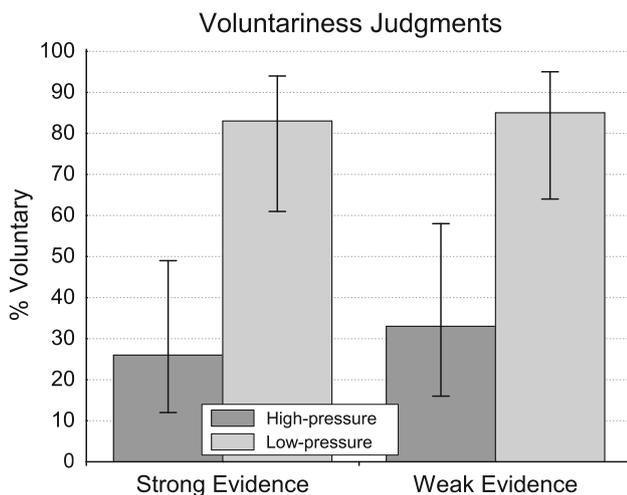
Evidence	Confession	<i>n</i>	Evidence strength		Verdict		Verdict-confidence	
			<i>M</i> ( <i>SD</i> )	95% CI	% Guilty	95% CI	<i>M</i> ( <i>SD</i> )	95% CI
Weak	None	26	3.69 (1.49)	[3.09, 4.29]	17	[07, 36]	-3.25 (4.08)	[-4.97, -1.53]
	High-pressure	16	6.25 (2.02)	[5.18, 7.32]	69	[44, 86]	2.94 (5.15)	[0.21, 5.67]
	Low-pressure	20	7.60 (1.67)	[6.82, 8.38]	95	[76, 99]	6.50 (3.92)	[4.67, 8.33]
Strong	None	25	7.72 (1.70)	[7.02, 8.42]	83	[64, 93]	4.88 (4.84)	[2.83, 6.92]
	High-pressure	19	8.42 (1.02)	[7.93, 8.91]	100	[83, 100]	7.71 (1.62)	[6.93, 8.49]
	Low-pressure	20	8.50 (1.57)	[7.77, 9.23]	100	[83, 100]	7.76 (2.13)	[6.74, 8.79]

CI confidence interval

seen as having weaker evidence than the low-pressure confession ( $M = 7.60$ ,  $SD = 1.67$ ),  $t(34) = 2.20$ ,  $p = .035$ ,  $d = 0.74$ ,  $M_{diff} = 1.35$ , 95% CI [0.10, 2.60]. In the strong evidence condition, however, ratings of evidence strength were already so high that there were no differences between the no-confession ( $M = 7.72$ ,  $SD = 1.70$ ) and high-pressure confession conditions ( $M = 8.42$ ,  $SD = 1.02$ ),  $t(42) = 1.59$ ,  $p = .120$ ,  $d = 0.48$ ,  $M_{diff} = 0.70$ , 95% CI [-0.19, 1.59]—which, in turn, did not differ from the low-pressure confession condition ( $M = 8.50$ ,  $SD = 1.57$ ),  $t(37) = 0.19$ ,  $p = .852$ ,  $d = 0.06$ ,  $M_{diff} = 0.1$ , 95% CI [-0.78, 0.94].

**Perceptions of Voluntariness**

Judges’ evaluations of the voluntariness of the confession (see Fig. 1) were subjected to a  $2 \times 2$  generalized linear model (for this analysis and others involving a DV that required the presence of a confession, the no-confession control condition was excluded from the design). The overall model was significant,  $\chi^2(3) = 23.68$ ,  $p < .001$ . For individual predictors, as one would expect, significantly more judges saw the confession as voluntary when it resulted from a low-pressure (84.2%) than a high-pressure interrogation (29.4%), Wald  $\chi^2(1) = 18.83$ ,  $p < .001$ , OR = 14.0, 95% CI [2.81, 69.8]. Importantly, evidence strength did not bias perceptions of the confession, with equal numbers finding the confession voluntary in the strong evidence (54.1%) and weak evidence (62.9%) conditions, Wald  $\chi^2(1) = 0.198$ ,  $p = .656$ , OR = 1.40, 95% CI [.32, 6.16]. The evidence strength  $\times$  confession pressure interaction was also not significant, Wald  $\chi^2(1) = 0.033$ ,  $p = .856$ , OR = 0.81, 95% CI [0.08, 7.98].



