Research Article

THE SOCIAL PSYCHOLOGY OF FALSE CONFESSIONS: Compliance, Internalization, and Confabulation

Saul M. Kassin and Katherine L. Kiechel
Williams College

Abstract—*An experiment demonstrated that false incriminating evidence can lead people to accept guilt for a crime they did not commit. Subjects in a fast- or slow-paced reaction time task were accused of damaging a computer by pressing the wrong key. All were truly innocent and initially denied the charge. A confederate then said she saw the subject hit the key or did not see the subject hit the key. Compared with subjects in the slow-pace/no-witness group, those in the fast-pace/witness group were more likely to sign a confession, internalize guilt for the event, and confabulate details in memory consistent with that belief. Both legal and conceptual implications are discussed.*

In criminal law, confession evidence is a potent weapon for the prosecution and a recurring source of controversy. Whether a suspect's self-incriminating statement was voluntary or coerced and whether a suspect was of sound mind are just two of the issues that trial judges and juries consider on a routine basis. To guard citizens against violations of due process and to minimize the risk that the innocent would confess to crimes they did not commit, the courts have erected guidelines for the admissibility of confession evidence. Although there is no simple litmus test, confessions are typically excluded from trial if elicited by physical violence, a threat of harm or punishment, or a promise of immunity or leniency, or without the suspect being notified of his or her Miranda rights.

To understand the psychology of criminal confessions, three questions need to be addressed: First, how do police interrogators elicit self-incriminating statements (i.e., what means of social influence do they use)? Second, what effects do these methods have (i.e., do innocent suspects ever confess to crimes they did not commit)? Third, when a coerced confession is retracted and later presented at trial, do juries sufficiently discount the evidence in accordance with the law? General reviews of relevant case law and research are available elsewhere (Gudjonsson, 1992; Wrightsman & Kassin, 1993). The present research addresses the first two questions.

Informed by developments in case law, the police use various methods of interrogation—including the presentation of false evidence (e.g., fake polygraph, fingerprints, or other forensic test results; staged eyewitness identifications), appeals to God and religion, feigned friendship, and the use of prison informants. A number of manuals are available to advise detectives on how to extract confessions from reluctant crime suspects (Aubry & Caputo, 1965; O'Hara & O'Hara, 1981). The most popular manual is Inbau, Reid, and Buckley's (1986) *Criminal Interrogation and Confessions*, originally published in 1962, and now in its third edition.

Address correspondence to Saul Kassin, Department of Psychology, Williams College, Williamstown, MA 01267.

After advising interrogators to set aside a bare, soundproof room absent of social support and distraction, Inbau et al. (1986) describe in detail a nine-step procedure consisting of various specific ploys. In general, two types of approaches can be distinguished. One is *minimization*, a technique in which the detective lulls the suspect into a false sense of security by providing face-saving excuses, citing mitigating circumstances, blaming the victim, and underplaying the charges. The second approach is one of *maximization*, in which the interrogator uses scare tactics by exaggerating or falsifying the characterization of evidence, the seriousness of the offense, and the magnitude of the charges. In a recent study (Kassin & McNall, 1991), subjects read interrogation transcripts in which these ploys were used and estimated the severity of the sentence likely to be received. The results indicated that minimization communicated an implicit offer of leniency, comparable to that estimated in an explicit-promise condition, whereas maximization implied a threat of harsh punishment, comparable to that found in an explicit-threat condition. Yet although American courts routinely exclude confessions elicited by explicit threats and promises, they admit those produced by contingencies that are pragmatically implied.

Although police often use coercive methods of interrogation, research suggests that juries are prone to convict defendants who confess in these situations. In the case of *Arizona v. Fulminante* (1991), the U.S. Supreme Court ruled that under certain conditions, an improperly admitted coerced confession may be considered upon appeal to have been nonprejudicial, or "harmless error." Yet mock-jury research shows that people find it hard to believe that anyone would confess to a crime that he or she did not commit (Kassin & Wrightsman, 1980, 1981; Sukel & Kassin, 1994). Still, it happens. One cannot estimate the prevalence of the problem, which has never been systematically examined, but there are numerous documented instances on record (Bedau & Radelet, 1987; Borchard, 1932; Rattner, 1988). Indeed, one can distinguish three types of false confession (Kassin & Wrightsman, 1985): *voluntary* (in which a subject confesses in the absence of external pressure), *coerced-compliant* (in which a suspect confesses only to escape an aversive interrogation, secure a promised benefit, or avoid a threatened harm), and *coerced-internalized* (in which a suspect actually comes to believe that he or she is guilty of the crime).

