

## Coerced Confessions, Judicial Instruction, and Mock Juror Verdicts<sup>1</sup>

SAUL M. KASSIN<sup>2</sup>  
*Williams College*

LAWRENCE S. WRIGHTSMAN  
*University of Kansas*

The present research assessed whether judicial instruction can curb jurors' inappropriate use of coerced-confession evidence. In Experiment 1, subjects read an auto theft trial in which the defendant had confessed on his own initiative (no constraint), after an offer of leniency (positive constraint), or after a threat of punishment (negative constraint). Subjects then received an instruction that simply directed them to ignore a coerced confession (short form), another that additionally defined both positive and negative inducement as coercive and hence unreliable (long form), or no instruction at all. As previously reported (Kassin & Wrightsman, 1980), subjects fully discounted the negatively constrained confession but not the positively induced one which, although judged involuntary, produced a high percentage of guilty verdicts. Neither form of instruction significantly reduced this latter tendency. In Experiment 2, subjects read an assault case involving a voluntary or positively coerced confession and one of four types of instruction. The positive coercion bias was replicated. An instruction that stressed both the unreliability and unfairness of an induced confession decreased voluntariness judgments but failed to lower the conviction rate. The theoretical basis for and practical implications of this phenomenon are discussed, and future research directions are proposed.

Pretrial confessions have historically played a significant role in criminal law and procedure. Legal practitioners have long recognized that testimony about a defendant's confession, even if it is subsequently retracted, is a most damaging type of evidence (Cohn & Udolf, 1979). In fact, it is reported that innocent people are sometimes convicted on the basis of questionable confessions, even when the balance of the case against them is weak (Zimbardo, 1967).

<sup>1</sup>The authors would like to thank Kurt Feich for his role as experimenter in Study 1 and Tracey Quick for her role as experimenter in Study 2.

<sup>2</sup>Requests for reprints should be sent to Dr. Saul M. Kassin, Department of Psychology, Williams College, Bronfman Science Center, Williamstown, MA 01267.

It is therefore not surprising that police officials and detectives are explicitly trained in interrogation techniques that are designed to elicit admissions of guilt from suspects who are under detention (e.g., Inbau & Reid, 1962).

Reik (1966) and others have documented instances of people who, for a variety of reasons, willingly confess to crimes they did not commit. In order to safeguard against unduly abusive interrogatory procedures and to reduce the number of "false confessions," the legal system maintains that evidence of a prior confession is admissible in court *only* if it was voluntarily given and not the result of coercion. This precondition is based on the fact that coerced confessions are (a) unconstitutional, violating a defendant's fifth amendment right to due process, and (b) untrustworthy or unreliable, increasing the chances of a false admission. What defines a confession as involuntary? Interestingly, the legal concept of coercion has evolved over time. At first, confessions were excluded only if they were induced by physical force or brutality (e.g., *Brown v. Mississippi*, 1936). Soon thereafter, the Supreme Court acknowledged that coercion could be psychological, resulting simply from a threat of harm or punishment (e.g., *Chambers v. Florida*, 1940). Today, the Court accepts an even broader view, defining coercion not only as a threat, but as a promise of leniency and immunity from prosecution (e.g., *Greenwald v. Wisconsin*, 1968).

In cases that involve disputed confessions, a pretrial hearing is held without the jury's presence, in which a factfinder, usually the presiding judge, determines whether the confession was voluntary and, hence, whether it is admissible in court. How certain must the factfinder be about voluntariness in order to admit a confession as evidence? Some states had adopted the stringent criterion that voluntariness must be proven "beyond a reasonable doubt," while others had maintained that it may be proven by a mere "preponderance of the evidence." Since judges have been shown to translate the reasonable doubt standard to mean a 89% certainty and the preponderance standard to mean only a 61% certainty (Simon & Mahan, 1971), this difference in criteria is noteworthy. In *Lego v. Twomy* (1972), the Supreme Court resolved the discrepancy. On the ground that jurors will naturally discount coerced confessions as potentially unreliable, the Court ruled that the pretrial factfinder may determine voluntariness by the lesser standard.

Kassin and Wrightsman (1980) recently tested the Supreme Court's stated assumption about jurors' ability and willingness to discount a coerced confession. In two experiments, subjects read a transcript of a criminal trial in which testimony revealed that the defendant had confessed to the arresting officer either on his own initiative (no constraint), in response to an offer of leniency (positive constraint), in response to a threat of punishment (negative constraint), or not at all. The Court's assumption was only partially supported. Instead, the results supported the social psychological phenomenon

that observers attribute more responsibility and freedom to an individual for actions aimed at achieving a positive outcome than for equivalent actions aimed at avoiding punishment (Bramel, 1969; Kelley, 1971; Kite, 1964; Wells, 1980). That is, a threat-elicited confession was viewed as involuntary and, hence, resulted in a low rate of conviction. When the confession was induced by a promise of leniency or immunity, however, subjects responded inconsistently—they conceded that the defendant had confessed involuntarily, but voted guilty anyway. A theoretically interesting question to emerge from this study is, what accounts for the asymmetry between positive and negative constraints? Wells (1980) recently found that people assume reward to be a weaker form of behavioral inducement than punishment, even when the two are objectively equivalent in their strength (i.e., in the amount of compliance they produce). In the context of coerced confessions, subjects may have failed to discount the positively constrained confession simply because it was perceived as a relatively weak inducement. Unfortunately, Kassin and Wrightsman's (1980) data did not address this important issue.