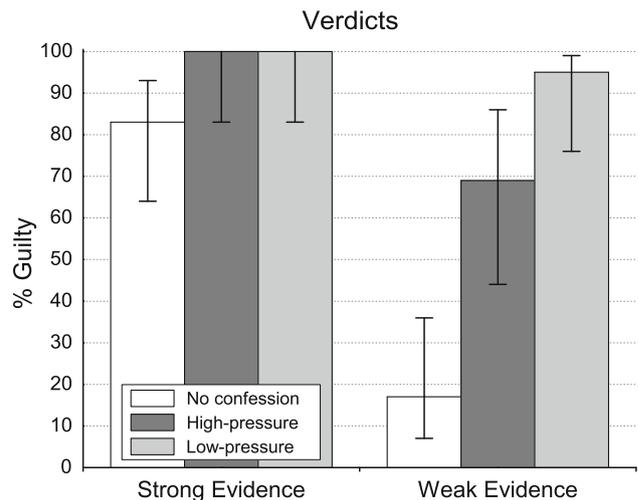
**Fig. 1** Percent of judges in each cell who perceived the confession to be voluntary. Error bars represent 95% confidence intervals

**Verdicts**

Dichotomous verdicts were analyzed as the outcome variable using the generalized linear model, with evidence strength and confession as predictors (see Fig. 2). The evidence strength  $\times$  confession interaction was not included in the model, as it induced a quasi-complete separation (maximum likelihood value could not be computed as the predictors perfectly predicted the interaction outcome). All confidence intervals for proportions were constructed using Newcombe and Altman’s (2000) method.

Overall, judges in the strong evidence condition were more likely to vote guilty than those in the weak evidence condition (93.5% vs. 56.7%), Wald  $\chi^2(1) = 21.26$ ,  $p < .001$ , OR = 30.00, 95% CI [7.07, 127.36]. Predictably, as well, those exposed to the low-pressure confession relative to the no-confession control were significantly more likely to vote for conviction (97.5% vs. 50%), Wald  $\chi^2(1) = 16.50$ ,  $p < .001$ , OR = 107.71, 95% CI [11.26, 1030.00]. Not predicted, but paralleling past research on mock juries, judges exposed to a high-pressure confession were also significantly more likely to find the defendant guilty (85.7% vs. 50%), Wald  $\chi^2(1) = 12.26$ ,  $p < .001$ , OR = 13.02, 95% CI [3.10, 54.79]. In fact, guilty verdicts were equally common in the low- and high-pressure confession conditions, Wald  $\chi^2(1) = 3.36$ ,  $p = .067$ , OR = 8.27, 95% CI [0.86, 79.14].

Inspection of the cell means shows that whereas conviction rates were uniformly high across cells in the strong evidence condition, verdicts varied dramatically with the confession manipulation in the weak evidence condition (see Fig. 2). In the absence of other evidence, far more guilty verdicts were produced by the low-pressure confession than in the no-confession control condition (95.5%



**Fig. 2** Percent of judges in each cell who rendered a guilty verdict. Error bars represent 95% confidence intervals

vs. 16.7%), Wald  $\chi^2(1) = 15.33$ ,  $p < .001$ , OR = 95.00, 95% CI [9.72, 928.31]. A similar effect was also obtained for the high-pressure confession, which produced significantly more guilty verdicts than in the no-confession condition (68.8% vs. 16.7%), Wald  $\chi^2(1) = 9.73$ ,  $p = .002$ , OR = 11.00, 95% CI [2.44, 49.63]. The higher conviction rate for the low-pressure confession (95.5%) than the high-pressure confession (68.8%) was not quite significant, Wald  $\chi^2(1) = 3.46$ ,  $p = .063$ , OR = 8.64, 95% CI [0.89, 83.75].