This last type of false confession seems most unlikely, but a number of recent cases have come to light in which the police had seized a suspect who was vulnerable (by virtue of his or her youth, intelligence, personality, stress, or mental state) and used false evidence to convince the beleaguered suspect that he or she was guilty. In one case that received a great deal of attention, for example, Paul Ingram was charged with rape and a host of satanic cult crimes that included the slaughter of newborn babies. During 6 months of interrogation, he was hypno-

tized, exposed to graphic crime details, informed by a police psychologist that sex offenders often repress their offenses, and urged by the minister of his church to confess. Eventually, Ingram "recalled" crime scenes to specification, pleaded guilty, and was sentenced to prison. There was no physical evidence of these crimes, however, and an expert who reviewed the case for the state concluded that Ingram had been brainwashed. To demonstrate, this expert accused Ingram of a bogus crime and found that although he initially denied the charge, he later confessed—and embellished the story (Ofshe, 1992; Wright, 1994).

Other similar cases have been reported (e.g., Pratkanis & Aronson, 1991), but, to date, there is no empirical proof of this phenomenon. Memory researchers have found that misleading postevent information can alter actual or reported memories of observed events (Loftus, Donders, Hoffman, & Schooler, 1989; Loftus, Miller, & Burns, 1978; McCloskey & Zaragoza, 1985)—an effect that is particularly potent in young children (Ceci & Bruck, 1993; Ceci, Ross, & Toglia, 1987) and adults under hypnosis (Dinges et al., 1992; Dywan & Bowers, 1983; Sheehan, Statham, & Jamieson, 1991). Indeed, recent studies suggest it is even possible to implant false recollections of traumas supposedly buried in the unconscious (Loftus, 1993). As related to confessions, the question is, can memory of one's own actions similarly be altered? Can people be induced to accept guilt for crimes they did not commit? Is it, contrary to popular belief, possible?

Because of obvious ethical constraints, this important issue has not been addressed previously. This article thus reports on a new laboratory paradigm used to test the following specific hypothesis: The presentation of false evidence can lead individuals who are vulnerable (i.e., in a heightened state of uncertainty) to confess to an act they did not commit and, more important, to internalize the confession and perhaps confabulate details in memory consistent with that new belief.

METHOD

Participating for extra credit in what was supposed to be a reaction time experiment, 79 undergraduates (40 male, 39 female) were randomly assigned to one of four groups produced by a 2 (high vs. low vulnerability) × 2 (presence vs. absence of a false incriminating witness) factorial design.

Two subjects per session (actually, 1 subject and a female confederate) engaged in a reaction time task on an IBM PS2/Model 50 computer. To bolster the credibility of the experimental cover story, they were asked to fill out a brief questionnaire concerning their typing experience and ability, spatial awareness, and speed of reflexes. The subject and confederate were then taken to another room, seated across a table from the experimenter, and instructed on the task. The confederate was to read aloud a list of letters, and the subject was to type these letters on the keyboard. After 3 min, the subject and confederate were to reverse roles. Before the session began, subjects were instructed on proper use of the computer—and were specifically warned not to press the "ALT" key positioned near the space bar because doing so would cause the program to crash and data to be lost. Lo and behold, after 60 s, the computer supposedly ceased to function, and a highly distressed experimenter accused the subject of having pressed the forbidden key. All subjects initially denied the charge, at which point the experimenter tinkered with the keyboard, confirmed that data had been lost, and asked, "Did you hit the 'ALT' key?"

