

Thin slice expert testimony and mock trial deliberations[☆]



Caroline Titcomb Parrott^{a,*}, Stanley L. Brodsky^b, Jennifer Kelly Wilson^a

^a Taylor Hardin Secure Medical Facility, Tuscaloosa, AL, United States

^b The University of Alabama, Tuscaloosa, AL, United States

ARTICLE INFO

Available online 4 September 2015

Keywords:

Thin slice observations
Juror deliberations
Jury decision-making
Expert witness testimony
Insanity defense bias

ABSTRACT

This study examined impressions of expert witness testimony in a not guilty by reason of insanity defense on two outcomes: witness's credibility and verdict. Borrowing in part from the "thin slice" methodology, we assessed outcomes in a 2 (deliberating vs. non-deliberating jurors) $\times$ 3 (length of videotaped testimony) between-subjects design. In 30 mock juries, 188 participants viewed the testimony by a forensic psychologist; then half of the juries deliberated. Thinner slices of the testimony were defined by the lower (30 s long) and upper (5 min long) temporal bounds in the literature. The third, fuller testimony condition was 10 min long and served as the accuracy marker for the shorter sliced exposures. We aimed to explore potential consequences to jurors relying on impressions of the expert, and his or her opinion, and to test that effect post deliberation. Accounting for deliberation, brief impressions of expert credibility generally exerted a similar influence on credibility to fuller considerations. The essential finding was that a two-way interaction emerged from time slice and deliberation on verdict for jurors in the 30 s condition. Overall, predictive accuracy was found in the 5 min slice, yet accuracy was not supported in the predictions based on the shortest slice. Individually-formed impressions are not likely to translate to the verdict ballot post-deliberation. Instead, brief impressions are likely to be heavily influenced by deliberation. Implications for understanding how impression-based testimony evaluations translate from the jury box to the deliberation room are discussed.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

Jurors are asked to make a fair judgment based on the evidence presented to them. These judgments are susceptible to preconceptions, subjectivity, and information processing biases that influence decision-making (Salerno & Diamond, 2010). For example, almost a century of social psychology research has concluded that first impressions provide a predictive mechanism for how social judgments are made (Allport, 1937; Ambady & Rosenthal, 1992; Funder, 1987; Gray, 2008; Kenny & Albright, 1987; Kruglanski, 1989; Swann, 1984). Impression formation is particularly relevant to understanding how jurors evaluate evidence, assess witness credibility, and decide verdicts (Gray, 2008; Greene & Ellis, 2007). This study examined the influences of brief impressions of expert testimony and subsequent deliberations on decision-making.

2. Evaluations of expert testimony

The court holds expert witnesses to a higher standard than it does lay witnesses. Once on the stand, accepted as experts and permitted to testify, experts are deemed qualified and implicitly credible by the court. Still, the triers of fact evaluate experts and their testimony for its merit and utility. This situation poses challenges for both the triers of fact and expert witnesses. The experts may present complex and cognitively taxing testimony. The jurors, already under a high cognitive load, may rely on cognitive shortcuts to simplify information (Tversky & Kahneman, 1974). Given that much information is presented during a trial, reliance on heuristically-based judgments can be helpful and in fact, can lead to accurate impressions (Gray, 2008). However, over-reliance on cognitive shortcuts may lead to erroneous and overly subjective conclusions (e.g., Greene & Ellis, 2007; Kamin & Rachlinski, 1995; Kelman, Rottenstreich, & Tversky, 1996; Moore, 1989). One such instance may be relying on impression-based evidence interpretations.

This study examines how reliance on impressions may introduce the potential for bias in evaluating experts and their testimony. Evidence interpretation may be influenced by *first* impressions of courtroom players such as defendants, lawyers, and experts. It is also possible that when listening to or thinking back on testimony, jurors rely on cognitive shortcuts to consider only *brief excerpts* of testimony when making sense of case facts and forming opinions. These brief excerpts form

[☆] The authors would like to thank Caroline Crocker for providing materials from which the current stimuli were derived. We also thank Clay Shealy and Steve Prentice-Dunn for their contributions to the project's development, as well as Kerry Johnson for assisting in data collection.

* Corresponding author at: Taylor Hardin Secure Medical Facility, Tuscaloosa, AL, United States. Tel.: +205 556 7060; fax: +1 205 556 1198.

E-mail address: caroline.parrott@hardin.mh.alabama.gov (C.T. Parrott).

the "impressions" examined in the current study. We borrow from the thin slice literature as it relates to person-impression formation and judgments to inform how impressions were defined and measured in this study.

2.1. Thinly sliced exposures

As Ambady, Bernieri, and Richeson (2000) explain, "a thin slice is defined... as a brief excerpt of expressive behavior sampled from the behavioral stream" (p. 203–4). While thin slices can be verbal or nonverbal, they are characterized by "reducing or sometimes eliminating" information presented in the verbal exchange, as well as the context of the behavioral stream and the target's history (Ambady et al., 2000, p. 204). In the operational definition of a thin slice, it is less than 5 min, typically presented in slices anywhere from 30 s to 5 min (Ambady & Rosenthal, 1992). Thin slices of observations have accounted for significantly accurate predictions and judgments ($r = .39$) of individuals and situations being rated (Ambady, Hallahan, & Rosenthal, 1995; Ambady & Rosenthal, 1992). For instance, ratings based on only brief, thinly sliced exposures have predicted targets' personality characteristics, interpersonal styles, subjective well-being, psychopathology deception, and sales effectiveness (e.g., Ambady & Rosenthal, 1992; Carney, Colvin, & Hall, 2007; Fowler, Lilienfeld, & Patrick, 2009; Funder & Colvin, 1988; Gray, 2008; Pretsch, Flunger, Heckmann, & Schmitt, 2013).

2.2. Thin slices and the expert witness paradigm

Within a legal context, expectations of trial outcome and a defendant's criminal history have been predicted based on thin slices (Ambady, Krabbenhoft, & Hogan, 2006; Funder, 1995; Kenny, 1991; Kraus & Keltner, 2009; Kruglanski, 1989). However, no research has examined thinly sliced expert testimony and jury decision-making. Accuracy in thin slices research is the correspondence between a judgment and a criterion or as consensus among individuals judging a particular attribute of a third party (Ambady & Rosenthal, 1992; Gray, 2008; Kruglanski, 1989). In jury decision-making context, for example, one might compare the thin slice judgments of an expert to a criterion such as agreement with the expert (i.e., verdict) or to a consensus such as a jury's view of an expert's credibility. Predictions based on thin slices have also been found comparable to longer observations (Ambady & Rosenthal, 1992; Funder & Ozer, 1983). Thus, it is also possible to use thin slice methodology to test if mock jurors make the same decisions for verdict and credibility when exposed to a brief slice of testimony as they would if exposed to longer slices.