Verdict-confidence scores were subjected to a  $2 \times 3$  ANOVA (see Table 1). Results closely mirrored those of the generalized linear model on verdicts. Significant main effects for evidence strength,  $d = 0.90$ ,  $F(1, 120) = 44.38$ ,  $p < .001$ ,  $\eta_p^2 = 0.277$  and the confession were obtained,  $F(1, 120) = 30.95$ ,  $p < .001$ ,  $\eta_p^2 = 0.348$ . Qualifying these main effects was a significant evidence strength  $\times$  confession interaction,  $F(1, 120) = 8.45$ ,  $p < .001$ ,  $\eta_p^2 = 0.127$ . When the evidence was strong, the confession manipulation had no additional impact on perceptions of guilt—the result of a ceiling effect. When the evidence was weak, however, the confession manipulation had a strong effect: The low-pressure confession ( $M = 6.50$ ,  $SD = 3.92$ ) produced significantly higher verdict-confidence scores than high-pressure confession ( $M = 2.94$ ,  $SD = 5.15$ ),  $t(34) = 2.36$ ,  $p = .024$ ,  $d = 0.79$ , which were in turn higher than in the no-confession control condition ( $M = -3.25$ ,  $SD = 4.08$ ),  $t(38) = 3.23$ ,  $p < .001$ ,  $d = 1.37$ .

Across confession-present conditions, we conducted a binary logistic regression to determine the extent to which perceptions of evidence strength and voluntariness judgments were predictive of verdicts. Together, these two measures accounted for more than half of the variance in verdict scores,  $R_{\text{Nagelkerke}}^2 = 0.53$ . Individually, ratings of evidence strength were a significant predictor of verdict-confidence scores,  $b_{\text{evidence}} = 1.07$ ,  $p = .006$ , whereas ratings of voluntariness were not,  $b_{\text{voluntary}} = 0.43$ ,  $p = .748$ .

### Implicit Harmless Error

As described earlier, an implicit harmless error measure was constructed by comparing the verdict-confidence scores of participants not exposed to the potential error (i.e., those in the no-confession conditions) with the counterfactual responses of those who were exposed (i.e., those in the two confession conditions). If judges are able to set aside an erroneous confession, then those in the confession conditions should produce verdict-confidence scores similar to those found in the control condition when asked to ignore the confession. Thus, an accurate harmless error response would be reflected in a lack of significant main effects or interactions involving the confession

manipulation, while maintaining the effects of evidence strength (see Table 2).

A  $2 \times 3$  ANOVA indicated that this harmless error pattern was indeed obtained. There was no significant confession main effect,  $F(1, 120) = 1.53$ ,  $p = .222$ ,  $\eta_p^2 = 0.026$ . The no-confession, high-pressure, and low-pressure conditions all elicited equivalent verdict-confidence scores ( $M_s = 0.81, 0.21$ , and  $-1.24$ , respectively); none of the pairwise mean differences were significant, all  $t_s < 1.45$ ,  $p_s > .15$ . The evidence strength  $\times$  confession interaction was also no longer significant,  $F(1, 120) = 14.53$ ,  $p = .582$ ,  $\eta_p^2 = 0.009$ . The main effect for evidence strength, as expected, remained intact,  $F(1, 120) = 95.94$ ,  $p < .001$ ,  $\eta_p^2 = 0.453$ ,  $d = 1.77$ .

### Explicit Harmless Error

Explicit measures of harmless error were obtained only from participants in the two confession conditions and comprised two dichotomous responses: (1) whether admittance of the confession into evidence constituted a trial error, and (2) whether that error was harmless (see Table 2 for means and proportions for each). Each measure was subjected to a  $2 \times 2$  generalized linear model with evidence strength (strong vs. weak) and confession type (high-pressure vs. low-pressure) as between-subject factors.

On the question of whether the admission of the confession constituted trial error, the pattern of responses was quite rational. Coercion influenced error judgments, with judges more likely to find the high-pressure confession to be erroneous than the low-pressure confession (65.7% vs. 17.5%), Wald  $\chi^2(1) = 10.82$ ,  $p = .001$ , OR = 5.09, 95% CI [1.93, 13.44]. As one would hope, this judgment was not influenced by the evidence strength manipulation, Wald  $\chi^2(1) = 1.91$ ,  $p = .167$ , OR = 1.965, 95% CI [0.75, 5.12] or by the interaction of evidence strength  $\times$  confession, Wald  $\chi^2(1) = 0.62$ ,  $p = .433$ .

Judges were previously asked to evaluate whether the confession was voluntary or not, so we analyzed the consistency of their voluntariness and trial error judgments (one would expect that those who perceived the confession to be voluntary would not see its ruled admittance into evidence as erroneous). Results showed that judges were highly consistent across these two responses, Cronbach's  $\alpha = 0.89$ ; 90.3% of judges saw the confession as either coerced and erroneous or as voluntary and admissible. Judges who were inconsistent were equally split: four (5.5%) labeled the confession coerced but admissible; three (4.2%) saw the confession as voluntary but erroneous.