Two forensically relevant factors were independently varied. First, we manipulated subjects' level of *vulnerability* (i.e., their subjective certainty concerning their own innocence) by varying the pace of the task. Using a mechanical metronome, the confederate read either at a slow and relaxed pace of 43 letters per minute or at a frenzied pace of 67 letters per minute (these settings were established through pretesting). Two-way analyses of variance revealed significant main effects on the number of letters typed correctly (Ms = 33.01 and 61.12, respectively; $F[1, 71] = 278.93$, $p < .001$) and the number of typing errors made (Ms = 1.12 and 10.90, respectively; $F[1, 71] = 38.81$, $p < .001$), thus confirming the effectiveness of this manipulation.

Second, we varied the use of *false incriminating evidence*, a common interrogation technique. After the subject initially denied the charge, the experimenter turned to the confederate and asked, "Did you see anything?" In the false-witness condition, the confederate "admitted" that she had seen the subject hit the "ALT" key that terminated the program. In the no-witness condition, the same confederate said she had not seen what happened.

As dependent measures, three forms of social influence were assessed: compliance, internalization, and confabulation. To elicit *compliance*, the experimenter handwrote a standardized confession ("I hit the 'ALT' key and caused the program to crash. Data were lost") and asked the subject to sign it—the consequence of which would be a phone call from the principal investigator. If the subject refused, the request was repeated a second time.

To assess *internalization*, we unobtrusively recorded the way subjects privately described what happened soon afterward. As the experimenter and subject left the laboratory, they were met in the reception area by a waiting subject (actually, a second confederate who was blind to the subject's condition and previous behavior) who had overheard the commotion. The experimenter explained that the session would have to be rescheduled, and then left the room to retrieve his appointment calendar. At that point, the second confederate turned privately to the subject and asked, "What happened?" The subject's reply was recorded verbatim and later coded for whether or not he or she had unambiguously internalized guilt for what happened (e.g., "I hit the wrong button and ruined the program"; "I hit a button I wasn't supposed to"). A conservative criterion was employed. Any reply that was prefaced by "he said" or "I may have" or "I think" was not taken as evidence of internalization. Two raters who were blind to the subject's condition independently coded these responses, and their agreement rate was 96%.

Finally, after the sessions seemed to be over, the experimenter reappeared, brought the subjects back into the lab, reread the list of letters they had typed, and asked if they could reconstruct how or when they hit the "ALT" key. This procedure was designed to probe for evidence of *confabulation*, to determine whether subjects would "recall" specific details to

fit the allegation (e.g., "Yes, here, I hit it with the side of my hand right after you called out the 'A' "). The interrater agreement rate on the coding of these data was 100%.

At the end of each session, subjects were fully and carefully debriefed about the study—its purpose, the hypothesis, and the reason for the use of deception—by the experimenter and first confederate. Most subjects reacted with a combination of relief (that they had not ruined the experiment), amazement (that their perceptions of their own behavior had been so completely manipulated), and a sense of satisfaction (at having played a meaningful role in an important study). Subjects were also asked not to discuss the experience with other students until all the data were collected. Four subjects reported during debriefing that they were suspicious of the experimental manipulation. Their data were excluded from all analyses.

RESULTS AND DISCUSSION

Overall, 69% of the 75 subjects signed the confession, 28% exhibited internalization, and 9% confabulated details to support their false beliefs. More important, between-group comparisons provided strong support for the main hypothesis. As seen in Table 1, subjects in the slow-pace/no-witness control group were the least likely to exhibit an effect, whereas those in the fast-pace/witness group were the most likely to exhibit the effect on the measures of compliance ($\chi^2[3] = 23.84, p < .001$), internalization ($\chi^2[3] = 37.61, p < .001$), and confabulation ($\chi^2[3] = 18.0, p < .005$).

Specifically, although 34.78% of the subjects in the slow-pace/no-witness group signed the confession, indicating compliance, not a single subject in this group exhibited internalization or confabulation. In contrast, the two independent variables had a powerful combined effect. Out of 17 subjects in the fast-pace/witness cell, 100% signed a confession, 65% came to believe they were guilty (in reality, they were not), and 35% confabulated details to support their false belief (via chi-square tests, the differences in these rates between the slow-pace/no-witness control group and fast-pace/witness group were significant at ps $< .001$, .001, and .005, respectively).

