

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

Faculty Publications, Department of Psychology

Psychology, Department of

2011

What are we studying? Student Jurors, Community Jurors, and Construct Validity

Stacie R. Keller

Richard L. Weiner

Follow this and additional works at: <https://digitalcommons.unl.edu/psychfacpub>



Part of the [Applied Behavior Analysis Commons](#), [Courts Commons](#), and the [Human Factors Psychology Commons](#)

This Article is brought to you for free and open access by the Psychology, Department of at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Faculty Publications, Department of Psychology by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

What are we studying? Student Jurors, Community Jurors, and Construct Validity

Stacie R. Keller, M.L.S. & Richard L. Wiener, Ph.D., M.L.S.

University of Nebraska–Lincoln

Corresponding author – Richard L. Wiener, Department of Psychology, 238 Burnett Hall,
University of Nebraska at Lincoln, Lincoln NE 68588, U.S.A. Email: rwiener2@unl.edu

Abstract

Jury researchers have long been concerned about the generalizability of results from experiments that utilize undergraduate students as mock jurors. The current experiment examined the differences between 120 students (55 males and 65 females, mean age = 20 years) and 99 community members (49 males and 50 females, mean age = 42 years) in culpability evaluations for homicide and sexual assault cases. Explicit attitude measures served as indicators of bias for sexual assault, defendant, and homicide adjudication. Results revealed that student and community participants showed different biases on these general explicit attitude measures and these differences manifested in judgments of culpability (guilt likelihood, convincingness of state's arguments, convincingness of defendant's arguments, and the defendants' criminal intentions) in sexual assault and homicide case scenarios. The results also showed that student mock jurors were more lenient when assigning guilt in homicide cases than were community members. The implications for future mock jury research are discussed.

Published in *Behavioral Sciences and the Law* 29 (2011), pp 376–394.

DOI: 10.1002/bsl.971

Copyright © 2011 John Wiley & Sons, Ltd. Used by permission.

Published 28 March 2011.

This research was funded by the National Science Foundation, Grant no. SES-0096347

Introduction

Graduate students, and indeed undergraduate psychology majors, spend a great deal of time learning about the differences between drawing causal inferences from research data and generalizing those inferences outside the sample data that produced them (Cook & Campbell, 1979). For jury researchers, the issue of generalizability goes beyond the usual concerns of time, locations, and research participants to include creating an accurate representation of the processes in the laboratory that we observe in the courtroom. To the extent that researchers are successful at modeling the factors that define jury decision-making at trial in their laboratories, the results of the research will be informative to the legal system. To the extent that they are not successful, the results will be less relevant to the law but may still be important to scholars as they try to develop theories of choice that capture the decision-making in culpability evaluations. The factors that researchers typically try to represent in the laboratory include, but are not limited to, elements of the law, trial procedures, rules of evidence, and the impact of trial outcomes on the judgments and decisions of lay jurors.

It seems to us that in jury research, the issues of external validity (generalizing across people, locations, and times) and construct validity (generalizing from laboratory processes to real jury rooms) are heavily interactive. In other words, the effectiveness of legal translation and trial simulations is very likely dependent upon the knowledge, background, and understanding of the individuals who serve as research participants, or mock jurors, in our analog studies. When researchers ask jurors to make civil liability decisions or criminal culpability decisions, the decision changes not only with the jury instructions, quality of the jury simulation, elements of the law, admissibility of evidence, and the like, but also with the nature of the decision-maker. Therefore, the central question at issue is the interaction between the measures and manipulations, on the one hand, and the type of decision-maker reaching decisions on the other.

We are hardly the first jury researchers to consider the interaction of people with measures and manipulations. At the same time that Cook and Campbell (1979) were writing their seminal paper elaborating on their approach to understanding different types of validities,

Weiten and Diamond (1979) were applying these same ideas to jury research. Weiten and Diamond (1979) surveyed the then existing simulated jury research and identified six threats to external validity (i.e., the ability of a research design to make valid causal inferences) and policy relevance. These included (not in their original order) inadequate trial simulations, lack of jury deliberations, inappropriate dependent variables (focusing too often on criminal punishment and lacking actual elements of charged crimes), corroboration with field research (i.e., convergence of findings between laboratory and field studies), problems in role-playing (i.e., the difference in consequences of decisions for real and mock jurors are substantial), and inadequate sampling (relying too heavily on college student mock jurors). While the first five of these categories are problems for construct validity, the last is clearly an issue of what we more traditionally refer to as external validity (i.e., the ability of a research design to generalize across people, locations, and times; Cook & Campbell, 1979).

Furthermore, when discussing the issues inherent in inadequate sampling, these authors consistently entertained the likelihood that “interactions between experimental treatments and subject variables may pose threats to valid cross-sample inference” (p. 76). For example, Weiten and Diamond (1979) considered an early study in which Berg and Vidmar (1975) showed that conservatism correlated positively with harsh jury decision-making and argued that student mock jurors may be less conservative, less authoritarian and therefore more likely to be lenient in criminal trials. Another way of thinking about this problem is that juror attitudes interact with case facts, charges, and trial procedure in such a way as to threaten the construct validity of the studies that rely solely on undergraduate mock jurors. Put bluntly, the question is simply this: do undergraduates come with attitudes and beliefs that alter their perceptions of legal constructs, such as the meaning of the elements of crime or a tort, beyond a reasonable doubt versus preponderance of the evidence, criminal intent, negligence, burden of proof, and the like? If so, then the interaction of type of participant and type of procedure may indeed become a subtle but important threat to the generalizability of mock jury research. Weiten and Diamond (1979) go on to suggest that other factors, such as differences in cognitive capacities, might make student jurors behave differently than jury-eligible citizens. If students have greater memory capacities and information-processing abilities than do typical jurors,

then they may reach very different judgments about the same case facts, or even if they reach similar judgments, they may do so using very different approaches, styles, and information. Indeed, in an independent assessment of the simulated jury research, Bray and Kerr (1979) shared Weiten and Diamond's concerns and called for researchers to systematically explore how well simulated research studies generalize to real trial results. Some, but not all, argue that this work is still lacking in the literature.

Eighteen years later, Diamond (1997) revisited her arguments about the threats that mock jury research present to external validity and policy relevancy and concluded that some progress was evident in the literature. Her review showed more studies with realistic measures, legally relevant trial simulations, and even more use of jury-eligible mock jurors. Still she found room for improvement, grounded, again, in concerns about the interactions between participant attributes and legal processes. She opined that there were mixed results in the literature comparing student samples with eligible jurors, with some showing parallel findings (Casper, Benedict, & Perry, 1989) but others finding major differences between the samples (Severance, Greene, & Loftus, 1984). The problem in 1997 remained the interaction between external and construct validity. According to Diamond (1997), we had no theories to help us understand when and how differences between student and general adult populations interacted with the experimental manipulations that researchers used to try to capture the jury experience in a simulated mock jury study. Diamond (1997) called for efforts at building a database to compare the differences between student and adult samples and to develop theories that guided researcher discretion about sampling. Our work in this paper and elsewhere studying generic prejudice (Wiener, Arnot, Winter, & Redmond, 2006) goes to those twin goals of developing a theory that offers an adequate model of jury behavior and that points the way to tests of the main effects and interactions of participant attributes in the context of legally relevant judgments and decisions. Diamond (1997) concluded that, "... until we can predict reliably when a student sample is likely to provide an adequate model of juror behavior, jury venires should remain the preferred source for maximizing both face and external validity" (p. 563).