On the second-level question of whether admission at trial of the confession—assuming it was erroneous—was harmless, judges' responses were uniform: On this measure, there were no main effects for confession pressure,

**Table 2** Harmless error measures

Evidence	Confession	<i>n</i>	Implicit harmless error (verdict-confidence)		Implicit harmless error (verdict)		Trial error		Harmless error	
			<i>M</i> ( <i>SD</i> )	95% CI	% Guilty	95% CI	% Error	95% CI	% Harmless	95% CI
Weak	None	26	-3.25 (4.08)	[-4.97, -1.53]	17	[7, 36]	N/A		N/A	
	High-pressure	16	-5.50 (5.91)	[-8.63, -2.37]	19	[7, 43]	63%	[39, 82]	13%	[03, 36]
	Low-pressure	20	-5.70 (5.55)	[-8.29, -3.11]	15	[5, 36]	10%	[03, 29]	05%	[01, 23]
Strong	None	25	4.88 (4.84)	[2.83, 6.92]	83	[64, 93]	N/A		N/A	
	High-pressure	19	5.03 (5.43)	[2.42, 7.63]	84	[62, 94]	68%	[46, 85]	16%	[06, 38]
	Low-pressure	20	3.45 (5.48)	[0.82, 6.08]	79	[57, 91]	26%	[12, 49]	05%	[01, 25]

*CI* confidence interval

Wald  $\chi^2(1) = 0.71$ ,  $p = .401$ , OR = 1.68, 95% CI [0.50, 5.63], or evidence strength, Wald  $\chi^2(1) = 0.01$ ,  $p = .942$ , OR = 1.05, 95% CI [0.28, 3.89], and no significant interaction, Wald  $\chi^2(1) = 0.01$ ,  $p = .908$ . On this explicit measure of harmless error, the results were clear: 91% of judges saw the inclusion of an erroneous confession as prejudicial and this pattern was not limited to those who he deemed the confession an error in the previous question. Only 9% perceived the error to be harmless—even in conditions in which the other evidence was demonstrably compelling and sufficient as a basis for conviction.

## Discussion

Past research has shown that juries see a confession as a powerful form of evidence—so powerful that they do not discount it when it is legally and logically appropriate to do so. In this study, judges exhibited a similar pattern of responses. When presented with a summary of a murder case, they made voluntariness judgments in rational accord with the conditions of interrogation. They overwhelmingly perceived as coercive the confession that was the product of a 15-h interrogation in which the investigator threatened the suspect with the death penalty and a drawn firearm; they saw as voluntary the confession that resulted from a brief interrogation that was free of pressuring tactics.

The effect of confession on judges' perceptions of guilt, however, was a different matter. When other evidence in the case was so weak that it yielded only a 17% conviction rate, the percentage of guilty verdicts rose to 95% in the low-pressure confession condition. Clearly, a voluntary confession—even one that stands alone, uncorroborated by other evidence—provides for judges a sufficient basis of conviction. In addition, however, the results showed that when the totality of other evidence was weak, the conviction rate was increased fourfold to 69% by a confession that was elicited by high-pressure interrogation tactics that were judged coercive by more than two-thirds of

participants. Two aspects of these results are noteworthy. First, it is clear that perceptions of coercion and guilt are independent of one another. Regression analysis results showed that while perceptions of evidence strength predicted verdict scores, voluntariness ratings did not. On the individual level, 38% of the judges exposed to a high-pressure confession scenario and little other evidence both saw the confession as coerced and saw the suspect as guilty. Second, three out of ten judges believed that an interrogation was *not* coercive in which the detective brandished his sidearm and threatened the death penalty for more than 15 h over the suspect's repeated denials—tactics that were said to have been captured on videotape and not in dispute. In short, the results confirm research conducted with lay people indicating that confession evidence can be prejudicial. Although more research is needed to assess the scope of this finding, it appears that judges—as with mock jurors in previous research—do not fully discount a coerced confession.