While jurors are rarely exposed to merely 30 s (or even just 5 min) of testimony, due to the cognitive demands of their task and the often complex nature of expert testimony, jurors may use impressions to incorporate expert evidence into their story of the events (Bennett & Feldman, 2003; Fiske & Taylor, 1991; Pennington & Hastie, 2003). While it is possible jurors may rely on only brief excerpts of testimony, what they actually do with that impression and how it influences their decisions has not yet been examined. It is not yet known if an expert can have just as much impact, and the same type of influence, in a few seconds of testimony as he or she would in longer testimonies. If so, trial strategists could use this knowledge to prepare witnesses and inform courtroom tactics, and the courts would better understand the potential impact of mere impressions of expert testimony on jurors' decision-making.

3. From the jury box to the deliberation room

Jurors are instructed to immerse themselves in deliberation to make legal sense out of often ambiguous or complicated evidence (Greene & Ellis, 2007). On the surface, this instruction buffers against potential errors that may arise from pre-existing beliefs and against jurors having

maximized their cognitive loads by the time deliberations begin. For example, deliberations have been shown to yield increased comprehension of trial information and instructions, and less attention to inadmissible testimony (e.g., Diamond, 1997; Diamond & Levi, 1996; Ellsworth, 1989; Kerwin & Shaffer, 1994). While jurors often rely on probative evidence when deliberating, heuristic processing that helps to shape jurors' evaluations of the evidence may also factor into deliberations (Gray, 2008; Greene & Ellis, 2007; Miller & Burgoon, 1982; Salerno & Diamond, 2010). However, generalizable findings regarding the effects of "deliberation" are limited due to methodological challenges and context-dependent issues (Diamond, 1997; Greene et al., 2002; Nietzel, McCarthy, & Kern, 1999; Nunez, McCrea, & Culhane, 2011; Salerno & Diamond, 2010). Nonetheless, pre-deliberation trial factors, such as expert testimony, have been linked with post-deliberation jury outcomes (see review by Devine, 2012; see also Diamond & Casper, 1992; Devine, Clayton, Dunford, Seying, & Pryce, 2001; Greene et al., 2002; Nietzel et al., 1999). It is not known whether jurors relying on brief impressions to interpret evidence while experts are on the stand are likely to carry those impressions into the deliberation room. Nor it is known how such impressions may interact with deliberation to influence decision-making.

4. Investigative framework and design

Research has yet to compare considerations of thinly sliced testimony with longer, fuller considerations of the testimony. The present study used different expert testimony slices to examine how reliance on cognitively shortened impressions may influence testimony interpretation and decision-making when a brief deliberation was included. We used a not guilty by reason of insanity (NGRI) case due to the intrinsic reliance on expert psychological testimony and documented heuristic processing of evidence in such cases (Louden & Skeem, 2007). The shortest slice included the 30 s of testimony in which the expert opined an unequivocal affirmation that the defendant's mental state at the time of the offense (MSO) was significantly impaired due to a mental disorder. We also included this same testimony in a 5-min long "thin slice" condition, as well as a 10-min long "fuller slice" condition (Ambady & Rosenthal, 1992, p. 3). We recognize that traditional thin slice exposures typically aim to minimize differences in information encoding between time slices (Ambady et al., 2000) and that testimony manipulations in this study are not "thin slices" in their purest form due to this influence. However, the influence of testimony content was minimized by providing all mock jurors with a case fact sheet that outlined the major aspects of the expert's testimony.

Our aim was to manipulate the degree to which the expert verbally presented this information as a function of time exposed to the testimony. Thus, the length of the testimony was varied to experimentally manipulate the behavioral stream upon which mock jurors may base their impressions. We tested for differences in mock juror ratings of expert credibility and verdict. We also included a deliberation condition to see if the thin slice phenomenon held after a deliberation. Finally, we tested the potential for impressions from each of these three time periods to interact differently with deliberation to influence credibility and verdict.

4.1. Hypotheses

Guided by the thin slice literature, we postulated that brief impressions would be as influential on decision-making as longer exposures. We hypothesized that witness credibility ratings and verdicts would not significantly differ between the 30 s and 5 min conditions, the 30 s and 10 min conditions, or between the 5 and 10 min conditions. We included a deliberation condition to see if the results hypothesized above would hold true for both non-deliberating and deliberating mock jurors. We generally anticipated that decisions in the 30-s and 5-min

conditions would be more impervious to deliberation effects than would those based on the 10-min condition.

5. Method

The study used a 2 (deliberation vs. no deliberation) $\times$ 3 (30 s, 5 min, or 10 min of expert testimony exposure) between-subjects factorial design. A between-subjects design was preferable to a within-groups design in order to isolate the potential effects of relying on a brief impression of testimony during deliberations. Moreover, measuring mock juror's expert credibility and verdict at several time points prior to deliberation would weaken the verisimilitude of the deliberation as real jurors may not enter deliberations with polarized verdicts (e.g., Hannaford-Agor, Hans, Mott, & Munsterman, 2002).

The dependent variables were witness credibility ratings and verdict. Mock jurors were randomly assigned to either the individual juror condition ($N = 92$), in which ratings were made independent of deliberations, or the jury deliberation condition ($N = 96$), in which ratings were made after deliberations. Mock jurors were also randomly assigned to one of the three lengths of testimony conditions: 30 s ($n = 60$), 5 min ($n = 65$), and 10 min ($n = 63$).

5.1. Sample size

The overall sample size was determined based on jury size and statistical power estimates for linear mixed modeling (LMM) analysis appropriate for the study design. The ideal mock jury size was deemed to be six, with a minimum of five members, appropriate for an equitable analysis across group size (e.g., McLafferty, 2004). A minimum of 20 to 30 groups was recommended for LMM analysis, preferably with at least five groups per cell (Maas & Hox, 2005), which was met or exceeded in this study. Each of the six conditions was comprised of data from five mock juries (at least 30 participants), for a total of 30 mock juries.