From a practical standpoint, the Kassin and Wrightsman findings suggest that the courts, currently assuming that various types of coercion are perceived by jurors as equivalent, should exercise special caution in cases that involve claims of a positively induced confession. One possible strategy for curbing this bias might be through the use of judicial instruction. Indeed, some states (e.g., Massachusetts) advocate that even after a judge has admitted a confession as evidence, he or she must instruct the jury that they too should decide the voluntariness issue before rendering a verdict. Toward this end, two variations of an approved instruction are currently available to judges (Mathes & Devitt, 1965): (a) a short form that simply asks jurors to reject any confession that they believe to have been coerced, and (b) a longer version that also defines both positive and negative constraints as coercive and, further, explains that such elicited confessions are unreliable. On the negative side, many legal scholars (Frank, 1949; Kalven & Zeisel, 1966) and researchers (Sue, Smith, & Caldwell, 1973) have observed that judges' instructions often have little impact on jury decisions. In fact, some investigators have reported a "boomerang effect" whereby jurors who are admonished to ignore a critical piece of inadmissible evidence attach *greater* weight to that testimony than those who are not so instructed (Broeder, 1959; Wolf & Montgomery, 1977).

The present study was designed with three goals in mind—to replicate the Kassin and Wrightsman results, to test the effects of two ecologically valid forms of judicial instruction, and to measure directly whether the positive and negative types of coercion differ in the perceived strength of their inducement value. Accordingly, subjects read a detailed transcript of a criminal trial in which testimony revealed that the accused person had confessed to the arresting officer on his own initiative, in response to an offer of leniency, or in

response to a threat of punishment. Subjects also received from the judge a brief confession-related instruction, the more detailed instruction, or no instruction at all. Afterwards, subjects judged the voluntariness of the confession, rendered their verdicts, and answered other case-related questions.

## EXPERIMENT 1

### *Method*

#### *Subjects and Design*

A total of 170 introductory psychology students (80 male, 90 female) were randomly assigned to one of 10 cells produced by the 3 (confession constraint—none, positive, negative)  $\times$  3 (judge's instruction—none, short form, long form) factorial design that included an uninstructed, no-confession control group ( $n = 17$  per cell). The experiment was conducted in small groups that ranged in size from 4 to 6. Within each session, the 10 transcripts were randomly distributed.

#### *The Trial*

A 22-page transcript of a previously employed (Kassin & Wrightsman, 1979, 1980) criminal case entitled "United States vs. Ronald Oliver, 1972" served as the stimulus trial. Based on an actual case that was reenacted by law students (cf. Juhnke, Vought, Pyszczynski, Dane, Losure, & Wrightsman, 1979), the trial involves a young male defendant who is charged with transporting a stolen car across state lines. A used car salesman from whom the vehicle was stolen and the police officer who stopped the defendant for speeding and arrested him testified for the prosecution. The defendant testified on his own behalf that he had borrowed the car from an acquaintance without knowing that it had previously been stolen. The transcript thus consisted of the attorneys' opening remarks, the direct and cross examination of three witnesses, closing arguments, and the judge's charge.

*Confession manipulation.* Information about the confession and its surrounding circumstances was manipulated through the testimony of the arresting officer to whom the defendant had allegedly confessed. In the *no constraint* condition, the patrolman testified that "as soon as I mentioned it (the accusation) to the defendant, he confessed that he had in fact stolen the car from the Parker Ford Company." The *positive constraint* condition revealed that the defendant initially denied the accusation. Then the officer "told Mr. Oliver that if he confesses to the crime, he would be treated well during his detention and that the judge would surely be a lot easier on him—maybe even a suspended sentence." The defendant responded to that offer by confessing to

having stolen the car. In the *negative constraint* condition, the defendant again denied the accusation until the officer "told Mr. Oliver that if he does not confess to the crime, he would be treated very poorly during his detention and that the judge would surely be very hard on him—maybe even the maximum sentence." At that point, Ronald Oliver confessed. In the *no-confession* control group, all the evidence remained intact except the patrolman testified simply that "as soon as I mentioned it to the defendant, he flatly denied having anything to do with stealing a car."<sup>3</sup>

*Instruction manipulation.* In the *no-instruction* condition, subjects received in writing, at the close of the trial, a brief and general charge that made no mention of the confession or the issue of voluntariness. Instead, it merely outlined the jury's duties, reiterated the charge, and explicated the requirements of proof (i.e., that the defendant is presumed innocent until the prosecution proves guilt beyond a reasonable doubt).

The voluntariness instructions used in the study were taken from approved instruction manuals from which judges typically select portions of their charge to the jury (LaBuy, 1963; Mathes & DeVitt, 1965). The relevant portion of the *short-form* instruction read:

Before you retire to the deliberation room, I must ask that you consider this: You will recall that the prosecution introduced testimony from Patrolman Alvin Matheson who testified that while under arrest, the defendant, Ronald Oliver, confessed that he had stolen the car. However, before you accept this fact that the defendant confessed, you must first consider the circumstances and decide for yourselves whether he confessed voluntarily and without coercion.