Shortly after Diamond revisited this issue, Bornstein (1999) also addressed the concerns of external and construct validity in mock jury

research. Bornstein took seriously the need for an empirical assessment of the relationship between student and jury-eligible citizens' performance in mock jury trials. He collected and displayed a table of 26 experiments, which included student and non-student samples, and tabulated how many studies found main effects comparing student and non-student samples and how many found interactions. While this analysis does not constitute a true meta-analysis with an exhaustive list of studies, measures of effect sizes and fail-safe statistics, Bornstein found only six out of the 26 reported main effects for type of participant and even fewer that found interactions with other study factors. He suggested that jury researchers could take solace from the findings that type of participant does not seem to matter a great deal in mock jury research. Indeed, many articles published after 1999 cite this paper to justify the use of convenient but less than realistic simulations and undergraduate participants in mock jury research.

While we welcome the Bornstein (1999) analysis as a first step in examining systematically what we know about student mock jurors and community samples, we must point out that it does not address several important factors. This initial work was unable to examine the specific types of interactions involving sample type and the factors that define specific construct validity issues in jury research (burdens of proof, standards of evidence, deliberations, types of charges, and the like). For example, at even the most elementary level, we wonder whether type of charge matters, and we suspect that it does. Very little, if any, data directly compare differences in how student and community jurors think about facts and evidence even for criminal cases with a common theme, such as violent crime.

Despite researchers' hopes that juror type has minimal consequence, there are, as of yet, no formal meta-analyses that we know of that tested differences in samples or interactions between samples and other study factors in mock jury research, with the exception of Mitchell, Haw, Pfeifer, and Meissner (2005), who examined race effects in mock jury research. The authors studied both jury verdicts and sentencing and found small effects for racial bias more among Black than among White research participants. For our purposes, the most interesting finding in the study was that the type of participant moderated the effect of the race of the defendant, so that community members were more likely to show racial bias in sentencing (but not verdict decisions). Thus for at least some purposes there is some formal

evidence that type of juror does modify judgments about defendants. Where else will we find evidence of interactions between external and construct validity factors?

Much more needs to be done. One of the most important findings in the Mitchell et al. (2005) study was the difficulty that the researchers had in studying the moderating effects of type of sample. For the verdict decision, the authors located 37 experiments that studied the effects of race in student samples, but only seven that did so with community samples. For the sentencing decision there were 15 student sample studies but only five community sample studies. In short, the Mitchell et al. (2005) meta-analysis suggests that the interaction between defendant race and type of sample is worth further investigation, but that until more community studies appear in the literature, formal comparisons using meta-analytic techniques to study this and other interactions between external and construct validity threats will remain difficult to complete.

In this paper, we try to further this type of work by asking whether student and community samples think differently about sexual assault charges and homicide charges. In our previous work, we conducted a number of community sample studies looking at legal decision-making involving sexual harassment (Maeder & Wiener, 2007; Wiener & Winter, 2007; Wiener & Hurt, 1999, 2000; Wiener, Hurt, Russell, Mannen, & Gasper, 1997; Wiener Hackney, Kadela, Rauch, Seib, Warren, & Hurt, 2002; Wiener, Voss, Winter, & Arnot, 2005), jury decisions in re-enacted death penalty cases (Wiener, Pritchard, & Weston, 1995; Wiener, Rogers, Winter, Hurt, Hackney, Kadela, Seib, Rauch, Warren, & Morasco, 2004; Wiener, Richter, Humke, Jones, Maeder, & Cantone, unpublished), and with this current paper, generic prejudice in sexual assault and homicide cases. In the current research, we made use of prior findings on generic prejudice with community jury samples (Vidmar, 1997) and experimental findings relying on student samples (Wiener et al., 2006). Vidmar (1997) defines generic prejudice as “general attitudes, beliefs, and biases held by the juror that prevents her or him from deciding the case with a fair and impartial mind” (p. 6). Generic prejudice arises from individual differences that jurors hold about the “nature of the crime or the type of parties involved [which] cause the juror to classify the case as having certain characteristics, thereby invoking stereotyped prejudices about any defendant accused of the crime” (Vidmar, 1997, p. 6; see also Vidmar, 2002). In

a field study of Canadian courts, Vidmar (1997) reported an average of 36% of venire members biased during *voir dire* in 25 trials (849 potential jurors) in which male defendants faced charges of sexually assaulting male and female complainants, who ranged from 5 to 23 years old (Vidmar, 1997). In these cases, had the courts seated these jurors it is likely they would have applied a heuristic, which holds that male defendants charged with sexual crimes are likely guilty. Vidmar (2002) argued that pretrial publicity interjected into litigation through the written and electronic media's coverage of criminal proceedings is one, but not the only, trigger of generic prejudice. He favored peremptory challenges to remove jurors who show signs of generic prejudice and who would apply prejudicial heuristics.

In a follow-up study in our laboratory (Wiener et al., 2006), we presented 10 homicide and 10 sexual assault scenarios to undergraduate mock jurors, for whom some of the scenarios were matched by charge and others were matched by fact similarity. We used regression analysis to study consistencies in verdicts as a function of specific charge, type of case (sexual assault vs. homicide), and fact similarities among cases and found that undergraduates judged the same charge types with different facts more consistently than they should by chance alone (generic prejudice). This was truer for the sexual assault cases than for the homicide cases. One interpretation of these findings is that undergraduates hold attitudes about sexual assault that influence their judgments of culpability beyond the facts of the case, as they understand them. In other words, the results of Wiener et al. (2006) were consistent with an untested interaction between type of sample and type of charge, which would suggest that specific attitudes may very well influence the judgments that different samples of mock jurors make about sexual assault cases even when the cases show little in the way of fact similarities. In the current research, we tested this interaction using a similar set of sexual assault and homicide cases to look more systematically for interactions between explicit attitudes as measured with several scales that other researchers have used successfully in the recent past. These were the Attitudes toward Rape (ATR) scale (Kovera, 2002), the Attitudes toward Sexual Abuse (ATSA) scale (Briere, Hehschel, & Smiljanich, 1992), and the Juror Bias Scale (JBS) (Kassin and Wrightsman, 1983).

