

# Testing the Waters: An Investigation of the Impact of Hot Tubbing on Experts From Referral Through Testimony

Jennifer T. Perillo<sup>1</sup>, Anthony D. Perillo<sup>1</sup>, Nikoleta M. Despodova<sup>2, 3</sup>, and Margaret Bull Kovera<sup>2, 3</sup>

<sup>1</sup> Department of Psychology, Indiana University of Pennsylvania

<sup>2</sup> Department of Psychology, The Graduate Center, City University of New York

<sup>3</sup> John Jay College of Criminal Justice, City University of New York



**Objective:** The present research examined whether concurrent expert testimony (“hot tubbing”) and court-appointed testimony reduced adversarial allegiance in clinical experts’ judgments compared with traditional adversarial expert testimony. **Hypotheses:** We predicted Hypothesis 1: Defense experts would render more not responsible judgments and lower ratings of criminal responsibility than would prosecution experts; Hypothesis 2: Adversarial allegiance effects on experts’ judgments would be heightened for adversarial experts and attenuated for concurrent experts over time; Hypothesis 3: Adversarial and concurrent experts would report higher dissonance than would court-appointed experts and adversarial experts’ ratings would increase over time, concurrent experts’ ratings would decrease, and court-appointed experts’ ratings would remain unchanged. **Method:** Clinicians and advanced clinical doctoral students conducted simulated criminal responsibility evaluations for the prosecution, defense, or court. We categorized participants as favoring the prosecution or defense based on their preexisting attitudes and randomly assigned them to the adversarial, concurrent, or court-appointed expert testimony conditions. Participants completed a dichotomous responsibility judgment, strength of responsibility ratings, and cognitive dissonance measure after initial evidence review ( $n = 93$ ), report completion ( $n = 52$ ), and testimony ( $n = 48$ ). Concurrent experts generated a joint report outlining areas of agreement and disagreement before providing testimony. **Results:** Concurrent testimony did not eliminate adversarial allegiance. Adversarial and concurrent experts’ perceptions of responsibility did not significantly differ ( $d = .04$ , 95% CI  $[-.64, .71]$ ) or change over time ( $\eta_p^2 = .03$ ); however, prosecution experts—across testimony types—rated the defendant as significantly more responsible than did defense experts ( $d = 1.87$ , 95% CI  $[1.06, 2.67]$ ). Concurrent and adversarial experts did not differ in their reports and minimally differed in testimony content. **Conclusions:** Experts who initially favored the prosecution or defense showed adversarial allegiance regardless of expert testimony method, and we observed no attenuation of this bias over the course of their case involvement.

## Public Significance Statement

Two alternatives to traditional adversarial expert testimony are concurrent testimony, in which opposing experts collaborate on a report to the court and testify together, and court-appointed testimony. Concurrent testimony did not reduce or eliminate experts’ initial bias in favor of the prosecution or defense when rendering judgments during a simulated criminal responsibility hearing; court-appointed experts’ judgments remained moderate, but that may reflect an aggregation of biased

David DeMatteo served as Action Editor.

Jennifer T. Perillo  <https://orcid.org/0000-0003-1634-2892>

Anthony D. Perillo  <https://orcid.org/0000-0001-6153-2236>

Nikoleta M. Despodova  <https://orcid.org/0000-0002-8714-778X>


Margaret Bull Kovera  <https://orcid.org/0000-0002-2551-8434>

We have no known conflicts of interest to disclose. This research was supported by a grant from the National Science Foundation (NSF) to Margaret Bull Kovera under Grant SBE-1023796. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF. We are also indebted to Patricia Zapf for providing the criminal responsibility training to our participants through CONCEPT and Thomas Kucharski for reviewing

the clinical case materials. Thanks are also due to Anthony Connors and Max Seltzer for serving as the attorneys in the research study. Finally, we wish to acknowledge the assistance of the following research assistants: Angela Jones, Lauren Clatch, and Jennifer Teitcher. Portions of the data were previously presented at the 2013, 2014, and 2015 meetings of the American Psychology-Law Society (Portland, Oregon, New Orleans, Louisiana, and San Diego, California, respectively).

Data and materials are available at <https://osf.io/u23tc/>.