If you decide that the confession was coerced or involuntary, then you should disregard this confession entirely and not allow it to influence your verdict. On the other hand, if you are convinced that Ron Oliver did confess freely and without coercion, then you should consider this fact as evidence in the case against the defendant.

The *long form* instruction *added* the following paragraph:

What does voluntary mean? If it appears from the evidence that the confession would not have been made without some threat of harm and punishment or some promise of immunity from prosecution or leniency in punishment, such a confession should not be considered voluntary. This is so because of the danger that an accused person might be persuaded to confess to things which are not true in an effort to avoid threatened punishment or to secure a promised reward.

Thus, in the latter version, the legal concept of voluntariness was defined and the basis for its unreliability was explicated.

<sup>3</sup>For the verbatim details of the constraint manipulations, see Kassin and Wrightsman (1980).

### *Procedure*

Upon entering, subjects were handed one of the 10 versions of the transcript. The title page read, "Enclosed is a transcript of a trial in which the defendant, Ronald Oliver, is charged with a violation of the Dire Act—transporting a stolen car across state lines. Read the trial carefully as if you were on the jury deciding the case. After considering the evidence, you will be asked to render your verdict." When all subjects in a session had completed their reading of the trial, they filled out a questionnaire individually and without deliberation. On it, experimental (i.e., confession) subjects first made a voluntariness judgment by answering, "Did Ron Oliver confess to Patrolman Matheson voluntarily and without coercion—yes or no?" and then indicated their confidence (0–8) in that decision. Next, all subjects rendered their verdicts (guilty or not guilty) and indicated their confidence (0–8) associated with that response. Note that voluntariness judgments always preceded verdicts in order to simulate the decision order that real jurors are faced with. Subjects then answered a number of other case-related questions. Specifically, they estimated the probability that the defendant had committed the crime (0 to 100% in multiples of 5) and the standard of proof they thought was necessary for conviction (i.e., "In this case, the defendant should be found guilty if there is at least a \_\_\_\_\_% chance that he committed the crime"), and they rated the extent to which their verdicts were influenced by the three witnesses' testimony (the used car salesman, the highway patrolman, the defendant) and the judge's closing instruction. Finally, to test whether the positive and negative constraints were perceived to differ in strength of inducement, experimental subjects rated "How much pressure did Patrolman Matheson exert on Ron Oliver to confess?". All of the above ratings were made on 9-point scales, where 0 = not at all and 8 = very much.

### *Results*

#### *Verdicts*

Overall, the stimulus trial elicited a perfect split in verdicts—85 subjects (50%) voted guilty and 85 (50%) voted not guilty. The overall pattern of judgments appears at the top of each cell in Table 1. It can be seen that across all levels of instruction, the conviction rates were 62.75% (32/51) in the no-constraint condition, 50.98% (26/51) in the positive-constraint condition, and 41.18% (21/51) in the negative-constraint condition. This difference between conditions was not quite significant ( $\chi^2(2) = 4.77, p < .10$ ). It can also be seen that the instruction manipulation had no effect on verdicts ( $\chi^2(2) < 1$ ).

In order to obtain a more sensitive between-groups comparison, a scalar variable was defined by combining subjects' verdicts with their 0–8 confidence

TABLE 1

PERCENTAGES OF GUILTY VERDICTS AND VOLUNTARINESS JUDGMENTS IN EXPERIMENT 1

Instruction conditions	Confession-constraint conditions			Marginals
	None	Positive	Negative	
None	58.82	52.94	41.18	50.98
	82.35	23.52	5.88	37.25
Short	64.70	58.82	41.18	54.90
	76.47	58.82	23.52	52.94
Long	64.70	41.18	41.18	49.02
	58.82	29.41	29.41	39.22
Marginals	62.75	50.98	41.18	
	72.55 <sub>a</sub>	37.25 <sub>b</sub>	19.61 <sub>c</sub>	

*Note.* Percentages on the top of each cell represent the proportion of guilty verdicts and those on the bottom represent the proportion of voluntary judgments ( $n = 17$  per cell). The conviction rate in the no-confession control group, not shown in the above table, was 35.29%. Numbers not sharing a common subscript differ at  $p < .05$  via  $\chi^2$  test.

levels. Specifically, positive confidence values were assigned to guilty verdicts and negative values to verdicts of not guilty. Scores thus ranged from -8 (maximum confidence in a not guilty verdict) to +8 (maximum confidence in a guilty verdict). Paralleling the dichotomous data, a 3 X 3 analysis of variance on these scores revealed a significant main effect for confession condition ( $F(2,144) = 2.95, p < .055$ ). Specifically, subjects were less likely to vote confidently for conviction in the negative constraint condition than in the no constraint condition ( $M_s = -.92$  &  $+ 2.08$ , respectively;  $p < .05$ ). The positive inducement ( $M = +.42$ ) fell between these extremes.<sup>4</sup>

The instruction manipulation had no effect either alone or in interaction with constraint type on these verdict-confidence scores. It did, however, have a significant effect on confidence per se ( $F(2,144) = 11.77, p < .01$ ). That is, subjects were generally more confident in their verdicts when they received the short-form instruction than no instruction at all ( $M_s = 6.51$  &  $5.55$ , respectively;  $p < .01$ ). The long-form instruction ( $M = 6.02$ ) did not differ significantly from either of these conditions.