In this manner, we applied our prior generic prejudice work to the current concern with interactions between construct and external

validity factors. We predicted that sample type (student vs. community mock jurors) would moderate the associations between these explicitly measured attitudes and mock juror evaluations of sexual assault culpability, and that the type of charges that the mock jurors evaluated would further qualify those associations. Because Wiener et al. (2006) found more evidence of generic bias in sexual assault cases than in homicide cases, we expected fewer, if any, interactions between explicit attitude measures toward homicide and type of sample in the 10 homicide cases from Wiener et al. (2006). To test that prediction, we administered an Attitude Toward Homicide Scale (ATHA), which we created for use in this experiment, along with the existing ATR, ATSA, and JBS, and calculated associations between those explicit measures and mock juror judgments of culpability for the sexual assault and homicide scenarios.

The current paper adds to the previous work the expectation that sexual assault charges and homicide charges may invoke different types of attitudes in students and community members as a function of both groups' sophistication with the social world. Put simply, we suspect that adult community members, many of whom have children of their own, likely hold different attitudes about sexual assault and apply them somewhat differently than do college students, who probably look at sexuality with much more ambiguity than do their grandparents, parents, uncles, and aunts. Because sexual assault involves behavior that is normally valued (i.e., consensual relations among adults) but sometimes not (i.e., forced or unwanted relations), charges in this area of criminal law require complicated value judgments and, as such, are more likely to result in generic prejudice. We expected that college students' decisions would show the ambiguity in their judgments about sexual behavior and therefore be more influenced by their attitudes about sexual conduct (and misconduct) than would those in a community sample. We did not expect these same types of effects for homicide. Because there is less room for interpretation in cases of homicide, we expected simple leniency effects in homicide cases as previously found in the jury decision-making literature, where college students are, overall, more lenient than community samples (Berman & Cutler, 1996; Schuller & Hastings, 1996; Simon & Mahan, 1971).

In summary, the problem with simple comparisons between students and nonstudents that do not take each sample's attitudes and

prior beliefs into consideration is that these studies oversimplify a complex social and psychological reality. If different types of jurors respond differently in an understandable manner to specific aspects of law and legal process with regard to violent crime, then it is incumbent upon researchers to look for other systematic differences in other areas. Other interactions between external validity issues (i.e., type of sample) and construct validity issues (i.e., effects of admonishments, burdens of proof, standards of proof, establishment of negligence, exculpatory clauses, and so on) may also be lurking beneath the surface of our research literature.

Method

Participants

This study included one sample of 120 undergraduates (55 males and 65 females, mean age = 20 years) at the University of Nebraska-Lincoln who participated in return for course credit. The sample comprised 81% Caucasian, 5% African-American, 5% Asian-American, and 9% other students. A second sample of 99 community members (49 males and 50 females, mean age = 42 years) responded to advertisements in local newspapers and completed the experiment for a \$30 stipend. Again, the majority of participants were Caucasian (85%). African-American participants comprised 6% of the sample and the remainder categorized themselves as “other” (9%). Prior to participation in the experiment, researchers screened community members to ensure juror eligibility. Each individual who answered a newspaper ad by leaving their name and number on our laboratory phone was required either to be a registered voter in the state of Nebraska or to hold a valid Nebraska driver’s license, be free from felony convictions that stripped civil rights, and be at least 19 years of age, the age of majority in Nebraska.¹ In this manner, we selected a jury-eligible community sample to compare against a sample of undergraduate students who were part of the typical university subject pool.

1. One community member feigned answers to these questions, as discovered in the demographics questionnaire, and was excluded from the analysis, resulting in a sample of 99 community members.

There was no deception in this experiment. The researchers simply explained to the participants that they would be reading and evaluating case scenarios and completing questionnaires about the culpability of the actors in a series of allegations.

Materials and Procedure

Similar to Wiener et al.'s (2006) method, participants evaluated 20 crime fact patterns or case scenarios that we loosely based on 10 homicide and 10 sex crime trials appearing as appellate cases in New York State case reporters. Participants evaluated two exemplars for each of five sexual assault charges (i.e., attempted sexual assault in the first degree, attempted sexual assault in the third degree, sexual assault against a minor, sexual assault in first degree, and sexual assault in the second degree) and two for each of five homicide charges (i.e., first-degree murder, second-degree murder, involuntary manslaughter, second-degree attempted murder, and motor vehicle homicide).² Each written scenario summarized the case facts, altering the arguments to create fact patterns that were neither weak nor strong, as established in prior pilot testing.

Each scenario began by naming the defendant and victim and then went on to present the facts of the case. The details consisted of testimony of the defendant, victim, and any available testimony from expert and lay witnesses. The scenarios concluded with a short definition of the charge, which gave the elements of the law, burden of proof, and standard of proof. The elements of the law for each scenario came from statutory language in the Nebraska Penal Code. All scenarios were approximately one page in length.

After reading each scenario, participants evaluated the culpability of the defendant against four indices, three of which presented evaluation ratings on nine-point Likert type scales: (i) the likelihood of the defendant's guilt (1, very unlikely, to 9, very likely); (ii) convincingness of the state's evidence (1, not at all convincing, to 9, very convincing); and (iii) convincingness of the defendant's evidence (same convincingness scale). Also included was a "*mens rea*" question, which followed

2. Copies of all scenarios may be obtained from the second author, Dr Richard Wiener, by email at rwieners2@unl.edu

Table 1. Statistics for guilt judgments

<i>Charge</i>	<i>n</i>	<i>M(SD)</i>	<i>95% CI</i>
Attempted sexual assault in the first degree	218	5.85(1.54)	[5.64, 6.05]
Attempted sexual assault in the third degree	218	5.25(1.68)	[5.03, 5.48]
Sexual assault of a Minor	217	5.32(1.55)	[5.11, 5.52]
Sexual assault in the first degree	218	6.44(1.35)	[6.26, 6.62]
Sexual assault in the second degree	217	5.51(1.51)	[5.31, 5.71]
Murder in the second degree	219	6.22(1.35)	[6.04, 6.40]
Murder in the first degree	219	6.74(1.36)	[6.56, 6.92]
Manslaughter	219	5.84(1.65)	[5.62, 6.06]
Attempted murder in the second degree	219	5.52(1.59)	[5.31, 5.73]
Motor vehicle homicide	219	5.48(1.70)	[5.25, 5.71]

CI, confidence interval.

the Model Penal Code (1962; 1981) and asked respondents whether the defendant acted with (1) no intention, (2) negligently, (3) recklessly, (4) knowingly or (5) purposefully. Participants read and evaluated each scenario in the order in which they appeared. We tested the calibration of the scenarios by calculating the means and standard deviations for each pair of scenarios with the same charges on the likelihood of guilt scale, the midpoint of which was 5.00. **Table 1** lists these statistics along with the 95% confidence intervals (CIs) for the means. Examination of these results reveals that guilt judgments were close to the midpoint of the means (ranging between 5.25 and 6.74), and the CIs (never including a value below 5.00 and never reaching 7.00). Table 1 shows that the furthest deviations for the midpoints of the scales were for murder in the first degree and sexual assault in the first degree, which tend to produce slightly higher ratings (more likelihood of guilt) than the other charges. Thus, the judgments on the scenarios, although slightly biased against the defendants, showed enough variability to allow for meaningful analyses of culpability.