5.2. Participants

The mock jurors were 188 undergraduates from a large state university. The sample was primarily female (64.9% female and 35.1% male) and self-identified as Caucasian (84% White, 13.8% African American, and 2.2% from a different racial background). Demographics across the mock juries were approximately evenly distributed for age ($M = 18.88$, $SD = 1.12$), gender, race, political orientation, and exposure to mental illness. The mock jurors were randomly assigned to a mock jury of between five and eight jurors ($M = 6.27$, $SD = 0.87$) for an equitable analysis across group size (e.g., McLafferty, 2004; Morgan, 1996), with a six-member jury being preferred. The plurality of the juries was comprised of six jurors, with 20% five-member, 40% six-member, 33.3% seven-member, and 6.7% eight-member mock juries.

5.3. Procedure

5.3.1. Initial procedures

Each session was conducted with participants seated in a mock jury box configuration. All participants received the same description outlining the nature of the case including the charges, the defendant's NGRI plea, and the burden of proof being on the defendant to demonstrate with a preponderance of the evidence that s/he was insane at the time of the offense. The mock jurors were informed that they were to view "a short segment" of expert testimony called to the stand by the defense and encouraged to treat the material as they would if they were actual jurors.

5.3.2. Testimony manipulation

The participants watched the expert testimony via pre-recorded video. The testimony was extracted from the New York Supreme

Court case of *People v. Goldstein* (2004). The defendant had a long-standing schizophrenia diagnosis and pled NGRI to second degree murder for pushing a woman onto a subway track, killing her instantly. Details of the crime were adjusted to depict a subway system and city near the research location. The case was chosen due to its use in previous research (Crocker & Kovera, 2010), its reliance on expert testimony, and its real-life ambiguity in case outcome (*People v. Goldstein*, 2004).

Each jury of participants watched one of three videos (i.e., 30 s, 5 min, or 10 min). Three separate videos were recorded using the same experienced expert witness who was a forensic psychologist. Each condition included the same 95-word scripted exchange between the defense attorney and the expert, including the expert's unequivocal affirmation that the defendant's mental state at the time of the offense was significantly impaired due to a mental disorder. This script served as the 30 s, conclusion-only condition. The 716-word 5-min condition included this conclusion, as well as the expert's credentials, a description of schizophrenia and its effects, the defendant's mental health history, and a description of the expert's MSO opinion. The 3012-word 10-min condition included the 5 min scripted material, but elaborated on details that were provided in the case fact sheet to all participants.

All mock jurors received a one-page case fact sheet with basic facts presented by both the prosecution and defense, which was also included in the expert's testimony in the fuller conditions. Thus, the basic evidentiary facts were held constant across conditions and only the elaboration of the testimony was adjusted across time-slice conditions. The mock jurors then received further instructions regarding their task to determine guilt options (i.e., guilty, not guilty, and NGRI).

5.3.3. Deliberation procedures

Deliberating mock-jurors had 15 min to deliberate regarding evidence and verdict, after which they were to vote and reach a unanimous outcome. Pre-deliberation voting was not permitted in an effort to avoid verdict-driven discussion and to promote an evidence-driven approach (Davis, Kameda, Parks, Stasson, & Zimmerman, 1989; Devine et al., 2001; Greene et al., 2002). Seven deliberating mock juries were unable to reach a unanimous decision and deliberated for an additional 15-min time period (i.e., given an "Allen Charge," *Allen v. United States*, 1896). After 30 min, deliberations concluded, at which time six of the seven Allen charged juries returned a second deadlocked verdict.

5.3.4. Non-deliberating procedures

To improve statistical uniformity across conditions, non-deliberating mock jurors also participated in a 15-min discussion in which no case-relevant discussion was permitted. Instead, they were instructed to share and discuss their earliest memories of smell and taste, which is a nonintrusive, neutral topic found to generate group discussion among strangers (Brodsky, 2011). Non-deliberating mock jurors were told that the study was investigating the effects of case-irrelevant jury deliberations.

5.4. Measures

After the deliberation or group activity was concluded, each participant completed the Witness Credibility Scale (WCS), verdict ratings, the Insanity Defense Attitudes—Revised (IDAR) Scale, a Participant Reaction Form (PRF), and a demographics questionnaire.

5.4.1. Witness Credibility Scale (WCS)

The WCS was used to assess expert witness believability and effectiveness along four subscales: Knowledge, Likeability, Trustworthiness, and Confidence (Brodsky, Griffin, & Cramer, 2010). The participants rated the expert on 20 items using a 10-point Likert-type scale with an adjective and its antonym (e.g., unfriendly to friendly). All four subscales ($\alpha = .92$ for Confidence, $\alpha = .91$ for Likeability, $\alpha = .98$ for Trustworthiness, and $\alpha = .95$ for Knowledge) and the total ($\alpha = .90$) have adequate internal reliability.

5.4.2. Verdict ratings

Verdict decisions were collected on a dichotomous and continuous measure adapted from previous mock juror research (e.g., Brodsky et al., 2010). In the dichotomous rating, participants voted either guilty or NGRI. Mock jurors also rated the probability of the defendant's guilt, as well as level of punishment, on a 7-point Likert-type scale, with 7 as the maximum confidence of guilt and most punitive punishment. The items assessed (1) the defendant's guilt, from NGRI to Guilty; (2) juror confidence in their verdict; (3) level of punishment assigned; and (4) juror commitment to their punishment.

5.4.3. Insanity Defense Attitudes—Revised (IDAR)

Potential for bias against the insanity defense was assessed as a covariate using the 22-item IDAR developed by Skeem, Loudon, and Evans (2004), $\alpha = .85$. The measure assesses beliefs that mental illness is germane to criminal responsibility, as well as perceptions of misuse of the insanity defense and subsequent threats to public safety.

5.4.4. Participant reaction form (PRF)

The PRF consisted of forced-choice and open-ended questions regarding mock-juror participation and self-reported experience. Participant comprehension of the case instructions and standard of proof were also examined. Six mock jurors were dropped from further analysis due to low comprehension.

5.4.5. Demographics questionnaire

Information was elicited on participant age, gender, race, ethnicity, parental education, political affiliation, previous jury participation, and degree of exposure to mental illness.