 The data are available at <https://osf.io/u23tc/>

 The experiment materials are available at <https://osf.io/u23tc/>

Correspondence concerning this article should be addressed to Jennifer T. Perillo, Department of Psychology, Indiana University of Pennsylvania, 1020 Oakland Avenue, Indiana, PA 15705, United States. Email: [jperillo@iup.edu](mailto:jperillo@iup.edu)

judgments rather than attenuation of bias. Additional research examining concurrent testimony and other potential alternatives, such as joint conferences between opposing experts *before* examining evidence and writing reports, is necessary to help inform the legal system's efforts to minimize the undesirable effects of adversarial allegiance in trials.

*Keywords:* expert testimony, adversarial allegiance, cognitive bias, forensic assessment

*Supplemental materials:* <https://doi.org/10.1037/lhb0000446.supp>

Undoubtedly there is a natural bias to do something serviceable for those who employ you and adequately remunerate you. It is very natural, and it is so effectual that we constantly see persons, instead of considering themselves witness, rather consider themselves as the paid agents of the persons who employ them.

(*Lord Arbinger v Ashton*, 1873, p. 374)

In legal cases, complex information is often presented to jurors who may not have the necessary background to use it effectively. The Federal Rules of Evidence (FRE) provide for expert witnesses in these circumstances to “help the trier of fact to understand the evidence or to determine a fact in issue” (*Federal Rules of Evidence*, 28 U.S.C., 2020, Rule 702). Although the FRE only apply in federal cases, many states use similar rules. By specifying testimony must be reliable, based on sufficient knowledge, and speak the “truth,” the FRE implicitly embed in this requirement the expectation that experts be neutral (*Saks*, 1990). In psychology, this expectation is explicit in the American Psychological Association’s *Specialty Guidelines for Forensic Psychology* (*American Psychological Association*, 2013). The guidelines state forensic psychologists must be impartial, fair, and independent and “strive to resist partisan pressures to provide services in any ways that might tend to be misleading or inaccurate” (*APA*, 2013, p. 8). This demand for expert neutrality is not limited to the psychology profession; professional societies in other disciplines make similar calls for the avoidance of partisanship (*Saks*, 1990).

Despite these calls for professional neutrality, judges and other legal professionals have long expressed concerns about experts being biased or “hired guns” (e.g., *Edens et al.*, 2012; *Krafka et al.*, 2002; *Stridbeck et al.*, 2016). Likewise, governmental review boards have expressed similar concerns in criticizing the usage of certain sciences in court (*National Research Council, & Committee on Identifying the Needs of the Forensic Science Community*, 2009; *PCAST*, 2016). Given experts’ “tremendous power” in the courtroom to influence judges and jurors (*DeMatteo et al.*, 2019, p. 129), these concerns are justified. Moreover, research supports these concerns, showing being hired by a particular side unintentionally biases experts toward that side (e.g., *Murrie et al.*, 2009; *Otto*, 1989; *Zusman & Simon*, 1983), and experts intentionally may agree only to represent the side that is consistent with their preexisting biases (*Neal*, 2016). Compounding this issue, most experts are unaware of their own biases, incorrectly assuming it is a problem for other experts but not for themselves (*Zapf et al.*, 2018). As a result, attention has shifted toward promoting interventions to prevent or reduce expert bias. These recommendations have included changing how experts perform their evaluations by using debiasing strategies (*Neal & Brodsky*, 2014, 2016), masking experts to the hiring party (*Robertson*, 2010; *Slobogin*, 2014), or

changing the process of expert testimony altogether through techniques like court-appointed experts or concurrent expert testimony (*Butt*, 2017; *Yarnall*, 2009).

Proponents of concurrent expert testimony, colloquially known as “hot tubbing,” have argued this technique reduces expert bias by enhancing neutrality through proximity with the experts on the opposing side (*Edmond*, 2008, 2009; *Garbis*, 2003) and increasing the clarity and quality of the expert testimony (*Downes*, 2004; *McClellan*, 2007). Despite a large number of legal professionals calling for its implementation (e.g., *Kristjanson*, 2012; *Wood*, 2007a, 2018), concurrent expert testimony has been the subject of almost no empirical study.