<sup>4</sup> All post hoc comparisons were conducted via Newman-Keuls tests.

### *Voluntariness Judgments*

Overall, 66 out of the 153 experimental subjects (42.48%) judged the defendant's confession to be voluntary. The pattern of judgments appears at the bottom of each cell in Table 1. It can be seen that over levels of instruction, the percentages of "voluntary" judgments were 72.55 (37/51) under no constraint, 37.25 (19/51) under positive constraint, and 19.61 (10/51) under negative constraint. The overall difference among groups was highly significant ( $\chi^2(2) = 30.23, p < .001$ ). Subsequent  $\chi^2$  revealed that more voluntariness judgments were made under no constraint than positive constraint ( $\chi^2(1) = 12.83, p < .01$ ) which, in turn, produced more voluntariness judgments than the negative constraint ( $\chi^2(1) = 3.64, p < .05$ ). Again, the judge's instruction did not have a significant impact ( $\chi^2(2) = 3.04, n.s.$ ).

As before, a scalar variable was created by combining subjects' voluntariness judgments with their 0-8 confidence levels. Positive confidence values were assigned to voluntary judgments and negative values to involuntary judgments, so scores ranged from -8 to +8. An analysis of these scores corroborated the pattern of results for the dichotomous decisions. A main effect for constraint condition ( $F(2,144) = 19.44, p < .001$ ) showed that the defendant's confession was seen as more voluntary when given in the absence of any inducement ( $M = +2.57$ ) than under positive or negative constraints ( $M_s = -1.49$  and  $-4.02$ , respectively, both at  $p < .01$ ). As in the Kassin and Wrightsman (1980) study, the positive inducement was seen as significantly more voluntary than the negative one ( $p < .05$ ). Again, the judge's instruction played no role in these decisions.

### *Additional Measures*

No significant effects appeared on subjects' probability-of-commission estimates, their standard-of-proof requirements, or their ratings of how influential the three witnesses' testimony were.<sup>5</sup> An interesting main effect for the influence of judicial instruction ( $F(2,144) = 4.0, p < .02$ ), however, revealed that subjects who were instructed about the confession claimed to have been influenced by the judge more ( $M_s = 4.86$  in the short form and  $5.06$  in the long form) than those who were not so instructed ( $M = 3.82$ , both at  $p < .05$ ).

Finally, recall that to assess the perceived strength of the positive and negative types of inducement, experimental subjects indicated how much pressure

<sup>5</sup>Overall, subjects' mean standard of proof (i.e., how certain they should be in order to convict) was 83.27, a figure that is comparable to previously reported estimates of reasonable doubt (Kassin & Wrightsman, 1979, 1980; Simon & Mahan, 1971).

to confess they thought the arresting officer had exerted on the suspect. As it turned out, type of constraint produced a significant main effect ( $F(2,144) = 24.21, p < .01$ ). The negative inducement was seen as stronger than the positive one ( $M_s = 5.73$  and  $4.33$ , respectively,  $p < .01$ ) and both, in turn, were rated as stronger than the no-constraint situation ( $M = 2.80$ , both at  $p < .01$ ). Thus, although the promise of reward was viewed as some form of inducement (i.e., compared to none at all), it was in fact perceived as weaker than a threat of punishment. An important main effect for instruction ( $F(2,144) = 3.81, p < .05$ ) showed, moreover, that subjects did perceive greater pressure ( $p < .05$ ) to confess when they received the long-form instruction that explicitly defined coercion ( $M = 4.84$ ) than when they were not instructed ( $M = 3.69$ ). The short-form instruction had only a nonsignificant effect on this measure ( $M = 4.33$ ).

### Discussion

The present study provided some interesting insights into the impact that a pretrial confession has on potential jurors. At the most basic level, a comparison of verdicts in the no-confession control group with those in the confession-no constraint condition reaffirms the time-honored suspicion that evidence about a prior confession is often sufficient to elicit a conviction. More importantly, these results replicate very closely those obtained by Kassin and Wrightsman (1980). When confronted with a defendant who had confessed in response to a threat of harm or punishment, subjects clearly discounted the confession. In line with the Supreme Court's expectations (*Lego v. Twomy*, 1972), subjects viewed the negatively coerced confession as involuntary and they exhibited a relatively low rate of conviction (i.e., lower than in the no-constraint condition). However, when presented with testimony indicating that the defendant had confessed in response to a reward offer, subjects did not fully discount the confession. Under these circumstances, they decided that the confession was coerced but nevertheless used the evidence and voted guilty (i.e., compared to the no-constraint condition). In short, positively coerced confessions pose an evidentiary problem for the courts.