5.5. Analyses

Mock juror decisions and ratings within an interacting group, such as a mock jury, may be dependent on the effect of being grouped together (Field, 2009). Therefore, a linear mixed model (LMM) analysis in SPSS 17.0 was used (Field, 2009) to account for the violation of the assumption of independence that leads to increase in probability of Type I errors (Tabachnick & Fidell, 2007). Deliberation condition and time exposed to testimony, as well as their interaction term, were designated as fixed factors. We anticipated that bias against the insanity defense, as measured by the IDAR scores, would predict significantly fewer NGRI verdicts and more punitive outcomes. Thus, bias against the insanity defense, as measured by the total IDAR score ($M = 80.35$, $SD = 19.85$; Range = 23 to 159), was included as a centered covariate in all models. The parametric assumptions, including equal variance and normal distributions, as well as the LMM assumption that intercepts in groups were normally distributed around the overall model, were not violated. The analyses confirmed that LMM procedures were appropriate due to the group's statistically significant effect on verdict (continuous outcome), $F(1, 13.99) = 55.82$, $p = .020$, with $d = 2.33$ (large effect) and $ICC = .54$ (roughly 54% of the variance in verdict being accounted for by the grouping variable).

LMM procedures were used to generate an F and corresponding p value for each possible main effect and interaction. Minimum-effect tests¹ (Murphy & Myors, 1999) were employed to test the presence of only negligibly small differences. Any generated F was compared to an F critical value to test for only “negligibly small” differences between each of the testimony conditions (Murphy & Myors, 1999, p. 235). Any effects accounting for 5% or less of the variance in outcomes were labeled *negligible* (Murphy & Myors, 1999).

¹ Although minimum effects tests are designated for use with general linear models, it is preferable to traditional null hypothesis testing (Murphy & Myors, 1999; Thompson, 1997), and theoretically, the LMM's accounting for violating the assumption of independence should permit use of such tests in the present study (Field, 2009).

6. Results

Bias against the insanity defense, as measured by total IDAR scores, failed to exert a main effect on witness credibility, $F(1, 171.87) = 1.67$, $p = .199$, a small effect compared to $F^{crit5\%}$ ($\alpha = .05$) of 20.52. However, this bias did have a significant main effect on defendant guilt, $b = .51$, $F(1, 168.73) = 23.92$, $p < .001$, so that as bias increased, so did the likelihood of guilty verdicts. IDAR scores failed to significantly interact with impression slice or deliberation to predict either outcome. Due to its effect on verdict, IDAR scores were included in all models.

Hypothesis 1. Negligible differences among all three testimony conditions on witness credibility.

Results yielded a statistically significant main effect of impression slice on overall witness credibility, $F(2, 25.48) = 9.02$, $p = .001$. LSD post-hoc comparisons showed that the jurors in the 30 s condition judged the expert as significantly less credible ($M_{30sec} = 141.180$, $SD = 4.7$) than jurors in the 5 and 10 min conditions ($M_{5min} = 160.56$, $SD = 4.5$, $p = .007$; $M_{10min} = 168.17$, $SD = 4.5$, $p < .001$), whose ratings were not significantly different from each other. A critical $F(\alpha = .05)$, or $F^{crit5\%}$ was established for witness credibility at 5.34. The $F^{crit5\%}$ is used to test the hypothesis that impression slice accounts for 5% or less of the variance in witness credibility. Thus, effects on witness credibility are large enough to reject the hypothesis that the true effects of impression slice are negligible, suggesting a meaningful (and perhaps encouraging) effect of impression slice on credibility. Moreover, witness credibility and likelihood of guilty verdicts were significantly negatively correlated, $r = -0.28$, $n = 188$, $p < .001$, such that increased credibility of the defense expert was associated with more lenient verdicts—as expected.

Hypothesis 2. Negligible verdict differences among all three testimony conditions.

Table 1 presents the means for all continuous verdict ratings grouped by independent variable. The primary verdict outcome asked mock jurors to rate defendant guilt from 1 (definitely NGRI) to 7 (definitely guilty). Defendant guilt did not differ significantly across impression slice, $F(2, 23.76) = 2.86$, $p = .077$, although there was a trend for mock jurors in the 30 s condition to select more punitive verdicts than those in the 5 and 10 min conditions. Compared to the $F^{crit5\%}$ ($\alpha = .05$) of 5.30, effects on verdict are *not* large enough to reject the hypothesis that the true effects of time slice are negligible. Thus, results *cannot* reject the hypothesis that time slice may only account for 5% or less of the variance in verdict, suggesting a small effect on guilt determinations. Regarding the additional verdict and sentencing outcomes, no significant main effects or interactions were found for confidence in verdict, commitment to punishment, or level of punishment assigned.

6.1. Exploratory analysis

6.1.1. Exploring deliberation as a main effect or interaction

When controlling for impression slice, there was no significant main effect of deliberation on juror verdict, $F(1, 23.98) = 0.11$, $p = .749$, nor was there a main effect of deliberation on witness credibility, $F(1, 26.01) = 1.25$, $p = .270$. However, for just the 30 s condition, deliberation condition did influence verdict, although not credibility. A significant two-way interaction emerged between deliberation and impression slice on defendant guilt, $F(2, 23.78) = 3.69$, $p = .040$.

As shown in Fig. 1, a test of simple effects indicated that the mock jurors exposed to the witness for only 30 s showed a significant difference in defendant guilt ratings between deliberation conditions, $F(1, 24.75) = 5.58$, $p = .026$. Based on the $F^{crit5\%}$ ($\alpha = .05$) of 8.74, effects of this interaction on guilt may only account for 5% or less of the variance in verdict, suggesting a small effect on guilt determinations.

Table 1
Means (and standard deviations) for juror verdict ratings by independent variable.

Conditions		Continuous verdict variables ^a			
		Item 1 <i>Is the defendant guilty?</i>	Item 2 <i>How confident are you about that verdict?</i>	Item 3 <i>What should the level of punishment be?</i>	Item 4 <i>How committed to your sentencing decision are you?</i>
Deliberation	Filler task	3.67 (.31)	5.54 (.16)	4.32 (.20)	5.43 (.14)
	Deliberation	3.81 (.31)	5.89 (.16)	4.64 (.20)	5.79 (.13)
Time slices	30 s	4.48 (.38)	5.78 (.20)	4.42 (.28)	5.48 (.17)
	5 min	3.36 (.37)	5.62 (.19)	4.60 (.24)	5.60 (.16)
	10 min	3.40 (.24)	5.73 (.19)	4.42 (.24)	5.75 (.16)

Item 1: *Definitely NGRI to Definitely Guilty.*

Item 2: *Not at all Confident to Very Confident.*

Item 3: *As Lenient as Possible to As Harsh as Possible.*

Item 4: *Not at all Confident to Very Confident.*

^a Items scored on a Likert-type scale from 1 to 7:

Nevertheless, the LSD post-hoc comparisons showed that in the 30 s exposure, non-deliberating jurors were more lenient and more likely to find the defendant NGRI than were deliberating jurors, while there were no significant differences in defendant guilt between deliberation conditions for the 5 min or 10 min conditions.