### Adversarial Allegiance

Adversarial allegiance refers to the unconscious tendency of experts to interpret evidence in line with the side that has hired them (*Murrie et al.*, 2009; *Murrie & Boccaccini*, 2015). Rather than the deliberate bias of “hired guns,” adversarial allegiance has primarily been explained via cognitive biases (*Neal & Grisso*, 2014). By committing to represent a side, tension may arise between an expert’s perceived role as an advocate and their professional and ethical duty of impartiality. This commitment to a course of action incompatible with the professional expectation of neutrality may lead to cognitive dissonance (*Brehm & Cohen*, 1962), further exacerbated by pressure from the hiring attorney and the knowledge that performance in the current case can affect the experts’ ability to be hired in future cases (*Krauss et al.*, 2018; *Murrie et al.*, 2013).

According to dissonance theory, resolving cognitive dissonance requires individuals to change their attitudes, trivialize the importance of their original attitudes, or change their thoughts or behaviors (*Festinger*, 1957). The chosen method often depends on which is least resistant to change (*Brehm & Cohen*, 1962). For an expert, there are few opportunities to reduce dissonance. Certainly, an expert may take back the agreement to testify, but this retraction is likely to have a negative impact on their ability to gain future work. Moreover, experts presumably will not trivialize the importance of neutrality, because most experts recognize its importance (*Zapf et al.*, 2018). Instead, experts may unknowingly reduce dissonance by searching for and interpreting evidence in a manner that will justify their position in the case (*Hart et al.*, 2009; *Neal & Grisso*, 2014). Indeed, such a defense motivation has a greater influence on search strategies than an accuracy motivation (*Hart et al.*, 2009). By engaging in a biased information search, experts can resolve their cognitive dissonance by deciding that they are acting in line with their professional ethics, because they came to the correct conclusion anyway.

Resolving cognitive dissonance in the above manner is consistent with *confirmation bias*, the tendency to search for or interpret information in a manner consistent with prior beliefs and attitudes (Jonas et al., 2001; Lundgren & Prislun, 1998; Wason & Johnson-Laird, 1972). The tendency to engage in confirmation bias can be exacerbated by multiple commitments (Joule & Azdia, 2003). Beyond the initial commitment to represent a particular side, experts actually engage in multiple points of commitment throughout the testimony process. As the experts review information in the case that challenges that side and exacerbates dissonance, experts must engage in more motivated reasoning to reduce that dissonance (Neal & Grisso, 2014). This escalating commitment is further strengthened by the overall testimony process: Experts first come to an opinion that their lawyers accept, then write expert reports explaining that opinion, then testify at court about those opinions. Experts recommit to their opinion at each step. This escalating commitment may drive experts' opinions to become even stronger over time, as research suggests dissonance effects are additive and lead to increased confirmation bias (Jonas et al., 2001; Joule & Azdia, 2003). Thus, the adversarial system encourages experts to become more extreme in their opinions as they progress through a case.

Adversarial allegiance occurs among experts in sexual risk assessment evaluations (e.g., Murrie et al., 2009), civil litigation assessments (Zusman & Simon, 1983), evaluations of child interviews in child sexual abuse cases (McAuliff & Arter, 2016), and insanity evaluations (Otto, 1989). Even supposedly objective evaluations such as actuarial risk assessments—which may rely purely on background information in someone's records—are impacted by adversarial allegiance (Chevalier et al., 2015; Murrie et al., 2008, 2009). With evidence of adversarial allegiance well-established, attention has now shifted to methods for reducing adversarial allegiance and expert bias (e.g., Neal & Brodsky, 2016; Neal & Grisso, 2014).

### Court-Appointed Experts

One alternative to the typical adversarial expert process is the use of court-appointed experts. Federal Rule of Evidence 706 specifically provides for the use of court-appointed experts (Federal Rules of Evidence, 28 U.S.C., 2020), which means they can be easily employed in the United States even if available evidence suggests they are not often used (Cecil & Willging, 1994; Champagne et al., 1996; Sward, 1989).

Proponents for court-appointed testimony argue that being hired by the court rather than a side naturally promotes expert impartiality (Malsch & Freckelton, 2005; Perez, 2016; Worthington et al., 2002), a view shared by legal professionals (Dattilio et al., 2006; Lindsay et al., 1990). Some evidence suggests, however, that court-appointed experts may instead function like prosecution experts (Blais & Forth, 2014), and court-appointed experts show preference for prosecutors over defense attorneys (Grøndahl et al., 2013). Furthermore, using a single court-appointed expert may make it more likely for experts' personal biases to influence their perception of the case evidence (Sonenshein & Fitzpatrick, 2013; Yarnall, 2009) or may inadequately convey the amount of dispute in a particular area (Edmond, 2009; Murrie & Boccaccini, 2015; Yarnall, 2009). Given court-appointed experts are supposedly in a neutral role, however, it is unlikely that they will experience dissonance that would lead to a strengthening of their position over the course of the case. Instead, personal biases would more readily

influence court-appointed experts' views of the case, allowing those views to persist throughout the process.