Following the 20 case scenarios and their evaluations, participants completed several explicit measures of adjudicative bias that measure trait-like qualities. The researchers presented these stable attitude measures after the case scenarios so that they would not inform the participants that the true purpose of the experiment was to examine differential relationships between attitudes toward different types of crimes and culpability evaluations. Three of the four measures were published scales with established and stable reliabilities:

the ATR scale ($\alpha = 0.87$), the ATSA scale ($\alpha = 0.72$) and the Juror Bias Scale ($\alpha = 0.83$). We also included the ATHA that we created for the purposes of this experiment.

The ATR is a 27-item questionnaire that measures responsibility and blame for sexual assault crimes. Examples of questions include, "In most cases when a woman was raped, she was asking for it" and "In order to protect the male, it should be difficult to prove that a rape has occurred" (1, strongly agree, to 7, strongly disagree). The ATSA measures culpability in sexual assault cases with child victims across 15 items such as, "Children sometimes say they have been molested to get attention, or to get back at someone" and "Children sometimes try to seduce adults" (1, strongly agree, to 5, strongly disagree). The JBS is a 22-item scale that measures jurors' pretrial attitudes by tapping into their opinions about the legal structure and about those who are key players in the system, such as judges, the police, and defendants. While it includes two, and sometimes three, scales, this study used a global measure of bias against the defense with questions such as, "Too often jurors hesitate to convict someone who is guilty out of pure sympathy" and "Out of every 100 people brought to trial, at least 75 are guilty of the crime with which they are charged" (1, strongly agree, to 5, strongly disagree).

The ATHA contains seven items that we developed to measure attitudes about homicide charges on which participants rated their agreement on scales ranging from 1 (agree) to 5 (disagree). The items were as follows: (i) defendants charged with homicide are very likely to be guilty; (ii) when police arrest someone for killing another, they only do so if they are sure the killer is guilty; (iii) too often people serve long sentences for homicide when they are not guilty; (iv) people who kill others while driving drunk should not be judged as harshly as those who kill for their own purposes; (v) manslaughter, murder, and attempted murder are all equally horrible crimes; (vi) people who try to kill someone but fail at the last minute should be punished as severely as those who succeed in their efforts; (vii) the courts make too much out of subtle differences in homicide cases, such as the difference between first- and second-degree murder. Coefficient alpha for the scale was 0.62. These four explicit measures appeared in random order separately determined for each participant.

To rule out the possibility that these scales were all measures of the same or similar constructs, we calculated correlations between

Table 2. Correlations between predictors

	<i>JBS</i>	<i>ATHA</i>	<i>ATSA</i>	<i>ATR</i>
JBS	—	—	—	—
ATHA	0.56***	—	—	—
ATSA	0.10	-0.07	—	—
ATR	0.28***	0.18*	0.42***	—

ATHA, Attitude Toward Homicide Adjudication Scale; ATR, Attitudes toward Rape scale; ATSA, Attitudes toward Sexual Abuse scale; JBS, Juror Bias Scale.

* $p < 0.05$; *** $p < 0.001$

the scales and displayed them in **Table 2**. The highest correlation among these predictors was 0.56, between the JBS and ATHA, indicating that respondents who generally favored the prosecution also showed some bias against defendants in homicide cases. Furthermore, the 0.42 correlation between the ATSA and ATR showed that respondents who thought children could be responsible for their own sexual assaults were also biased against complainants in rape cases. Despite these significant relationships (and two other lower ones that included the ATR), the squared correlation coefficients in these data show that, on average, the predictors shared only about 10% of the variance and that the highest r^2 was 0.31. Thus, while there were some overall relationships between these variables, they were not measures of the same constructs.

Finally, participants completed a standard demographic questionnaire. Students and community members worked independently within a group setting of no more than 10 respondents to evaluate the cases, complete the attitude measures, and fill out the demographic survey. At no time did the researchers permit the participants to communicate about the materials.

Results

To simplify the analysis and focus on the examination of differences between students and community members, we combined cases that shared a charge, so that we had 10 pairs of cases, five pairs with homicide-based charges and five with sexual assault-based charges instead of 20 individual cases. For example, the analyses used the averages of the guilt ratings for both attempted sexual assault in the

second-degree cases for each participant, to create a single attempted sexual assault in the second-degree score. We did the same for state convincingness, defense convincingness, and criminal intent.

Analyses of covariance measured main effects and interactions for the type of participant (student or community member) separately for the sexual assault and homicide cases. The dependent variables in those cases were the guilt, state convincingness, defendant convincingness, and criminal intent ratings, and the between-subject independent variables were type of participant and participant gender. The five case-specific charges within sexual assault cases and the five specific charges within homicide cases were repeated measures factors. The continuous explicit attitude measures were covariates in the analyses reported below. In summary, we calculated 2 (participant type) \times 2 (gender) \times 5 (charge) ANCOVAs with appropriate explicit attitude scales as covariates and evaluated those models separately for the sexual assault cases and the homicide cases.

Sexual Assault

Guilt likelihood

Each ANCOVA model examined the relationship between the five sexual assault charges by participant type and gender with the appropriate explicit measures, ATSA, ATR, and JBS. We report only main effects or interactions involving the type of participant for each of the four indices of culpability. The analyses resulted in two interactions for the sexual assault cases, demonstrating that the relationships between the attitudes toward defendants overall (JBS), and attitudes toward rape (ATR) were different depending upon the type of participant and the type of charge. For guilt in sexual assault cases, there were interactions for type of participant by charge by JBS [$F(4, 800) = 2.54, p = 0.039, \eta^2 = 0.013$], and type of participant by charge by gender by ATR [$F(4, 800) = 2.90, p = 0.021, \eta^2 = 0.014$]. Follow-up analysis for the three-way JBS interaction showed that the JBS was a stronger predictor of guilt verdicts in sexual assault cases for community members [$F(4, 364) = 3.73, p = 0.005, \eta^2 = 0.039$] than for students [$F(4, 436) = 1.57, p = 0.18, \eta^2 = 0.014$]. **Table 3** lists the beta weights and significance levels for each type of charge.³ The JBS predicted guilty

3. Note that the beta values reported in this manuscript are all standardized.

Table 3. Regression weights, sexual assault guilt and JBS by type of participant

Charge	Beta weight	
	Students	Community members
Attempted sexual assault in the first degree	0.076	-0.20 ^a
Attempted sexual assault in the third degree	-0.047	-0.23 [*]
Sexual assault of a minor	-0.13	-0.45 ^{***}
Sexual assault in the first degree	-0.12	-0.056
Sexual assault in the second degree	-0.13	-0.079

a. $p < 0.065$; * $p < 0.05$; *** $p < 0.001$

verdicts for attempted sexual assault in the first degree, attempted sexual assault in the third degree, and sexual assault of a minor for community members. Thus, community members who showed less bias against defendants were the ones who were least likely to find the defendants guilty of these three sexual assault charges. We found no evidence of a relationship between bias and guilt ratings for the sexual assault ratings in the student sample.