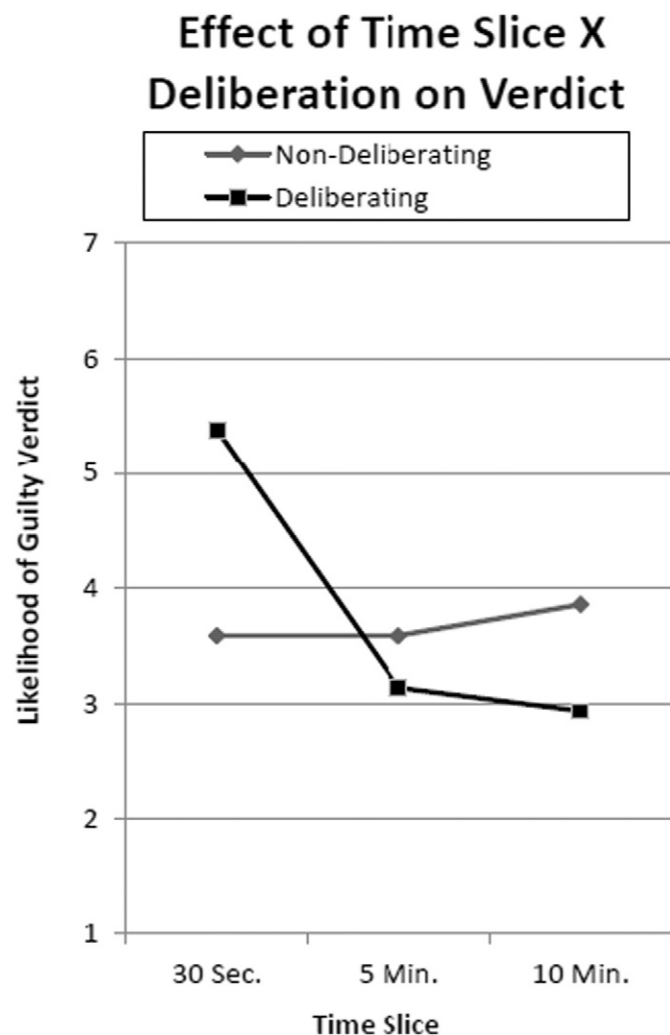


Fig. 1. Significant two-way interaction between time slice and deliberation on verdict (defendant guilt). Verdict assessed as likelihood of guilty rating from 1 (definitely NGRI) to 7 (definitely guilty).

6.1.2. A closer look at witness credibility

The four facets of the WCS (i.e., Trustworthiness, Confidence, Likeability, and Knowledge) were also subjected to analyses within the LLM framework. The results yielded main effects for time slice on Confidence, $F(2, 24.73) = 5.91, p = .008$, on Knowledge, $F(2, 24.74) = 8.24, p = .002$, and on Trustworthiness, $F(2, 25.01) = 7.63, p = .003$, suggesting more than small effects with an $F^{crit5\%} (\alpha = .05)$ of 5.34. There was also a main effect for time slice on Likeability, $F(2, 182) = 17.88, p < .001$, suggesting a large enough effect to be meaningful with an $F^{crit5\%} (\alpha = .05)$ of 10.86. Table 2 shows that mock jurors in the 30 s group rated the witness as significantly lower on each of these four positive WCS variables than did jurors in the 5-min and 10-min condition. For Trustworthiness and Likeability, the differences between ratings were significant between all three time slices. A significant two-way interaction between deliberation and time slice, $F(2, 25.01) = 4.47, p = .022$ emerged for Trustworthiness only. Results yielded a small effect with a significant difference between deliberation conditions for the 5 min group, $p = .047$, and a trend in difference for the 30 s group, $p = .072$. When compared to non-deliberating counterparts, deliberating mock jurors rated the expert as *more* trustworthy after 5 min, but as potentially *less* trustworthy if only shown 30 s of testimony.

6.2. The PRF: Convergence of self-reported data

A significant two-way interaction was found in additional LLM analyses between deliberation and time slice on jurors' self-reported use of testimony on verdict, $F(2, 22.53) = 6.42, p = .006$. A test of simple effects confirmed that this interaction exists for the 30 s condition, $F(1, 24.18) = 7.22, p = .013$, as well as for the 5 min condition, $F(1, 22.46) = 5.24, p = .033$. LSD post-hoc comparisons indicated that deliberating jurors exposed to the witness for only 30 s relied significantly less on the testimony when compared to the non-deliberating jurors—suggesting that perhaps deliberation enhanced discussion of brief impressions (Fig. 2).

Table 2

Means and standard deviations for mock juror ratings by experimental condition on Confidence, Knowledge, Trustworthiness, and Likeability.

WCS Facet				
Time Slice	Confidence	Knowledge	Trustworthiness	Likeability
30 s	35.25 (1.16)	36.60 (1.37)	34.20 (1.48)	33.50 (1.09)
5 min	40.42 _a (1.12)	43.32 _b (1.34)	38.91 (1.43)	38.31 (1.03)
10 min	41.54 _a (1.13)	43.43 _b (1.35)	41.26 (1.43)	41.81 (1.03)

Note. Comparison of means between time slices can be interpreted within WCS facet columns. Means in the same column that share subscripts do not differ significantly, $p > .05$. All other means within the same column do differ significantly, $p < .05$.