### Concurrent Expert Testimony

The long-established mechanism of court-appointed testimony is suggested to reform expert testimony with only modest changes to implement. In contrast, concurrent expert testimony represents a relatively novel alternative with substantive changes to the typical adversarial expert process. Concurrent expert testimony, or hot tubbing, is a process that began in Australia and has spread to several other countries (Butt, 2017). In its traditional form, hot tubbing involves experts from opposing sides directly interacting at several stages in the testimony process (Edmond, 2008, 2009). The experts first meet in a pre-trial conference, ideally without attorneys present (Edmond, 2003, 2008, 2009), to identify areas of agreement and outstanding issues to be discussed at trial (Edmond, 2003; McClellan, 2007). After these issues have been laid out and agreed to by the experts, the experts produce a joint report that delineates those areas of agreement as well as the areas of dispute. The experts note their respective positions on any areas of disagreement, and testimony is limited at trial to these areas of disagreement (Edmond, 2008, 2009). By narrowing the scope of testimony to just the areas of disagreement, hot tubbing is purported to improve the ability of fact finders to understand and evaluate trial evidence (Sanders, 2007).

The hot tubbing process continues after the pre-trial conference, and it is set up to continue reinforcing proximity with the opposing experts. Opposing experts sit and testify together on the stand during trial (Downes, 2004; McClellan, 2007). The first phase of testimony functions more freely and less formally than typical direct examination, as experts can simply provide their interpretation of the case and the reasons for it (Edmond, 2008, 2009; Garbis, 2003). The experts may also question each other and comment on the testimony of the other experts while on the stand. The judge may also pose questions during this part of the proceedings or ask the experts to comment on specific topics or issues. The second phase of concurrent testimony functions more like typical cross-examination, with lawyers leading the questioning process during that phase (Edmond, 2008, 2009; Freckelton, 2005).

Proponents have argued that the close proximity created through hot tubbing promotes expert neutrality by increasing experts' independence from their retaining parties and greater reliance on their professional identities (Edmond, 2008, 2009; Garbis, 2003). When an issue is relevant to an individual's ingroup, that individual is more likely to engage in systematic processing of the message content (Crano, 2000; Mackie, 1986; Mackie et al., 1990). In the context of expert testimony, one would therefore expect concurrent experts engaging in hot tubbing to experience less dissonance between their roles as advocates and their ethical responsibility to be nonpartisan. If experts experience less dissonance and their identity as an impartial expert becomes the forefront identity, experts should be less likely to engage in confirmation bias and adversarial allegiance and should instead engage in more thorough processing of the evidence. Surveys of judges, experts, and legal professionals have generally been quite favorable and support this view, suggesting experts are perceived to be less partisan when using concurrent procedures (e.g., AAT, 2005; Civil Justice Council, 2016; Downes, 2004).

On the one hand, this shift toward expert impartiality should mean experts' reports would be less partisan. These reports would seemingly be more comprehensive, including more acknowledgments of other sides, counterarguments, limitations of experts' ultimate opinions. On the other hand, it is unclear how this shift may affect the testimony process. Concurrent expert testimony is specifically set up to restrict testimony at court to the areas of disagreement and the reasons for the disagreement (Kristjanson, 2012). Even if experts' reports are less biased and display less adversarial allegiance, their testimony is nonetheless likely to skew toward justifying one's own position as opposed to acknowledging its weaknesses, which may undermine its usefulness. This view is consistent with some of the stated criticisms of concurrent testimony that have said it may still produce a skewed view of the case on the stand (Davies, 2004; Kristjanson, 2012; Martire & Edmond, 2017).