Follow-up analysis for the four-way interaction with the ATR revealed that, overall, the attitudes toward rape behaved just the opposite to overall jury bias because it was a better predictor of guilt likelihood in sexual assault cases for students [$F(4, 436) = 2.59, p = 0.036, \eta^2 = 0.023$] than for community members [$F(4, 364) = 1.49, p = 0.21, \eta^2 = 0.016$]. Making the results more complicated, there were differences among the types of cases in which the ATR was successful for predicting judgments for men and women. **Table 4** regression weights showed significant effects for four of the five charge types. For male students, the ATR was related to guilt verdicts for attempted sexual assault in the first and third degrees and sexual assault in the second degree, and predicted sexual assault of a minor and sexual assault in the second degree for female students. Furthermore, the ATR was related to sexual assault guilt likelihood scores among female community members when they evaluated guilt in attempted sexual assault in the third degree and sexual assault of a minor. Thus, for male students and women, regardless of community or student status, those who scored higher on negative attitudes toward rape were more likely to find the defendants guilty in one or more of the sexual assault cases. Importantly, the relationship between the ATR and guilt verdicts failed to reach significance for any of the five sexual assault charge types among male community members.

Table 4. Regression weights, sexual assault guilt and ATR by gender and type of participant

Charge	Beta weight			
	Male students	Female students	Male community members	Female community members
Attempted sexual assault in the first degree	0.40*	0.11	0.088	0.15
Attempted sexual assault in the third degree	0.38*	-0.052	-0.082	0.38*
Sexual assault of a minor	0.053	0.33*	0.099	0.35*
Sexual assault in the first degree	0.24	0.23	0.15	0.022
Sexual assault in the second degree	0.31 ^a	0.41**	0.037	0.21

ATR, Attitudes toward Rape scale.

a. $p < 0.065$; * $p < 0.05$; ** $p < 0.01$

Convincingness of state's evidence

Assessments of state convincingness in sexual assault scenarios produced a significant interaction for type of participant by JBS [$F(1, 199) = 3.95, p = 0.048, \eta^2 = 0.019$]. Follow-up analyses showed that the JBS was associated with the convincingness of the state's arguments in community members [$F(1, 91) = 10.63, p = 0.002, \eta^2 = 0.11$], but not in students [$F(1, 108) = 2.22, p = 0.14, \eta^2 = 0.02$]. Simple correlations showed that those community members with less bias against defendants were less likely to find the state's evidence convincing ($r = -0.21, p = 0.042$). The same relationship for students was not significant ($r = -0.056, p = 0.54$).^{4,5}

4 There was a significant interaction among charge \times type of participant \times gender \times ATSA [$F(4, 800) = 2.49, p = 0.042, \eta^2 = 0.012$]. However, follow-up analyses showed that only a single correlation was driving this interaction. The ATSA was predictive of female student state convincingness evaluations for the attempted sexual assault in the first degree cases only. We discuss this isolated effect no further.

5 The indicator of convincingness of the defendant's evidence in sexual assault cases yielded a significant interaction among type of participant, gender, and ATSA [$F(1, 200) = 4.35, p = 0.038, \eta^2 = 0.021$]. However, follow-up regressions showed that the ATSA was not significantly related to defendant convincingness in any of the four groups (i.e., male or female students or community members). There were no significant main effects or interactions involving type of participant for criminal intent in the sexual assault cases.

Homicide

We constructed ANCOVA models to examine the relationship between the five homicide charges, again including type of participant and gender as between-subject factors, and the non-sex-based explicit attitude measures, the ATHA and the JBS, as covariates. The ATR and ATSA are largely irrelevant to homicide verdicts so we did not include them in the statistical models. Again, we report only main effects or interactions involving the type of participant for each of the four indices of guilt.

Guilt Likelihood

For guilt likelihood, we found a between-groups main effect for type of participant [$F(1, 205) = 4.40, p = 0.037, \eta^2 = 0.021$], which did not involve charge as a moderator so that the effect occurred across all five pairs of homicide cases. Students had a mean verdict of 5.88 (SD = 0.82), while community members rated homicide guilt certainty higher, at 6.06 (SD = 0.90). Consistent with the previous literature, these results show that students were more lenient when assigning guilt in homicide cases than were community members. The ANCOVAs also showed a significant interaction for type of participant \times gender \times charge \times ATHA [$F(4, 820) = 3.79, p = 0.005, \eta^2 = 0.018$]. Follow-up regressions showed that the ATHA was significantly associated with likelihood of guilt for only female students in one crime category, namely, murder in the second degree ($\beta = -0.35, p < 0.01$). For female students, but not other participants, decreased bias against the defendants was associated with a lower likelihood of finding the defendant guilty in the second-degree murder cases.

Convincingness of state's evidence

The convincingness of the state's evidence against the defendant provided one between groups interaction between type of participant and the ATHA [$F(1, 205) = 3.75, p = 0.054, \eta^2 = 0.018$]. Follow-up analysis revealed that the ATHA was a better predictor of state convincingness in homicide cases for students [$F(1, 112) = 6.10, p = 0.015, \eta^2 = 0.052$] than for community members [$F(1, 93) = 0.075, p = 0.79, \eta^2 =$

Table 5. Regression weights, homicide defendant convincingness and JBS by gender and type of participant

Explicit measure	Beta weight			
	Male students	Female students	Male community members	Female community members
JBS	-0.20	0.30*	-0.013	-0.16

JBS, Juror Bias Scale.

* $p < 0.05$

0.001]. The beta weights and significance levels in follow-up regressions for this interaction explain this finding. Student assessments of state convincingness were significantly related to the ATHA ($\beta = -0.26, p = 0.014$), while community member assessments were not ($\beta = 0.024, p = 0.85$). That is, for students, but not for community participants, decreased bias against the defendants was associated with a lower likelihood of finding that the state had a more convincing case.

Convincingness of defendant's evidence

There was also one between-groups interaction of type of participant on defendant convincingness for homicide cases, a type \times gender \times JBS [$F(1, 205) = 4.62, p = 0.033, \eta^2 = 0.022$]. Again, follow-up regression analyses revealed that the JBS was a better predictor of student judgments of defendant convincingness [$F(1, 112) = 6.15, p = 0.015, \eta^2 = 0.052$] than community member assessments [$F(1, 93) = 0.42, p = 0.52, \eta^2 = 0.005$]. Analysis of **Table 5**, which provides the beta weights and significance levels for each type and gender of participant, shows that the JBS was predictive of defendant convincingness judgments for only female students. Surprisingly, female students who were higher in bias against defendants were more likely to find that the defense has a convincing case. Perhaps they were searching for evidence to offset their biases against the defense in order to be able to find the defendants less likely to be guilty in the homicide cases.