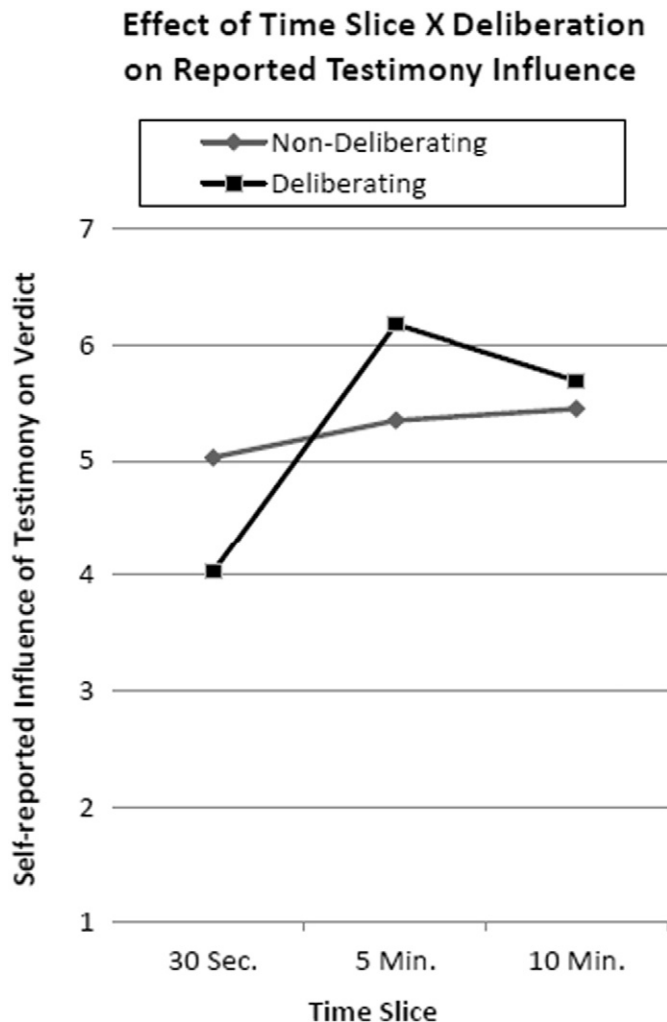


Fig. 2. Significant two-way interaction between time slice and deliberation on participant perceptions of degree of influence testimony had on verdict (mock juror self-report). Influence rated from 1 (not at all) to 7 (very much).

7. Discussion

This study experimentally manipulated the mock jurors' exposure to brief impressions of expert testimony compared to longer testimony. We sought to test the predictive utility of relying on impressions of testimony in decision-making and to see how such impressions held up after deliberation. Results yielded a complex picture indicating that while reliance on brief impressions of testimony could introduce bias in evaluating experts and their testimony, such reliance may be heavily influenced by deliberation. This is the first experimental manipulation borrowing from thin slice methodology to examine these constructs.

7.1. Witness credibility and thin slices

We did not find negligible differences among all impression slice conditions as had predicted. When deliberation condition was controlled, the briefest slice of expert exposure (30 s) generated significantly lower credibility ratings than the other conditions. This difference does not run completely counter to the literature. Recall that a thin slice is defined as anywhere from 30 s to 5 min of testimony. For total credibility, as well as expert confidence and knowledge, only the 30 s conditions revealed significant lower scores, while differences between the 5 and 10 min conditions were negligible. Thus, it appears a useful "thin slice" of expert credibility, and in particular expert confidence and knowledge, may translate to between 30 s and 5 min of exposure.

Total credibility, confidence, and knowledge improved from 30 s to 5 min across both deliberation conditions. The direction of this result suggests that the briefest slice of testimony generated a negative impression of the expert's credibility relative to the other conditions. Expert witnesses may view this finding as good news as their credibility may not hinge on very brief impervious impressions of testimony (i.e., 30 s clips).

Experts are viewed differentially when jurors rely on very brief impressions of testimony. Jurors are charged with fairly and justly evaluating the expert's credibility and the relative weight of his or her testimony. Mock jurors relying on the briefest sliced impressions of an expert may form a different, and in this case a poorer, opinion of that expert's credibility than if they had considered a fuller segment of the testimony. Expert likeability and trustworthiness revealed even more robust increases across all three time points.

Keep in mind that when the presence or absence of deliberation was considered, the effect of impression slice on credibility dissipated. Deliberation appears to mitigate the difference between brief impression-based credibility judgments and fuller, more informed judgments. The one exception was related to trustworthiness. This variable emerged as the only credibility component susceptible to an interaction between time slice and deliberation. Compared to the "full" 10-min testimony, post deliberation trustworthiness ratings based on brief impressions potentially decreased in the 30 s condition yet significantly increased in the 5 min condition. Thus, ratings of the other three witness credibility facets, even when based on very brief impressions, were more impervious to group influence.

7.2. Verdict and thin slices

Verdict ratings between briefly sliced and longer testimony exposures were best illuminated by the two-way interaction of time slice and deliberation on verdict. For non-deliberating mock jurors, no difference in verdict ratings emerged across impression slices, which supports the thin slice phenomena. If jurors were never to deliberate, then a brief impression of testimony could have just as much influence on verdict as longer considerations of the testimony.

However, jurors deliberate and according our findings, this may assist in mitigating the influence impressions of expert testimony may otherwise have on verdicts. In this study, jury deliberations were shown to influence the accuracy of briefly formed judgments. Deliberating mock jurors in the 30 s condition yielded significantly more punitive verdicts than their 5 and 10 min counterparts. This finding has two corollaries. First, deliberation matters when seeking to understand how impressions of testimony may be used in decision-making. Second, the 5 min condition yielded "accurate" verdicts as compared to the longer 10 min condition. The briefest impressions lose their accuracy after deliberation, while slightly longer, yet still briefly sliced impressions may yield more accurate and unbiased evidence interpretations.

These findings inform our understanding of how jurors interpret testimony and make decisions. Brief impressions of expert witness testimony may be formed for individually determined judgments just as we manipulated in this experiment. However, if brought to the deliberation room, these individually-formed impressions interact with deliberation to modify their influence. In this study, a leniency shift was shown for deliberating jurors permitted more exposure to the testimony.

7.3. Empirical strengths

A primary strength of this exploration of the jury process is its theoretical support from the social psychological and psycho-legal literatures (Wiener, Krauss, & Lieberman, 2011). We adapted our methodology from thin slice research to vary length of exposure to expert testimony and to assess predictive utility of impression-based credibility ratings in the courtroom. Ecological validity was improved by use

of a forensic clinical psychologist as the expert and the use of an actual trial transcript. Standardized scripts, assessments of mock juror experiences and instruction comprehension, and counter-balancing of questionnaires all strengthened the internal reliability of the findings. Other aspects of the design (e.g., filler task inclusion) and analysis (LMM procedures) were also followed in order to account for nested mock juries, which allowed for the inclusion of deliberation, perhaps the study's most noteworthy strength (Nunez et al., 2011).