Additional pair-wise comparisons revealed that the presence of a witness alone was sufficient to significantly increase the rates of compliant and internalized confessions, even in the slow-pace condition ($\chi^2[1] = 12.18, p < .005$, and $\chi^2[1] = 16.39, p < .001$). There were no sex differences on any measures (i.e., male and female subjects exhibited comparable confession rates overall, and were similarly influenced by the independent variables).

Table 1. *Percentage of subjects in each cell who exhibited the three forms of influence*

Form of influence	No witness		Witness	
	Slow pace	Fast pace	Slow pace	Fast pace
Compliance	35_a	65_b	89_{bc}	100_c
Internalization	0_a	12_{ab}	44_{bc}	65_c
Confabulation	0_a	0_a	6_a	35_b

Note. Percentages not sharing a common subscript differ at $p < .05$ via a chi-square test of significance.

The present study provides strong initial support for the provocative notion that the presentation of false incriminating evidence—an interrogation ploy that is common among the police and sanctioned by many courts—can induce people to internalize blame for outcomes they did not produce. These results provide an initial basis for challenging the evidentiary validity of confessions produced by this technique. These findings also demonstrate, possibly for the first time, that memory can be altered not only for observed events and remote past experiences, but also for one's own recent actions.

An obvious and important empirical question remains concerning the external validity of the present results: To what extent do they generalize to the interrogation behavior of actual crime suspects? For ethical reasons, we developed a laboratory paradigm in which subjects were accused merely of an unconscious act of negligence, not of an act involving explicit criminal intent (e.g., stealing equipment from the lab or cheating on an important test). In this paradigm, there was only a minor consequence for liability. At this point, it is unclear whether people could similarly be induced to internalize false guilt for acts of omission (i.e., neglecting to do something they were told to do) or for acts that emanate from conscious intent.

It is important, however, not to overstate this limitation. The fact that our procedure focused on an act of negligence and low consequence may well explain why the compliance rate was high, with roughly two thirds of all subjects agreeing to sign a confession statement. Effects of this sort on overt judgments and behavior have been observed in studies of conformity to group norms, compliance with direct requests, and obedience to the commands of authority. But the more important and startling result—that many subjects privately internalized guilt for an outcome they did not produce, and that some even constructed memories to fit that false belief—is not seriously compromised by the laboratory paradigm that was used. Conceptually, these findings extend known effects of misinformation on memory for observed events (Loftus et al., 1978; McCloskey & Zaragoza, 1985) and for traumas assumed to be buried in the unconscious (Loftus, 1993). Indeed, our effects were exhibited by college students who are intelligent (drawn from a population in which the mean score on the Scholastic Aptitude Test is over 1300), self-assured, and under minimal stress compared with crime suspects held in custody, often in isolation.

At this point, additional research is needed to examine other common interrogation techniques (e.g., minimization), individual differences in suspect vulnerability (e.g., manifest anxiety, need for approval, hypnotic susceptibility), and other risk factors for false confessions (e.g., blood alcohol level, sleep deprivation). In light of recent judicial acceptance of a broad range of self-incriminatory statements, increasing use of videotaped confessions at the trial level (Geller, 1993), and the U.S. Supreme Court's ruling that an improperly admitted coerced confession may qualify as a mere "harmless error" (*Arizona v. Fulminante*, 1991), further research is also needed to assess the lay jury's reaction to this type of evidence when presented in court.

Acknowledgments—This research was submitted as part of a senior honor's thesis by the second author and was funded by the Bronfman Science Center of Williams College.

REFERENCES

Arizona v. Fulminante, 59 U.S.L.W. 4235 (1991).

Aubry, A., & Caputo, R. (1965). *Criminal interrogation*. Springfield, IL: Charles C. Thomas.

Bedau, H., & Radelet, M. (1987). Miscarriages of justice in potentially capital cases. *Stanford Law Review, 40,* 21–179.

Borchard, E.M. (1932). *Convicting the innocent: Errors of criminal justice*. New Haven, CT: Yale University Press.