### Overview of Present Study and Hypotheses

Although there have been repeated calls to incorporate concurrent expert testimony into American trials (e.g., Prescott & Fadgen, 2019; Reifert, 2011; Welch, 2010) and those calls have been successful in limited areas (Devitt, 2012; Emmerig et al., 2013; Wood, 2018); little is known about the actual impact of the concurrent testimony process on experts' decision-making processes. Using the context of a criminal responsibility (insanity) evaluation and examining expert opinions across phases of the evaluation (from recruitment through testimony), the current study was designed to test whether concurrent expert testimony would lead experts to generate less biased opinions than the typical adversarial process. We manipulated whether experts testified using the adversarial, concurrent, or court-appointed process (type). We assigned the non-court-appointed experts to testify for the prosecution or defense (side) based on a measure of their preexisting leaning toward one side or the other. We examined the effects of these variables on experts' responsibility decisions, ratings of the case evidence, content of expert reports and testimony, and level of cognitive dissonance. The City University of New York Human Research Protections Program (HRPP) provided ethical approval for the study.

Hypothesis 1: Given strong existing evidence for adversarial allegiance, we predicted that defense experts would render more *not criminally responsible* judgments and rate the defendant lower in criminal responsibility than would prosecution experts.

Hypothesis 2: We expected that over time, adversarial allegiance effects on experts' responsibility judgments and responsibility ratings would be heightened for adversarial experts and attenuated for concurrent experts. We expected court-appointed experts' ratings would not change over time.

Hypothesis 3: Given adversarial allegiance is likely driven by cognitive dissonance, we predicted that testimony type and phase would influence participants' self-reported level of dissonance, with adversarial and concurrent experts reporting higher levels of dissonance than court-appointed experts. We expected that over time dissonance ratings would increase for adversarial experts, decrease for concurrent experts, and remain unchanged for court-appointed experts.

We also hypothesized testimony type (adversarial or concurrent) would influence the content of experts' reports and testimony. These hypotheses (Hypotheses 4–8) and the results related to them are available in the [online supplemental materials](#).

## Method

### Participants

Participants ( $N = 103$ ) were 42 clinicians ( $M$  age = 46.05 years;  $SD = 11.63$ ; 64.29% female; 92.68% White; 4.76% Hispanic;  $M$  years since doctorate = 16.17 years;  $SD = 11.63$ ) and 61 clinical psychology doctoral students ( $M$  age = 29.45 years;  $SD = 5.30$ ; 85.25% female; 78.69% White; 19.67% Hispanic;  $M$  year-long externships completed = 2.48;  $SD = 1.35$ ) recruited from New York, Pennsylvania, and Texas (see Table 1 for additional demographic information and professional qualifications for both samples). We required the doctoral students to have at least 1 year of doctoral-level clinical practicum experience and have completed at least one report on a full clinical evaluation. We compensated participants for their participation in the study.

Eighty-seven participants completed the first set of measures in the study during Phase 2, and 51 participants completed the second set of measures during Phase 3. Forty-eight participants completed all four phases of the study and were included in the analyses exploring attitude change across the phases of the study. For an overview of the study flow and participant exclusion across the study phases, see Figure 1.