7.4. Limitations and empirical considerations

Several limitations were present in this study's design and others were imposed by logistical constraints. First, to maximize control across conditions, one expert was chosen for all conditions, which limits generalizations. Second, mock jurors only viewed testimony from the defense. To lessen the effect of this limitation (e.g., Salerno & McCauley, 2009), mock jurors were provided study materials that offered an equivalent degree of case facts presented by both the defense and the prosecution. Third, the "full" testimony condition was limited to 10 min. While double the length of the upper bound of a traditional thin slice exposure, this length of time is shorter than the length of most testimonies. Fourth, the lack of a live trial simulation and the uniformly limited time allotment for deliberation are noted. This study shares in the common criticism that post-deliberation outcomes yielded from brief deliberations may be more reflective of a first ballot vote than an outcome evinced by a full deliberation (Salerno & Diamond, 2010). Finally, future studies could benefit from a more diverse, community-based participant pool (McCabe & Krauss, 2011).

7.5. Conclusions and future directions

The current results suggest that reliance on brief impressions of expert testimony may not influence decision-making as much as fuller considerations of the evidence. Borrowing from thin slice methodology emerged from this investigation as an informative way to study the influence of shorter and longer evidence presentations. Additional research may explore a more stringent "thin slice" methodology to focus on even shorter slices of testimony or narrow stimuli to nonverbal behavior. Future research to understand the course of impression formation in expert testimony interpretation is warranted. Within subjects, real-time studies may reveal critical time points in testimony and when jurors are or are not centrally processing the evidence being presented. Studies should also consider the possible impact of individual differences in need for cognition and expert testimony impression formation, content, influence and fortitude.

The study of impression formation with testable legal scenarios allows for theory-driven investigations into the processes behind jury decision-making and may lead to better informed generalizations of empirical findings (Wiener et al., 2011). Much is still unknown about the influence of group deliberation effects and how information-processing biases play out in this context, particularly in cases involving mental health testimony. Continued research in this area should consider mixed methodologies to assess potential deliberation influences on evidence interpretation and impression formation or malleability. As Ambady et al. (1995) emphasize, "Whereas judgments of others may be formed very quickly and often unwittingly, they may also be consequential and long-lasting" (p. 4). No place may this be truer than when justice and the lives of others lie in the hands of those unwittingly and perhaps prejudicially making such judgments.

References

- Allen v. United States, 164 U.S. 492 (1896).
 Allport, G.W. (1937). *Personality: A psychological interpretation*. New York: Holt.
 Ambady, N., Bernieri, F.J., & Richeson, J.A. (2000). Toward a histology of social behavior: Judgmental accuracy from thin slices of the behavioral stream. *Advances in Experimental Social Psychology*, 32, 201–271.
 Ambady, N., Hallahan, M., & Rosenthal, R. (1995). On judging and being judged accurately in zero-acquaintance situations. *Journal of Personality and Social Psychology*, 69, 518–529. <http://dx.doi.org/10.1037/0022-3514.69.3.518>.
 Ambady, N., Krabbenhoft, M.A., & Hogan, D. (2006). The 30-sec scale: Using thin-slice judgments to evaluate sales effectiveness. *Journal of Consumer Psychology*, 16, 4–13.
 Ambady, N., & Rosenthal, R. (1992). Thin slices of behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological Bulletin*, 2, 256–274.
 Bennett, W.L., & Feldman, M.S. (2003). Storytelling in the courtroom. In D.D. Koski (Ed.), *The jury trial in criminal justice* (pp. 282–291). Durham, North Carolina: Carolina Academic Press.
 Brodsky, S.L. (2011). *Therapy with coerced and reluctant clients*. Washington, DC: American Psychological Association Book.
 Brodsky, S.L., Griffin, M.P., & Cramer, R.J. (2010). The witness credibility scale: An outcome measure for expert witness research. *Behavioral Sciences & the Law*, 28, 892–907.
 Carney, D.R., Colvin, C.R., & Hall, J.A. (2007). A thin slice perspective on the accuracy of first impressions. *Journal of Research in Personality*, 41, 1054–1072. <http://dx.doi.org/10.1016/j.jrp.2007.01.004>.
 Crocker, C.B., & Kovera, M.B. (2010). The effects of rehabilitative voir dire on juror bias and decision making. *Law and Human Behavior*, 34, 212–226. <http://dx.doi.org/10.1007/s10979-009-9193-9>.
 Davis, J.H., Kameda, T., Parks, C., Stasson, M., & Zimmerman, S.K. (1989). Some social mechanics of group decision-making: The distribution of opinion, polling sequence, and implications of consensus. *Journal of Personality and Social Psychology*, 57, 1000–1012.
 Devine, D.J. (2012). *Psychology and crime series: Jury decision-making: The state of the science*. New York: New York University Press.
 Devine, D.J., Clayton, L.D., Dunford, B.B., Seying, R., & Pryce, J. (2001). Jury decision-making: 45 Years of empirical research on deliberating groups. *Psychology, Public Policy, and Law*, 7, 622–727. <http://dx.doi.org/10.1037/1076-8971.7.3.622>.
 Diamond, S.S. (1997). Illuminations and shadows from jury simulations. *Law and Human Behavior*, 21, 561–571. <http://dx.doi.org/10.1023/A:1024831908377>.
 Diamond, S.S., & Casper, J.D. (1992). Blindfolding the jury to verdict consequences: Damages, experts, and the civil jury. *Law and Society Review*, 26, 513–564.
 Diamond, S.S., & Levi, J.N. (1996). Improving decisions on death by revising and testing jury instructions. *Judicature*, 79, 224–232.
 Ellsworth, P.C. (1989). Are twelve heads better than one? *Law and Contemporary Problems*, 52, 207–224.
 Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). Thousand Oaks, California: Sage Publications, Inc.
 Fiske, S.T., & Taylor, S.E. (1991). *Social cognition* (2nd ed.). New York: McGraw-Hill.
 Fowler, K.A., Lilienfeld, S.O., & Patrick, C.J. (2009). Detecting psychopathy from thin slices of behavior. *Psychological Assessment*, 21, 68–78.
 Funder, D.C. (1987). Errors and mistakes: Evaluating the accuracy of social judgment. *Psychological Bulletin*, 101, 75–91.
 Funder, D.C. (1995). On the accuracy of personality judgment: A realistic approach. *Psychological Review*, 102, 652–670.
 Funder, D.C., & Colvin, C.R. (1988). Friends and strangers: Acquaintanceship, agreement, and the accuracy of personality judgment. *Journal of Personality and Social Psychology*, 55, 149–158.
 Funder, D.C., & Ozer, D.J. (1983). Behavior as a function of the situation. *Journal of Personality and Social Psychology*, 44, 107–112.
 Gray, H.M. (2008). To what extent, and under what conditions, are first impressions valid? In N. Ambady, & J.J. Skowronski (Eds.), *First Impressions* (pp. 106–128). New York: The Guilford Press.
 Greene, E., Chopra, S. R., Kovera, M. B., Penrod, S. D., Rose, V. G., Schuller, R., et al. (2002). Jurors and juries: A review of the field. In R. Roesch (Series Ed.) & J. R. P. Ogloff (Vol. Ed.), *Perspective in law and psychology: Vol. 14. Taking psychology and law into the twenty-first century* (pp. 225–284). New York: Kluwer Academic/ Plenum Publishers.
 Greene, E., & Ellis, L. (2007). Decision making in criminal justice. In D. Carson, R. Milne, F. Pakes, K. Shalev, & A. Shawyer (Eds.), *Applying psychology to criminal justice* (pp. 183–200). England: John Wiley & Sons Ltd.
 Hannaford-Agor, P.L., Hans, V.P., Mott, N.L., & Munsterman, G.T. (2002). Are hung juries a problem? The National Center for State Courts. Retrieved on August 30, 2012 from http://www.ncsconline.org/WC/Publications/Res_Juries_HungJuriesProblemPub.pdf
 Kamin, K., & Rachlinski, J. (1995). Ex post ≠ ex ante: Determining liability in hindsight. *Law and Human Behavior*, 19, 89–104.
 Kelman, M., Rottenstreich, Y., & Tversky, A. (1996). Context-dependence in legal decision-making. *Journal of Legal Studies*, 25, 287–297.
 Kenny, D.A. (1991). A general mode of consensus and accuracy in interpersonal perception. *Psychological Review*, 98, 155–163.
 Kenny, D.A., & Albright, L. (1987). Accuracy in interpersonal perception: A social relations analysis. *Psychological Bulletin*, 102, 390–402.
 Kerwin, J., & Shaffer, D.R. (1994). Mock jurors versus mock juries: The role of deliberations in reactions to inadmissible testimony. *Personality and Social Psychology Bulletin*, 20, 153–162. <http://dx.doi.org/10.1177/0146167294202002>.
 Kraus, M.W., & Keltner, D. (2009). Signs of socioeconomic status: A thin-slicing approach. *Psychological Science*, 20(1), 99–106.
 Kruglanski, A.W. (1989). The psychology of being "right": The problem of accuracy in social perception and cognition. *Psychological Bulletin*, 106, 395–409.
 Loudon, J.E., & Skeem, J.L. (2007). Constructing insanity: Jurors' prototypes, attitudes, and legal decision-making. *Behavioral Sciences and the Law*, 25, 449–470.
 Maas, C.J.M., & Hox, J.J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology*, 1(3), 86–92.
 McCabe, J.G., & Krauss, D.A. (2011). The effect of acknowledging mock jurors' feelings on affective and cognitive biases: It depends on the sample. *Behavioral Sciences and the Law*, 29, 331–357. <http://dx.doi.org/10.1002/bsl.990>.