The primary question posed by this experiment was, can the often employed instruction effectively curb juror's use of the positively coerced confession? As it turned out, the instruction manipulation had two interesting effects (or lack thereof). First, compared to the uninstructed subjects, those who had received the elaborated instruction generally conceded that more pressure to confess was exerted on the suspect. Yet the instruction did not affect the practically important variable—judgments of voluntariness. Second, and perhaps more disturbing, these instructions also had no influence on verdicts despite the finding that subjects *claimed* that it *had* influenced their guilty-not guilty decisions. This pattern thus reveals a fascinating discrepancy between

the actual impact of the judge's charge and subjects' self-reported beliefs about that effect.<sup>6</sup> On the positive side, it should be noted that although the judge's instruction did not achieve its full purpose, it also did not produce the boomerang effect reported by others (Broeder, 1959; Wolf & Montgomery, 1977).

## EXPERIMENT 2

Overall, the present results suggest that judicial instruction does not mitigate the positive coercion bias. It is premature, however, to dismiss totally the potential utility of instructions, since they did affect certain dependent measures. Instead, it might be helpful to speculate about why they failed and to test how they could be improved.

Recall that there are two reasons why coerced confessions are deemed inadmissible as evidence—(1) they are unconstitutional and unfair to the accused, and (b) they are unreliable and untrustworthy. A close look at the elaborated (long-form) instruction shows that it emphasizes the latter and neglects to advance the “fairness” justification. Yet Kalven and Zeisel (1966), citing real world examples, suggested that “the jury may not so much consider the credibility of the confession as the impropriety of the method by which it was obtained” (p. 320). This observation implies that one promising approach to improving the elaborated instruction is to shift its emphasis—perhaps an argument which emphasizes what Kalven and Zeisel call the “sympathy hypothesis” rather than the “credibility hypothesis” might prove effective.<sup>7</sup>

A second experiment was conducted in order to (a) test the conceptual replicability of our results using a different stimulus trial, and (b) compose and test a “sympathy instruction,” i.e., one that makes salient the unconstitutionality and unfairness of a coerced confession. Specifically, subjects read a hypothetical assault case involving either an unconstrained or positively coerced

<sup>6</sup>Although beyond the scope of this paper, the question raised by this self report-behavior discrepancy is whether it results simply from subjects' ignorance about the causes of their own behavior (Nisbett & Wilson, 1977) or whether there is a more dynamic, perhaps motivational reason for this “bias.” One possibility is a self presentation explanation—that subjects reported an instruction effect because they believed the judge's instruction *should* be influential. A related alternative interpretation of the self report effect is that subjects simply invoked their normative theories about human behavior.

<sup>7</sup>An alternative strategy for improving the instruction, though not the focus of Experiment 2, might be to bolster the credibility argument. Since positive forms of coercion appear to be problematic partly because people underestimate their power to induce compliance-in-general and confessions in particular, judges might attempt to deal with this misperception. Wells (1980), for example, found that this bias can be eliminated by providing subjects with information about the actual base rates for compliance. An instruction that is designed to convey to jurors the fact that many people confess in response to a reward might conceivably lower the credibility and hence persuasiveness of that elicited confession.

confession, and received either no special instruction, the standard credibility instruction, a sympathy instruction, or one that encompassed both arguments. It was hypothesized that although the positive coercion bias would appear, it would be mitigated by the delivery of a sympathy instruction.

### *Method*

One hundred eighty-eight introductory psychology students (93 male, 95 female) were randomly assigned to one of 10 groups ( $n = 18$  to 20 per cell) produced by a 2 (confession: no constraint, positive constraint)  $\times$  4 (instruction: none, sympathy, credibility, both) factorial design that included a confession-negative constraint and a no-confession control group. As in Experiment 1, the 10 transcript versions were distributed in small group settings.

An 18–19 page adaptation of the Adams-Zemp assault case, originally created by Walker, Thibaut, and Andreoli (1972), served as the stimulus trial. The transcript was written with the pre-scaled facts provided by Walker et al. (1972) and presented as a criminal trial entitled “Adams v. Illinois.” In this case, Samuel Adams is charged with assault for stabbing and seriously injuring Michael Zemp with a piece of broken glass during a heated argument in a tavern. The defense claimed that Adams, feeling threatened and endangered, had acted in self defense. The entire transcript contained the examination of seven witnesses, including the defendant and the victim. Pretesting revealed that the no-confession version of the trial elicited a relatively low (22%) rate of conviction.

As in the first experiment, information about the confession and the circumstances surrounding it was introduced through the testimony of an arresting police officer to whom the defendant had allegedly confessed. In the confession evidence conditions (i.e., no constraint, positive constraint, and the negative constraint control group), the officer testified that when he questioned Adams about the stabbing, the defendant confessed “that he had stabbed Michael Zemp without provocation.” In the no-confession control group, the officer testified that in response to his inquiry, the defendant said “that he was afraid Mr. Zemp was about to attack him.” The coercion manipulations were nearly identical to those of the first experiment. That is, Adams confessed either on his own initiative, in response to a threat, in response to an offer, or not at all.