Criminal intent

Intention of the defendant determinations produced two significant interactions, type of participant \times charge [$F(4, 820) = 5.26, p = 0.001, \eta^2 = 0.025$] and type of participant \times charge \times JBS [$F(4, 820) = 4.18,$

Table 6. Difference tests, homicide intention by type of participant

Charge pair	<i>t</i> (df)	
	Students	Community members
Murder in the second degree and murder in the first degree	6.96 (119)***	5.16 (98)***
Means ^a	4.11, 3.40	4.06, 3.52
Murder in the second degree and manslaughter	9.78 (119)***	7.42 (98)***
Means	4.11, 3.13	4.06, 3.21
Murder in the second degree and attempted murder in the second degree	5.30 (119)***	5.16 (98)***
Means	4.11, 3.52	4.06, 3.44
Murder in the second degree and motor vehicle homicide	22.10 (119)***	18.43 (98)***
Means	4.11, 1.93	4.06, 1.93
Murder in the first degree and manslaughter	2.51 (119) *	2.50 (98)*
Means	3.40, 3.13	3.52, 3.21
Murder in the first degree and attempted murder in the second degree	-1.00 (119)	0.59 (98)
Means	3.40, 3.52	3.52, 3.44
Murder in the first degree and motor vehicle homicide	12.59 (119)***	12.96 (98)***
Means	3.40, 1.93	3.52, 1.93
Manslaughter and attempted murder in the second degree	-3.51 (119)**	-1.73 (98)
Means	3.13, 3.52	3.21, 3.44
Manslaughter and motor vehicle homicide	13.46 (119)***	11.71 (98)***
Means	3.13, 1.93	3.21, 1.93
Attempted murder in the second degree and motor vehicle homicide	12.64 (119)***	11.06 (98)***
Means	3.52, 1.93	3.44, 1.93

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

a. Means in each row for each column are intention means for the student or community samples for the different cases in the first column.

$p = 0.002$, $\eta^2 = 0.020$]. Follow-up ANOVAs for the type of participant \times charge interaction showed that there were differences in the intention verdicts for the five charges of homicide cases for both students [$F(4, 448) = 9.14$, $p = 0.001$, $\eta^2 = 0.075$] and community members [$F(4, 372) = 2.40$, $p = 0.050$, $\eta^2 = 0.025$]. **Table 6** presents the 10 pairs of different charges that we listed by comparing each of the five homicide charges with each of the other homicide charges. It displays t -tests and significance levels for each pair. The results show that, for students, the only pair of cases that were not significantly different on intention assessments comprised murder in the first degree and attempted murder in the second degree. This pair of charges was also not significantly different for community members. However, for students there was an additional pair, manslaughter and attempted murder in the second degree, which was not significantly different in

intention ratings. The t -values show that for some of the pairs, the differences between the means of the students' ratings are more separated than were the community sample ratings.

Inspection of the means of the intention ratings for each charge revealed that students and community members ranked the cases on intention in a similar, but not the same, manner.

For all pairs, except for first-degree murder and attempted second-degree murder, both the students and the community members perceived criminal intention differences in the charges, but not always finding greater levels of criminal intent in the more serious charges (see, e.g., murder in the second degree, murder in the first degree, and manslaughter). Cases sharing a charge of murder in the second degree had the highest intention ranking, with means that corresponded to a *mens rea* rating of slightly above knowing. For students, manslaughter had the next highest ranking, followed closely by murder in the first degree, with a ranking between reckless and knowing. Community members rated murder in the first degree as slightly higher than attempted murder in the second degree. Both samples rated manslaughter as slightly above reckless. Finally, defendants facing a charge of motor vehicle homicide were assigned the lowest intention, slightly below negligence. Regardless of the direction of the differences in rated criminal intent, in all comparisons but one (manslaughter and attempted murder in the second degree), the students' level of perceived intent was similar to that of the community sample. Therefore, whatever drove the leniency effect in students on guilt likelihood ratings was very likely not a difference in perceptions of criminal intent.

The follow-up analysis for the type of participant \times charge \times JBS interaction demonstrated that the JBS was a better predictor of intention verdicts in homicide cases for students [$F(4, 448) = 5.21, p = 0.001, \eta^2 = 0.044$] than for community members [$F(4, 372) = 0.76, p = 0.55, \eta^2 = 0.008$]. The results of follow-up regression analyses showed that for one type of charge, attempted murder in the second degree, the JBS was associated with lower levels of inferred criminal intent. Thus students, but not community members, who showed less bias against defendants were likely to find lower levels of criminal intent ($\beta = -0.21, p < 0.05$).

See **Table 7** for a complete summary of the significant main effects and interactions involving type of sample in the sexual assault and homicide cases.

Table 7. Summary of findings

<i>Community sample</i>	<i>Students</i>
Sexual assault cases	
<i>Guilt certainty</i>	
The JBS was a significant predictor of certainty in sexual assault guilty verdicts	No effect for JBS
The ATR was a significant predictor of guilt certainty for some sexual assault cases for female community members but not for males	The ATR was a significant predictor of guilt certainty for some sexual assault cases for male students and other cases for female students
<i>State convincingness</i>	
The JBS was a significant predictor of how convincing the state's case was in sexual assault cases for community members	No effect for this group
Homicide cases	
<i>Guilt certainty</i>	
Community members more likely to be certain of guilty verdicts	The student sample was more likely to be certain of not guilty verdicts
No effect for ATHA	ATHA was a significant predictor of guilt certainty for female students
<i>State convincingness</i>	
No effect for this group	The ATHA was a significant predictor of how convincing the state's case was in homicide cases for students
<i>Defense convincingness</i>	
No effect for this group	The JBS was a significant predictor of how convincing the defense's case was for female student ratings of defense convincingness in homicide cases
<i>Intention</i>	
No effect for this group	The JBS was a significant predictor of level of intention for attempted murder in the second degree for students

ATHA, Attitude Toward Homicide Adjudication Scale; ATR, Attitudes toward Rape scale; ATSA, Attitudes toward Sexual Abuse scale; JBS, Juror Bias Scale.

Discussion

Sexual assault

Based upon our prior work in generic prejudice (Wiener et al., 2006), we predicted that the most meaningful interactions between type of participant attitudes and guilt ratings would emerge in sexual assault cases. Indeed, our results showed that the relationships between the attitudes toward defendants overall (JBS) and toward rape (ATR)

showed very different associations with sexual assault guilt judgments depending upon the type of participant and the type of charge. Specifically, community members who showed less bias against defendants were the ones who were least likely to find the defendants guilty of these sexual assault charges. Community members' explicit attitudes played a significant role in their verdicts. Importantly, we found no evidence of a relationship between bias and guilt ratings for the sexual assault judgments in our student sample. The results for the relationship between the JBS and the convincingness of the state's evidence were similar in that those community members with less bias against defendants were less likely to find the state's evidence convincing.

However, a different pattern emerged for the ATR measure, which was related to guilt for women mock jurors, regardless of community or student status. Those women who scored higher on negative attitudes toward rape were more likely to find the defendants guilty in one or more of the sexual assault cases. Importantly, the relationship between the ATR and guilt verdicts failed to reach significance for any of the five sexual assault charge types among male community members. In short, the overall adjudication bias in community members only was related to guilt ratings in the sexual assault cases, while attitudes toward rape were associated with guilt ratings in all samples except male community members. At the same time, the interaction between type of participant and the JBS provided additional evidence of the complexity of attitudes in different populations. Specifically, community members with less bias against defendants were less likely to find the state's evidence convincing, but this was not true for the student sample.