The fact that judges exhibited the same bias that mock jurors have in past studies is interesting and adds to a growing body of research suggesting that some of the biases observed in lay decision makers are rooted in such basic social cognition processes that they tend to afflict professional judges as well. In one series of studies, Wistrich et al. (2005) found that judges, like juries, often fail to disregard information ruled inadmissible despite instructions to do so. In a second line of research, Lassiter, Diamond, Schmidt, and Elek (2007) found that judges exhibit the camera perspective bias consistently found among laypeople by which their judgments of voluntariness and coercion in a videotaped confession is influenced by whether the camera is focused on the interrogator, suspect, or both. In the present case, wherein judges, like juries, used confessions they saw as coerced in their verdicts, we believe that the basic phenomenon involves the well-established fundamental attribution error, the tendency for social perceivers to make dispositional attributions for a person's actions, taking behavior at face value

and underestimating the causal role of situational factors (Gilbert & Malone, 1995; Jones, 1990; Ross, 1977). In these studies, as in the present experiment, people typically infer a disposition (e.g., guilt) from an actor's behavior (e.g., confession) even while recognizing that the behavior was induced by the situation (e.g., coercive interrogation). Indeed, we would now argue that the process is so "fundamental" that no one is immune—even in their own areas of expertise.

Although participants exhibited the coerced confession effect, and in contrast to much speculation to the contrary, they performed in the legally prescribed manner on our measures of harmless error. The harmless error doctrine presumes that appellate courts can determine both that the admission of a coerced confession was erroneous and that the error had no effect on the jury when other evidence already formed a sufficient basis for conviction. We devised two measures of harmless error. On the implicit measure, a "correct" assessment required that judges who were exposed to a confession ignore it if they had to and produce a verdict similar to those who never knew that a confession existed. The results showed that participants did set aside the confession in their verdicts, exhibiting sensitivity only to the other non-confession evidence. On the explicit measure, judges were asked directly if the ruling to admit the disputed confession constituted trial error. In response to this question, they were sensitive to coercion, seeing the admissible confession as erroneous more in the high-pressure condition than in the low-pressure condition. Further, when asked to assume that the confession was admitted in error, judges in all conditions reported that the error was prejudicial, not harmless. In short, it appears that judges fully appreciated the extent to which juries would be impacted by confessions—even in the strong evidence condition that already contained sufficient grounds for conviction.

On the question of whether it is reasonable, in light of hindsight biases, to expect that appellate courts can perform the harmless error analysis prescribed in *Arizona v. Fulminante* (1991), this study suggests that judges indeed have the capacity to do so. Specifically, they determined both that the admission of a coerced confession was erroneous and that the error was prejudicial in its effect on the jury when the totality of other evidence did not form a sufficient basis for conviction. That judges are cognitively capable of this decision, not bound by hindsight and other biases, is an important finding that should help frame the ongoing debate concerning the wisdom of harmless error. It does not, however, address criticisms concerning other unintended consequences of a harmless error analysis in confession cases (e.g., the fear that it will encourage increasingly coercive methods of police interrogation). Absent more empirical research, involving an investigation

of the appellate record, it also does not indicate that appeals court judges necessarily perform this prescribed analysis in actual cases—where they may face pressure from prosecutors, the electorate, the news media, and others; and where their rulings are public, on the record, and have consequences for defendants, victims, trial judges, and the law.

Apart from the question of whether appellate judges can and do rule on cases involving coerced confessions in an appropriate manner, there is an additional problem. The harmless error doctrine—specifically, the notion that an erroneously admitted confession can prove harmless in cases containing other evidence sufficient to support a jury's conviction—rests on an assumption that the other evidence is independent of that confession. In fact, however, recent studies have shown that confessions can corrupt other evidence such as the judgments of polygraph examiners (Elaad, Ginton, & Ben-Shakhar, 1994) and latent fingerprint experts (Dror, Charlton, & Peron, 2006) as well as eyewitness identifications (Hasel & Kassin, 2009). Looking at Innocence Project case files, Kassin (2009) found that 78% of DNA exonerations involving a false confession also contained one or more other errors as basis for conviction (e.g., eyewitness misidentifications, informant statements, invalid or improper forensic science), compared to only 44% in non-confession cases. And in two-thirds of these multiple error cases, the confession was elicited before the other erroneous evidence was collected. In short, appellate courts must consider the additional possibility that the confessions they perceive to have been coerced may have corrupted the other evidence that otherwise leads them to see the confession as cumulative and, therefore, harmless.