Ceci, S.J., & Bruck, M. (1993). Suggestibility of the child witness: A historical review and synthesis. *Psychological Bulletin, 113,* 403–439.

Ceci, S.J., Ross, D.F., & Toglia. M.P. (1987). Suggestibility of children's memory: Psycholegal implications. *Journal of Experimental Psychology: General, 116,* 38–49.

Dinges, D.F., Whitehouse, W.G., Orne, E.C., Powell, J.W., Orne, M.T., & Erdelyi, M.H. (1992). Evaluating hypnotic memory enhancement (hypermnesia and reminiscence) using multitrial forced recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 1139–1147.

Dywan, J., & Bowers, K. (1983). The use of hypnosis to enhance recall. *Science, 222,* 184–185.

Geller, W.A. (1993). *Videotaping interrogations and confessions* (National Institute of Justice: Research in Brief). Washington, DC: U.S. Department of Justice.

Gudjonsson, G. (1992). *The psychology of interrogations, confessions, and testimony*. London: Wiley.

Inbau, F.E., Reid, J.E., & Buckley, J.P. (1986). *Criminal interrogation and confessions* (3rd ed.). Baltimore, MD: Williams & Wilkins.

Kassin, S.M., & McNall, K. (1991). Police interrogations and confessions: Communicating promises and threats by pragmatic implication. *Law and Human Behavior, 15,* 233–251.

Kassin, S.M., & Wrightsman, L.S. (1980). Prior confessions and mock juror verdicts. *Journal of Applied Social Psychology, 10,* 133–146.

Kassin, S.M., & Wrightsman, L.S. (1981). Coerced confessions, judicial instruction, and mock juror verdicts. *Journal of Applied Social Psychology, 11,* 489–506.

Kassin, S.M., & Wrightsman, L.S. (1985). Confession evidence. In S.M. Kassin & L.S. Wrightsman (Eds.), *The psychology of evidence and trial procedure* (pp. 67–94). Beverly Hills, CA: Sage.

Loftus, E.F. (1993). The reality of repressed memories. *American Psychologist, 48,* 518–537.

Loftus, E.F., Donders, K., Hoffman, H.G., & Schooler, J.W. (1989). Creating new memories that are quickly accessed and confidently held. *Memory and Cognition, 17,* 607–616.

Loftus, E.F., Miller, D.G., & Burns, H.J. (1978). Semantic integration of verbal information into visual memory. *Journal of Experimental Psychology: Human Learning and Memory, 4,* 19–31.

McCloskey, M., & Zaragoza, M. (1985). Misleading postevent information and memory for events: Arguments and evidence against memory impairment hypotheses. *Journal of Experimental Psychology, 114,* 3–18.

Ofshe, R. (1992). Inadvertent hypnosis during interrogation: False confession due to dissociative state; misidentified multiple personality and the satanic cult hypothesis. *International Journal of Clinical and Experimental Hypnosis, 40,* 125–156.

O'Hara, C.E., & O'Hara, G.L. (1981). *Fundamentals of criminal investigation*. Springfield, IL: Charles C. Thomas.

Pratkanis, A., & Aronson, E. (1991). *Age of propaganda: The everyday use and abuse of persuasion*. New York: W.H. Freeman.

Rattner, A. (1988). Convicted but innocent: Wrongful conviction and the criminal justice system. *Law and Human Behavior, 12,* 283–293.

Sheehan, P.W., Statham, D., & Jamieson, G.A. (1991). Pseudomemory effects and their relationship to level of susceptibility to hypnosis and state instruction. *Journal of Personality and Social Psychology, 60,* 130–137.

Sukel, H.L., & Kassin, S.M. (1994, March). *Coerced confessions and the jury: An experimental test of the "harmless error" rule*. Paper presented at the biennial meeting of the American Psychology-Law Society, Sante Fe, NM.

Wright, L. (1994). *Remembering Satan*. New York: Alfred A. Knopf.

Wrightsman, L.S., & Kassin, S.M. (1993). *Confessions in the courtroom*. Newbury Park, CA: Sage.

(RECEIVED 12/21/94; ACCEPTED 2/22/95)