While there is no identifiable theory that explains the differential responses that students and community members made to these sexual assault cases, it would be very useful to conduct research that could explain these interactions. We cannot determine from these data whether the differences in processing style are the result of variation in experience, motivation, emotion, or other psychological mechanisms. However, one reasonable explanation might be that students are accustomed to systematically processing information looking for arguments and counter-arguments that support specific verdicts, while community members who are not currently engaged in academic pursuits seek heuristic evidence that is consistent with their preexisting attitudes and biases. Future research to pinpoint the

psychological mechanisms that underlie these effects has the potential to do more than warn against overgeneralization. The conclusions reached in such work would help researchers to determine when it is best to rely on community residents as participants in mock jury research and when it is sufficient to present materials to undergraduate research participants. Even though we are not in a position to explain with theory the differences that we found, we can conclude that if the goal of the mock jury research is to model accurately the type of decisions that jurors reach and the processes by which they reach them, mock jury studies of sexual assault ought to take seriously the sample composition.

Homicide

Based upon our previous work in generic prejudice (Wiener et al., 2006), we predicted that the interactions emerging between participant attitudes and guilt ratings would paint a less clear picture in the homicide cases, and, indeed, the major story here seems to be the main effect of type of participant. Across all five charges within the homicide block of cases, students rated the likelihood of guilt lower than did community members. Helping to explain these results was an interaction between our ATHA scale and type of participant, such that the students but not community members were sensitive to their own attitudes toward this type of adjudication when rating the convincingness of the state's evidence. For students, but not for community participants, decreased bias against the defendants was associated with a lower likelihood of finding that the state had a more convincing case. Community members were less discriminating when evaluating the state's case, which may explain why they were more willing to rate the defendants guilty. Surprisingly, differences in the guilt ratings were likely not attributable to the differences in perceptions of criminal intent. While there were some slight differences in the way students and community members analyzed intent, overall they did so similarly, even showing comparable errors in thinking about intent in the homicide charges. Overall, we conclude that mock jury studies of homicide ought to consider the sample composition if researchers are to describe accurately the role of overall adjudication bias and make accurate predictions of when jurors will find defendants guilty.

Interactions of External Validity and Construct Validity

Our comparison of the manner in which undergraduate participants and community participants think about sexual assault and homicide cases shows that for some cases (here sexual assault cases) external validity factors, such as type of participant and attitude factors, interact with construct validity factors, such as the specific charges. Using some well established explicit measures, the ATR scale, the ATSA scale, and the JBS, and a new measure created for this study, the ATHA, we showed clear differences in the associations between attitudes and case outcomes for students and community members. For some cases with sexual assault charges, the associations were the same for both samples, but for other cases, the way in which the mock jurors invoked their own attitudes was very different according to gender and community status. Furthermore, while we found some limited interactions between the JBS and the ATHA with type of participant for the homicide cases, the major finding was that there was a main effect across all charges, such that the students were more lenient in their judgments than were community-eligible jurors. Furthermore, that difference was not explainable by judgments of criminal intent but instead by overall leniency that influenced all of the homicide charges in the same way. Here the chief limiting factor is simply the main effect of one external validity factor, type of respondent.

Cook and Campbell (1979) provided us with a convenient and powerful way to think about the ability of our research results to generalize beyond the immediate data that we collect. Our work suggests that it is time for jury researchers to take seriously the issues of external validity and especially the way it may limit construct validity in jury studies. In this work, had we relied only on a sample of undergraduates, we would have found very different attitude relationships and guilt likelihood effects in both kinds of cases. Perhaps more importantly, we would have missed entirely the fine distinctions in the way in which mock jurors evaluate these types of cases and we would have been unable to retrieve those subtleties.

The use of undergraduate mock jurors is efficient and in many cases very suitable for the initial tests of important issues, even issues with some practical significance. For example, Rose and Ogloff (2001) showed that researchers could make extensive use of undergraduate participants to test the comprehensibility of jury instructions, pointing

out areas where they need improvement and areas in which people understand the plain meaning of the instructions. However, given the kinds of findings that we show here (leniency effects with undergraduates and more impact of jury bias), it might still be necessary to verify the final results of such testing with jury-eligible respondents. Other work that adds to the veracity of the student findings includes Wiener et al. (2004), who documented high levels of miscomprehension in a community sample of mock jurors. The poor comprehension in a community sample may have been better in a sample of undergraduate college students who were more familiar with mastering new and complex information and taking tests to measure comprehension. Our main point is that when there is some expectation that behavior will not be identical from sample to sample, jury researchers can only be sure of the generalizability of their findings when they compare results from different samples. Based on our work (Wiener et al., 2006) and Vidmar's (1997, 2002) work in generic prejudice we expected that students and community members would show different patterns of judgments across a sample of sexual assault and homicide cases and, in fact, they did. We suspect there are many other areas in which sample characteristics are important.

Vidmar (2008) commented that the most important structures to study in jury research are the effects of the judge's preliminary comments, the attorneys' opening statements, the direct and cross-examinations of plaintiffs' and defendants' attorneys, the actual testimony of the witnesses, the testimony of experts, attorneys' closing arguments, and jury instructions. Each of these important structural issues could influence the substantive judgments that are the products of both civil and criminal trials. Vidmar (2008) complains, justifiably, that jury simulation studies often fail to examine the relationship between the structural elements of trials and the judgment outcomes. We argue that relying only on the results of studies that use undergraduate students as mock jurors compounds this problem because students and eligible jurors may view charges, trial process, and trial procedure from some very different perspectives. At the very least, jury researchers should apply different theoretical models to study multiple sample testing for convergence of findings across populations before they place too much confidence in findings with undergraduate mock jurors.

Limitations

It is ironic that the major limitations in our study concern the construct validity of our methods. We did not use real trial procedures, jurors did not deliberate, and the consequences of judgments were very different for our mock jurors than they would be for jurors in real cases. Furthermore, the participants in this research, primarily White Europeans living in the Midwest, analyzed case scenarios rather than more realistic trial evidence. However, like so many others before us, we argue that our results are still important to consider in the context of mock jury trials and, as Bornstein (1999) has argued so effectively, simulations like our own can and do make important contributions to the literature. Furthermore, our work examined empirically a potential limitation to the generalizability of mock jury findings, type of participant, and we have shown that jury researchers ought to consider the samples that they collect in the context of the types of charges, burdens of proof, standards of evidence, and other procedures that make up the trials we study.