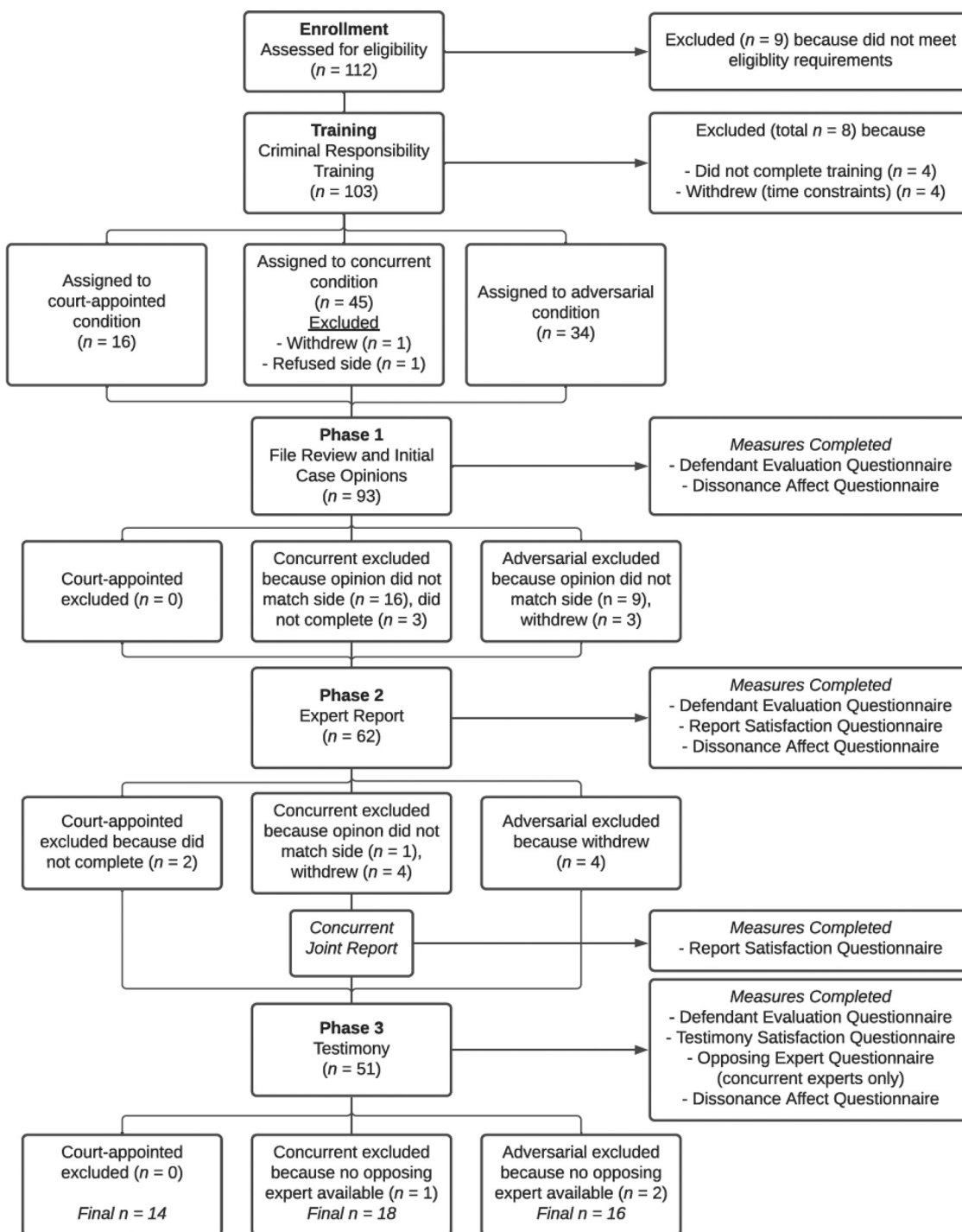
There was no significant difference in the proportion of participants excluded for their opinion not matching their hiring side between the adversarial (29.03%) and concurrent (42.50%) conditions,  $\chi^2(1) = .85, p = .36, V = .13, 95\% \text{ CI } [.00, .37]$ .

**Table 1**  
*Demographic Information by Participant Type*

	Clinicians ( $n = 42$ )	$n$	%
Educational background			
PhD in psychology		29	69.05
PsyD in psychology		12	28.57
MD		1	2.38
Licensed clinician		35	83.33
Forensic training experience			
Postdoctoral fellowship in forensic psychology		9	21.43
Continuing education in forensic evaluation		29	69.05
Forensic concentration in doctoral program		12	28.57
Psychology practica in forensic setting		14	33.33
Employment setting			
State hospital		7	16.67
Private practice		22	52.38
Prison		13	30.95
Faculty		11	26.19
Testimony experience			
Testify for prosecution		10	23.81
Testify for defense		18	42.86
	Students ( $n = 61$ )		
Number of practica completed		$n$	$M$
At least 8 months long		61	2.48
Fewer than 8 months		61	0.41
With adult populations		61	2.41
Number completed		$n$	$Mdn$
Clinical reports		58	13.00
Forensic evaluations		61	3.00
			$IQR$
Clinical reports			18.25
Forensic evaluations			10.00

*Note.* Clinician employment setting does not equal to 100%, because some participants were employed in multiple settings.

**Figure 1**  
Survey Flow Showing Participant Completion and Materials by Study Phase



This document is copyrighted by the American Psychological Association or one of its allied publishers. This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

**Design**

The study had a 2 (side: prosecution vs. defense) × 2 (testimony type: adversarial vs. concurrent) + 1 (testimony type control: court-appointed) factorial design. We assigned participants to side based

on their responses to the forensic bias questionnaire. We randomly assigned all participants to a testimony type condition.

Our power analysis indicated a sample size of 48 would give power of .80 to detect a medium effect ( $f = .20$ ) for the test of the 2 × 3 design across the three points of measurement if the

dependent measures had correlations of .70 (calculated using G\*Power, Faul et al., 2007).