The transcript concluded with one of four versions of the judge’s instruction to the jury. Subjects in the no-instruction condition received a general charge that made no special reference to the confession issue. Subjects in the credibility instruction condition read the long-form instruction that was employed in Experiment 1—they were thus informed of the danger that an accused person might be persuaded to confess to acts he/she did not commit in an effort to avoid punishment or secure a reward (e.g., a suspended sentence). In the sympathy condition, the basic confession/voluntariness instruction was

embellished as follows: "... because it is constitutionally unfair to an accused person who is under arrest for an officer of the law to pressure him through threats or trick him through offers of immunity into admitting to something against his will. Such tactics violate the individual's constitutional right to due process of law." A fourth instruction condition was included in which the credibility and sympathy arguments were combined.

After reading a version of the transcript, subjects completed a questionnaire individually and without deliberation. The major dependent measures from Experiment 1 were included: voluntariness judgments and verdicts were followed by reasonable doubt and probability-of-commission estimates, and witness ratings. Additionally, subjects in the nine confession cells were asked, in two separate questions, "How many out of 100 truly *guilty (innocent)* people do you think would have confessed to the arresting officer in this case?" This question was designed to assess lay beliefs about the normativeness of true and false confessions under the different instructions and constraint circumstances. Finally, all subjects rated, on a 9-point scale, how fairly the defendant was treated upon his arrest. This question was designed to assess subjects' attitudes along the sympathy dimension.

## Results

### Verdicts

An overall significant difference ( $\chi^2(3) = 13.80, p < .005$ ) between the four no-instruction groups corroborates the pattern of results repeatedly obtained for the Ron Oliver case (Kassin & Wrightsman, 1980). That is, the conviction rate was highest in the confession-no constraint group (.667) and lowest in the no-confession group (.105). Moreover, whereas the negatively coerced confession (.368) did not significantly increase the proportion of guilty verdicts (i.e., compared to the no-confession group,  $\chi^2(1) = 2.33, n.s.$ ), the positively induced confession (.556) did have that damaging effect ( $\chi^2(1) = 6.05, p < .01$ ). A one-way analysis of variance on the combined verdict-confidence measure (-8 to +8, as in Experiment 1) yielded a similarly significant difference ( $F(3,70) = 4.86, p < .005$ )—post hoc tests showed that these conviction scores were increased significantly only by the unconstrained and positively constrained confessions (both at  $p < .01$ ).

The pattern of verdicts for the full  $2 \times 4$  design appears at the top of each cell in Table 2. Note that across all types of instruction, the conviction rates were .573 in the unconstrained confession condition and .507 in the positive inducement condition ( $\chi^2(1) < 1$ ). It can also be seen that the instruction manipulation did not significantly affect verdicts either. A two-way analysis of variance on the verdict-confidence measure similarly revealed no significant

TABLE 2

PERCENTAGES OF GUILTY VERDICTS AND VOLUNTARINESS JUDGMENTS IN EXPERIMENT 2

Instruction conditions	Constraint conditions		Marginals
	None	Positive	
None	66.67	55.56	61.12
	100.00	66.67	83.33 <sub>a</sub>
Credibility	50.00	63.16	56.76
	72.22	57.89	64.86 <sub>ab</sub>
Sympathy	63.16	47.37	55.27
	78.95	47.37	63.16 <sub>ab</sub>
Combined	50.00	36.84	43.59
	75.00	21.05	48.72 <sub>b</sub>
Marginals	57.33	50.67	
	81.33 <sub>a</sub>	48.00 <sub>b</sub>	

*Note.* Percentages at the top of each cell represent the proportion of guilty verdicts and those on the bottom represent the proportion of voluntariness judgments. Numbers not sharing a common subscript differ at  $p < .05$  via  $\chi^2$  test. In the negative constraint (no instruction) confession group, the percentages of guilty verdicts was 36.8 and voluntariness judgments 52.6. The conviction rate in the no-confession group was .105.

effects. In short, subjects were uniformly as likely to vote guilty for a defendant who confessed in response to a positive form of inducement as they were for one who confessed on his accord.

*Voluntariness Judgments*

An inspection of the differences between the three no-instruction confession groups ( $\chi^2(2) = 10.95, p < .005$ ) shows that whereas 100% of our subjects judged the unconstrained confession to be voluntary, only 66.67% found the positively induced confession to have been voluntarily given ( $\chi^2(1) = 5.0, p < .01$ ) and only 52.63% decided as such for the negatively coerced confession ( $\chi^2(1) = 9.33, p < .01$ ). In contrast to Experiment 1, the latter two groups did not differ significantly from each other ( $\chi^2(1) < 1$ ).

Table 2 (the bottom number of each cell) presents the pattern of voluntariness judgments for the full 2 X 4 design. It can be seen, first, that subjects clearly

did distinguish between the unconstrained and positively constrained confessions (81.33% v. 48% judged voluntary across types of instruction;  $\chi^2(1) = 18.23, p < .001$ ). Second, the type of instruction delivered by the judge affected the overall frequency of voluntariness judgments ( $\chi^2(3) = 11.36, p < .01$ ). Subsequent tests indicated a lower proportion of voluntariness decisions with the combined instruction (.487) than with no instruction (.83),  $\chi^2(1) = 9.90, p < .005$ . Neither the sympathy nor credibility instruction *alone* (.632 and .649, respectively) significantly reduced the perception of voluntariness ( $\chi^2(1) = 3.82$  and  $3.23$ , respectively). An analysis of the voluntariness-confidence measure corroborated the above result. Main effects were obtained for constraint ( $F(1,142) = 24.65, p < .001$ ) and type of instruction ( $F(3,142) = 4.82, p < .003$ ) the latter resulting again reflecting a significant difference ( $p < .01$ ) between the no-instruction and combined instruction conditions. The interaction between constraint and instruction was not significant ( $p < .50$ ).