## References

- Appleby, S., Hasel, L., Shlosberg, A., & Kassin, S. (2009). *False confessions: A content analysis*. Paper presented at the American Psychology-Law Society, San Antonio, TX.
- Arizona v. Fulminante, 11 S. Ct. 1246 (1991).
- Bilaisis, V. (1983). Harmless error: Abettor of courtroom misconduct. *Journal of Criminal Law & Criminology*, 74, 457–475. doi:10.2307/1143084.
- Chapman v. California, 386 U.S. 18 (1967).
- Dror, I. E., Charlton, D., & Peron, A. (2006). Contextual information renders experts vulnerable to making erroneous identifications. *Forensic Science International*, 156, 74–78. doi:10.1016/j.forsciint.2005.10.017.
- Elaad, E., Ginton, A., & Ben-Shakhar, G. (1994). The effects of prior expectations and outcome knowledge on polygraph examiners' decisions. *Journal of Behavioral Decision Making*, 7, 279–292.
- Garrett, B. L. (2010). The substance of false confessions. *Stanford Law Review*, 62, 1051–1119.
- Gilbert, D. T., & Malone, P. S. (1995). The correspondence bias. *Psychological Bulletin*, 117, 21–38. doi:10.1037/0033-2909.117.1.21.

- Hasel, L. E., & Kassin, S. M. (2009). On the presumption of evidentiary independence: Can confessions corrupt eyewitness identifications? *Psychological Science*, *20*, 122–126. doi:[10.1111/j.1467-9280.2008.02262.x](https://doi.org/10.1111/j.1467-9280.2008.02262.x).
- Hirsch, A. (2007). Confessions and harmless error: A new argument for the old approach. *Berkeley Journal of Criminal Law*, *12*, 1.
- Jones, E. E. (1990). *Interpersonal perception*. New York: Freeman.
- Kamisar, Y. (1995). On the “fruits” at *Miranda* violations, coerced confessions, and compelled testimony. *Michigan Law Review*, *93*, 979–1010. doi:[10.2307/1289987](https://doi.org/10.2307/1289987).
- Kassin, S. M. (2009). *On the doctrine of harmless error: Myths and misconceptions*. Division 41 Presidential Address presented at the American Psychological Association, Toronto.
- Kassin, S. M., & Sukel, H. (1997). Coerced confessions and the jury: An experimental test of the ‘harmless error’ rule. *Law and Human Behavior*, *21*, 27–46. doi:[10.1023/A:1024814009769](https://doi.org/10.1023/A:1024814009769).
- Lassiter, G. D., Diamond, S. S., Schmidt, H. C., & Elek, J. K. (2007). Evaluating videotaped confessions: Expertise provides no defense against the camera-perspective effect. *Psychological Science*, *18*, 224–226. doi:[10.1111/j.1467-9280.2007.01879.x](https://doi.org/10.1111/j.1467-9280.2007.01879.x).
- Leo, R. A., & Liu, B. (2009). What do potential jurors know about police interrogation techniques and false confessions? *Behavioral Sciences and the Law*, *27*, 381–399. doi:[10.1002/bsl.872](https://doi.org/10.1002/bsl.872).
- Mueller, C. B., & Kirkpatrick, L. C. (1995). *Modern evidence*. Boston: Little, Brown & Co.
- Newcombe, R. G., & Altman, D. G. (2000). Proportions and their differences. In D. G. Altman, D. Machin, T. N. Bryant, & M. J. Gardner (Eds.), *Statistics with confidence* (2nd ed., pp. 45–56). London: BMJ Books.
- Ogletree, C. J. (1991). *Arizona v. Fulminante*: The harm of applying harmless error to coerced confessions. *Harvard Law Review*, *105*, 152–175.
- Ross, L. (1977). The intuitive psychologist and his shortcomings: Distortions in the attribution process. *Advances in Experimental Social Psychology*, *10*, 174–221. doi:[10.1016/S0065-2601\(08\)60357-3](https://doi.org/10.1016/S0065-2601(08)60357-3).
- Wistrich, A. J., Guthrie, C., & Rachlinski, J. J. (2005). Can judges ignore inadmissible information? The difficulty of deliberately disregarding. *University of Pennsylvania Law Review*, *153*, 1251–1345. doi:[10.2307/4150614](https://doi.org/10.2307/4150614).