- McLafferty, I. (2004). Focus group interviews as a data collecting strategy. *Journal of advanced nursing*, 48(2), 187–194. <http://dx.doi.org/10.1111/j.1365-2648.2004.03186.x>.
- Miller, G.R., & Burgoon, J.K. (1982). Factors affecting witness credibility. In N.L. Kerr, & R.M. Bray (Eds.), *The psychology of the courtroom*. San Diego, CA: Academic Press, Inc.
- Moore, A.J. (1989). Trial by scheme: Cognitive filters in the courtroom. *UCLA Law Review*, 37, 273–341.
- Morgan, D. (1996). Focus groups. *Annual Review of Sociology*, 22, 129–152.
- Murphy, K.R., & Myers, B. (1999). Testing the hypothesis that treatments have negligible effects: Minimum-effect tests in the general linear model. *Journal of Applied Psychology*, 84, 234–248. <http://dx.doi.org/10.1037/0021-9010.84.2.234>.
- Nietzel, M.T., McCarthy, D.M., & Kern, M.J. (1999). Juries: The current state of the empirical literature. In R. Roesch, S.D. Hart, & J.R.P. Ogloff (Eds.), *Psychology and law: The state of the discipline* (pp. 23–42). New York: Kluwer/Plenum.
- Nunez, N., McCreary, S.M., & Culhane, S.M. (2011). Jury decision making research: Are researchers focusing on the mouse and not the elephant in the room? *Behavioral Sciences and the Law*, 29, 439–451. <http://dx.doi.org/10.1002/bsl.967>.
- Pennington, N., & Hastie, R. (2003). The story model for juror decision-making. In D.D. Koski (Ed.), *The jury trial in criminal justice* (pp. 292–300). Durham, North Carolina: Carolina Academic Press.
- People v. Goldstein*, 14 A.D.3d 32, 786 N.Y.S.2d 428 (N.Y. 2004).
- Pretsch, J., Flunger, B., Heckmann, N., & Schmitt, M. (2013). Done in 60 seconds? Inferring teachers' subjective well-being from thin slices of nonverbal behavior. *Social Psychology of Education*, 16, 421–434. [http://dx.doi.org/10.1016/S0065-2601\(00\)80006-4](http://dx.doi.org/10.1016/S0065-2601(00)80006-4).
- Salerno, J.M., & Diamond, S.S. (2010). The promise of a cognitive perspective on jury deliberation. *Psychonomic Bulletin and Review*, 17, 174–179.
- Salerno, J.M., & McCauley, M.R. (2009). Mock jurors' judgments about opposing scientific experts: Do cross-examination, deliberation, and need for cognition matter? *American Journal of Forensic Psychology*, 27, 37–60.
- Skeem, J.L., Louden, J.E., & Evans, J. (2004). Venirepersons' attitudes towards the insanity defense: Developing, refining, and validating a scale. *Law and Human Behavior*, 28(6), 623–648.
- Swann, W.B. (1984). Quest for accuracy in person perception: A matter of pragmatics. *Psychological Review*, 91, 457–477.
- Tabachnick, B.G., & Fidell, L.S. (2007). *Using multivariate statistics* (5th ed.). Boston: Allyn & Bacon.
- Thompson, B. (1997, August). If statistical significance tests are broken/misused, what practices should supplement or replace them? *Paper presented at the 105th Annual Convention of the American Psychological Association, Chicago*.
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 27, 1124–1131.
- Wiener, R.L., Krauss, D.A., & Lieberman, J.D. (2011). Mock jury research: Where do we go from here? *Behavioral Sciences and the Law*, 29, 467–479. <http://dx.doi.org/10.1002/bsl.989>.