### *Additional Measures*

As in Experiment 1, no significant differences were obtained for probability-of-commission estimates or for subjects' ratings of how influential the seven witnesses were. However, a main effect for constraint was obtained for subjects' definitions of reasonable doubt ( $F(1,142) = 7.50, p < .01$ ). Specifically, subjects reported a higher standard of proof in the positively constrained than unconstrained confession condition. Compared to previously obtained estimates (Kassin & Wrightman, 1979, 1980; Simon & Mahan, 1971), it appears not that the positive constraint subjects adopted an unduly stringent standard ( $M = 88.0$ ), but that subjects in the unconstrained confession condition asserted a uniquely lenient estimate ( $M = 78.07$ ).

Recall that subjects were asked to estimate the percentages of truly guilty and innocent people who could confess under the circumstances of the case they had read. Although the independent variables had no effect on these responses, the overall estimates proved interesting. Specifically, subjects projected that confessions would be given by 45.67% of all guilty people and, surprisingly, by 35.58% of all innocent people. This latter figure suggests that many subjects believe that an accused person might indeed confess to acts he or she did not commit.

Finally, a significant main effect for instruction on the fairness measure ( $F(3,142) = 2.71, p < .05$ ) was obtained. Subsequent tests revealed, quite simply, that the defendant was seen as having been treated more unfairly ( $p < .05$ ) in the sympathy and combined instruction conditions ( $M_s = 4.12$  and  $3.98$ , respectively) than in either the credibility or no instruction sets ( $M_s = 6.55$  and  $6.20$ , respectively). The main effect for constraint, although in the expected direction, was not quite significant ( $F(1,142) = 3.18, p < .10$ ).

*Discussion*

The major results of Experiment 2 may be recapitulated as follows:

1. The positive coercion bias was replicated. That is, even though subjects acknowledged that the positively constrained confession was relatively involuntary (and even though they asserted a relatively high standard of proof as necessary for conviction), they did not discount that evidence when rendering their verdicts.

2. The sympathy appeal, present in the sympathy and combined instruction conditions, significantly increased subjects' perception that the defendant was unfairly treated. Of the four levels of instruction, however, only the combined condition was even partially effective—it successfully lowered the frequency of voluntary judgments but it failed to lower the conviction rate.

It was noted earlier that the legal definition of coercion has essentially progressed through three stages: (a) negative/physical pressure, (b) negative/physical or psychological pressure, and (c) positive or negative/physical or psychological pressure. Within this historical guideline, the present research suggests that the layperson, the potential juror, is, in a sense, fixated at the second stage. Subjects readily acknowledged that a mere threat, even without signs of physical brutality, is coercive enough to elicit an unreliable confession. Yet they seem unable or unwilling to excuse a defendant whose admission is induced by the promise of immunity or reward—this despite the recognition that accused persons might plausibly confess to acts they did not commit. The basis for this anomaly is clarified somewhat by subjects' responses to the strength-of-pressure question in Experiment 1 which showed that despite their assumed equivalence, subjects viewed the promise of reward as simply a weaker form of inducement than a threat of punishment. This interpretation is supported by Wells' (1980) finding that "the principal locus of this error is in the reward-contingency conditions where subjects underestimated the proportion of compliance rather than in the punishment-contingency conditions where the subjects' estimates of compliance were reasonably accurate" (p. 59).

*Judicial Instruction*

The present research was designed primarily to assess the utility of judicial instruction as a mechanism for reducing the positive coercion bias. As it turned out, our results on this issue are mixed. Experiment 1 demonstrated quite clearly that the currently available forms of the instruction are ineffective. Instruction effects were obtained on certain dependent variables, but not on the two practically important judgments. Experiment 2 revealed that although no instruction significantly affected verdicts, the dual instruction (i.e., emphasizing both the unfairness and the unreliability of an induced confession) did signifi-

cantly alter subjects' voluntariness judgments. The potential importance of this latter result should not be overlooked. After all, in numerous state courts, jurors are explicitly instructed to consider or ignore a confession in their verdicts on the basis of their prior decisions about voluntariness. Indeed, our own data suggest that the two decisions were highly interrelated. Across all confession groups, the correlation between voluntariness-confidence scores and verdict-confidence scores was .55 in Experiment 1 ( $N = 153, p < .001$ ) and .48 in Experiment 2 ( $N = 169, p < .001$ ).

### Conclusions

According to police and court statistics, pretrial confessions—both disputed and undisputed—are prevalent in criminal cases (Kalven & Zeisel, 1966; Zimbardo, 1967). For years, legal scholars and philosophers have debated about the *actual* probative value of voluntary and involuntary confessions. The present research suggests that it is equally important to investigate jurors' *beliefs* about their probative value.