References

- Berg, K. S., & Vidmar, N. (1975). Authoritarianism and recall of evidence about criminal behavior. *Journal of Research in Personality*, 9, 147-157. DOI: 10.1016/0092-6566(75)90025-2.
- Berman, G. L., & Cutler, B. L. (1996). Effects of inconsistencies in eyewitness testimony on mock-juror decision making. *Journal of Applied Psychology*, 81, 170-177. DOI: 10.1037/0021-9010.81.2.170.
- Bornstein, B. H. (1999). The ecological validity of jury simulations: Is the jury still out? *Law and Human Behavior*, 23, 75-91. DOI: 10.1023/A:1022326807441.
- Bray, R. M., & Kerr, N. L. (1979). Use of the simulation method in the study of jury behavior. *Law and Human Behavior*, 3, 107-119. DOI: 10.1007/BF01039151.
- Briere, J., Henschel, D., & Smiljanich, K. (1992). Attitudes toward sexual abuse: Sex differences and construct validity. *Journal of Research in Personality*, 26, 398-406. DOI: 10.1016/0092-6566(92)90067-E.
- Casper, J. D., Benedict, K., & Perry, J. L. (1989). Juror decision making, attitudes, and the hindsight bias. *Law and Human Behavior*, 13, 291-310. DOI: 10.1007/BF01067031.
- Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Boston, MA: Houghton Mifflin.
- Diamond, S. S. (1997). Illuminations and shadows from jury simulations. *Law and Human Behavior*, 21, 561-571. DOI: 10.1023/A:1024831908377.

- Kassin, S. M., & Wrightsman, L. S. (1983). The construction and validation of a juror bias scale. *Journal of Research in Personality*, *17*, 423-442. DOI: 10.1016/0092-6566(83)90070-3.
- Kovera, M. B. (2002). The effects of general pretrial publicity on juror decisions: An examination of moderators and mediating mechanisms. *Law and Human Behavior*, *26*, 43-72. DOI: 10.1023/A:1013829224920.
- Maeder, E. M., Wiener, R. L., & Winter, R. (2007). Does a truck driver see what a nurse sees? The effects of occupation type on perceptions of sexual harassment. *Sex Roles*, *56*, 801-810. DOI: 10.1007/s11199-007-9244-y.
- Mitchell, T. L., Haw, R. M., Pfeifer, J. E., & Meissner, C. A. (2005). Racial bias in mock juror decision making: A meta-analytic review of defendant treatment. *Law and Human Behavior*, *29*, 621-637. DOI: 10.1007/s10979-005-8122-9.
- Model Penal Code. x 2.02 (ULA, 1962; 1981).
- Rose, V. G., & Ogloff, J. R. P. (2001). Evaluating the comprehensibility of jury instructions: A method and an example. *Law and Human Behavior*, *25*, 409-431. DOI: 10.1023/A:1010659703309.
- Schuller, R. A., & Hastings, P. A. (1996). Trials of battered women who kill: The impact of alternative forms of expert evidence. *Law & Human Behavior*, *20*, 67-187. DOI: 10.1007/BF01499353.
- Severance, L. J., Greene, E., & Loftus, E. F. (1984). Toward criminal jury instructions that jurors can understand. *Journal of Criminal Law and Criminology*, *75*, 198-233. DOI: 10.2307/1143210.
- Simon, R. J., & Mahan, L. (1970). Quantifying burdens of proof: A view from the bench, the jury, and the classroom. *Law & Society Review*, *5*, 319-330. DOI: 10.2307/3052837.
- Vidmar, N. (1997). Generic prejudice and the presumption of guilt in sex abuse trials. *Law and Human Behavior*, *21*, 5-25. DOI: 10.1023/A:1024861925699.
- Vidmar, N. (2002). Case studies of pre- and midtrial prejudice in criminal and civil litigation. *Law and Human Behavior*, *26*, 73-105. DOI: 10.1023/A:1013881208990.
- Vidmar, N. (2008). Civil juries in ecological context: Methodological implications for research. In B. H. Bornstein, R. L. Wiener, R. Schopp, & S. L. Willborn (Eds.), *Civil juries and civil justice: Psychological and legal perspectives* (pp. 35-65). New York, NY: Springer Science+Business Media.
- Weiten, W., & Diamond, S. S. (1979). A critical review of the jury simulation paradigm: The case of defendant characteristics. *Law and Human Behavior*, *3*, 71-93. DOI: 10.1007/BF01039149.
- Wiener, R. L., Arnot, L., Winter, R., & Redmond, B. (2006). Generic prejudice in the law: Sexual assault and homicide. *Basic and Applied Social Psychology*, *28*, 145-155. DOI: 10.1207/s15324834basps2802_4.
- Wiener, R. L., Hackney, A., Kadela, K., Rauch, S., Seib, H., Warren, L., & Hurt, L. E. (2002). The fit and implementation of sexual harassment law to workplace evaluations. *Journal of Applied Psychology*, *87*, 747-764. DOI: 10.1037/0021-9010.87.4.747.

- Wiener, R. L., & Hurt, L. E. (1999). An interdisciplinary approach to understanding social sexual conduct at work. *Psychology, Public Policy, and Law*, 5, 556–595. DOI: 10.1037/1076-8971.5.3.556.
- Wiener, R. L., & Hurt, L. E. (2000). How do people evaluate social-sexual conduct: A psycholegal model. *Journal of Applied Psychology*, 85, 75–85. DOI: 10.1037/0021-9010.85.1.75.
- Wiener, R. L., Hurt, L., Russell, B., Mannen, K., & Gasper, C. (1997). Perceptions of sexual harassment: The effects of gender, legal standard, and ambivalent sexism. *Law and Human Behavior*, 21, 71–93. DOI: 10.1023/A:1024818110678.
- Wiener, R. L., Pritchard, C. C., & Weston, M. (1995). Comprehensibility of approved jury instructions in capital murder cases. *Journal of Applied Psychology*, 80, 455–467. DOI: 10.1037/0021-9010.80.4.455.
- Wiener, R. L., Richter, E., Humke, A. M., Jones, M. B., Maeder, E. M., & Cantone, J. (unpublished). Regulatory Focus and Jury Instruction Comprehension in Capital Murder Trials, *Motivation and Emotion*.
- Wiener, R. L., Rogers, M., Winter, R., Hurt, L. E., Hackney, A., Kadela, K., Seib, H., Rauch, S., Warren, L., & Morasco, B. (2004). Guided jury discretion in capital murder cases: The role of declarative and procedural knowledge. *Psychology, Public Policy, and Law*, 10, 516–576. DOI: 10.1037/1076-8971.10.4.516.
- Wiener, R. L., Voss, A. M., Winter, R. J., & Arnot, L. (2005). The more you see it, the more you know it: Memory accessibility and sexual harassment judgments. *Sex Roles*, 53, 807–820. DOI: 10.1007/s11199-005-8294-2.
- Wiener, R. L., & Winter, R. J. (2007). Totality of circumstances in sexual harassment decisions: A decision making model. In R. L. Wiener, B. H. Bornstein, R. Schopp, & S. Willborn (Eds.), *Social consciousness in legal decision making: Psychological perspectives* (pp. 171–195). New York: Springer Press. DOI: 10.1007/978-0-387-46218-9_9