From a practical standpoint, the results of Experiments 1 and 2 are discouraging because they collectively suggest that positively coerced confessions are a problem and that judicial instruction might not be an effective solution. Still, at least two issues should be addressed before drawing any firm negative conclusions. First, can the timing of the confession-voluntariness instruction mediate its impact? In the present research, the judge's instruction followed the presentation of evidence, as is common practice in most courts. Yet two studies have shown that certain types of judicial instruction affect mock jurors' decisions only when they *precede* the evidence (Elwork, Sales, & Alfini, 1977; Kassin & Wrightsman, 1979). Perhaps this finding holds for the confession instruction. It is thus possible that subjects had tentatively decided on their *verdicts* before the instruction was delivered and so were subsequently influenced only in their *voluntariness* judgments.

A second particularly important class of questions is similarly a basis for mere speculation—does the positive coercion bias persist or disappear after a jury deliberates and, if it persists, is judicial instruction any more effective a device at this group level? Thus far, our own research has focused on the judgments and beliefs of the individual, nondeliberating juror. Yet Kaplan and Miller (1978) have reported that jury discussion may correct for certain pretrial and mid-trial biases. Whether this group decision-making phenomenon operates in this instance remains to be seen.

### REFERENCES

- Bramel, D. Determinants of beliefs about other people. In J. Mills (Ed.), *Experimental social psychology*. New York: Macmillan, 1969.

- Broeder, D. The University of Chicago jury project. *Nebraska Law Review*, 1959, **38**, 744-760.
- Brown v. Mississippi*, 297 U.S. 278, 1936.
- Chambers v. Florida*, 309 U.S. 227, 1940.
- Cohn, A., & Udolf, R. *The criminal justice system and its psychology*. New York: Van Nostrand Reinhold, 1979.
- Elwork, A., Sales, B. D., & Alfini, J. J. Juridic decisions: In ignorance of the law or in light of it? *Law and Human Behavior*, 1977, **1**, 163-189.
- Frank, J. *Courts on trial: Myth and reality in American justice*. Princeton, NJ: Princeton University Press, 1949.
- Greenwald v. Wisconsin*, 390 U.S. 519, 1968.
- Inbau, F. E., & Reid, J. E. *Criminal interrogation and confessions*. Baltimore: Williams & Wilkins, 1962.
- Juhnke, R., Vought, C., Pyszczyński, T. A., Dane, F. C., Losure, B. D., & Wrightsman, L. S. Effects of presentation mode upon mock jurors' reactions to a trial. *Personality and Social Psychology Bulletin*, 1979, **5**, 36-39.
- Kalven, H., & Zeisel, H. *The American jury*. Boston: Little, Brown, & Co., 1966.
- Kaplan, M. F., & Miller, L. E. Reducing the effects of juror bias. *Journal of Personality and Social Psychology*, 1978, **36**, 1443-1455.
- Kassin, S. M., & Wrightsman, L. S. On the requirements of proof: The timing of judicial instruction and mock juror verdicts. *Journal of Personality and Social Psychology*, 1979, **37**, 1877-1887.
- Kassin, S. M., & Wrightsman, L. S. Prior confessions and mock juror verdicts. *Journal of Applied Social Psychology*, 1980, **10**, 133-146.
- Kelley, H. H. *Attribution in social interaction*. Morristown, NJ: General Learning Press, 1971.
- Kite, R. W. *Attribution of causality as a function of the use of reward and punishment*. Ph.D. dissertation, Stanford University, 1964.
- LaBuy, W. J. *Jury instructions in federal criminal cases*. St. Paul, MN: West Publishing Co., 1963.
- Lego v. Twomy*, 30 L.Ed. 2nd 618, 1972.
- Mathes, W. C., & DeVitt, E. J. *Federal jury practice and instructions*. St. Paul, MN: West Publishing Co., 1965.
- Nisbett, R. E., & Wilson, T. D. Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 1977, **84**, 231-259.
- Reik, T. *The compulsion to confess*. New York: John Wiley, 1966.
- Simon, R. J., & Mahan, L. Quantifying burdens of proof: A view from the bench, the jury, and the classroom. *Law and Society Review*, 1971, February, 319-330.
- Sue, S., Smith, R. E., & Caldwell, C. Effects of inadmissible evidence on the decisions of simulated jurors: A moral dilemma. *Journal of Applied Social Psychology*, 1973, **3**, 345-353.
- Walker, L., Thibaut, J., & Andreoli, V. Order of presentation at trial. *The Yale Law Journal*, 1971, **82**, 216-226.
- Wells, G. L. Assymetric attributions for compliance: Reward vs. punishment.

*Journal of Experimental Social Psychology*, 1980, **16**, 47-60.

Wolf, S., & Montgomery, D. A. Effects of inadmissible evidence and level of judicial admonishment to disregard on the judgments of mock jurors. *Journal of Applied Social Psychology*, 1977, **7**, 205-219.

Zimbardo, P. G. The psychology of police confessions. *Psychology Today*, 1967, June 1(2), 17-27.