## Procedure

### Training

We recruited participants to participate in a study of clinical decision making in the legal system focusing on criminal responsibility; we told them that they would be evaluating a mock case rather than conducting a real evaluation. We pulled names of potential doctoral-level clinicians in the greater New York area through professional association directories, state expert lists, and Internet searches. We then emailed these clinicians with invitations to participate. We recruited doctoral students from doctoral programs in clinical psychology in the greater New York City area, Pennsylvania, and Texas. The advertisements directed participants to a website where participants provided informed consent and completed the clinical background questionnaire and forensic bias questionnaire. Participants also provided contact information so they could be contacted to participate if eligible. Of the 112 participants who fully completed the screener questionnaire, nine were excluded from participation because they did not meet eligibility requirements.

Participants next completed an APA-approved self-paced online continuing education course on criminal responsibility. We provided all participants with a copy of *Evaluation of Criminal Responsibility* (Packer, 2009) as study material for the course and a reference to use while completing the study. We required participants provide their completion certificate for the course before beginning Phase 1 of the study.

### Phase 1

Once we categorized participants as prosecution or prodefense (based on their responses to the forensic issues and background questionnaires) and they completed the training, we randomly assigned them to the concurrent, adversarial, or court-appointed expert condition. We contacted participants via email to inform them that expert profiles had been generated based on their responses to the questions in the clinical background questionnaire and these files had been given to the attorneys or judge (only for court-appointed experts) in the case. We asked participants to provide ideal times for contact via telephone should they be selected as experts in the case.

Two practicing attorneys working on the project contacted the adversarial and concurrent experts assigned to their respective sides. Using a script, the attorneys described the case and requested the participants' assistance in the case as an expert witness. If the participants agreed to participate, we sent them an instruction sheet and case materials via e-mail to review for the case from the attorney for their side. A research assistant posing as an assistant to the mock judge in the case contacted participants in the court-appointed condition to request their assistance as court-appointed experts. These participants received the instruction sheet and case materials from the judge's assistant on behalf of the judge.

Prior to reviewing materials, we told participants in the adversarial and concurrent conditions that the attorneys had the right to terminate experts at any time, so their opinions would determine whether the attorneys asked the experts to write an expert report or testify in the case. We also informed participants they would be

participating in a novel concurrent trial process at this time if assigned to that condition. Participants provided informed consent online and then reviewed the case materials and videotaped interviews with the defendant in the order specified in the instruction sheets. Participants then completed the defendant evaluation questionnaire and the dissonance affect questionnaire.

### Phase 2

We then invited all court-appointed experts and those adversarial and concurrent experts whose opinions matched the side for which they had been hired to complete an expert report by their respective attorney or the judicial assistant. After participants provided informed consent online, we gave participants a 75-min timeframe to complete the report. Given time constraints, we asked them to focus on the clinical formulation and diagnosis, the defendant's mental state at the time of the alleged offense, and their overall impression regarding the defendant's criminal responsibility at the time of the alleged offense. We also asked participants to include psychosocial history and current mental status, time permitting. We told participants they could provide the latter portions of the report in bullet points to save time if necessary. We asked participants to write their report and email a copy to their respective attorney or the judicial assistant (depending on condition). We also reminded participants in the adversarial and concurrent conditions that the attorneys' satisfaction with their reports would determine whether they would be invited to provide testimony. After completing the report, participants completed the defendant evaluation questionnaire, the report satisfaction questionnaire, and the dissonance affect questionnaire.

### Phase 3

We then invited via email all court-appointed experts and those adversarial and concurrent participants whose opinions matched the side for which they had been hired to come in to provide testimony. We scheduled court-appointed experts to testify individually. We randomly paired adversarial and concurrent experts with an expert in the same testimony condition but from the opposing side. After testimony was scheduled, we sent the adversarial and concurrent participants the report of their opposing experts via e-mail for review before the testimony phase.

We scheduled participants in the concurrent condition in pairs. Only one pair in the study involved a clinician-student pairing. Upon arrival to the research laboratory, we gave concurrent expert pairs copies of their individual reports and asked them to generate a joint report for the court outlining areas of agreement and disagreement. We gave the pairs a 30-min timeframe, and we video recorded the sessions. Participants completed the report satisfaction questionnaire upon completion of the joint report and before providing testimony.

We video recorded all testimony, which the experts provided in mock courtrooms. Participants in the concurrent condition testified together on the stand. Participants sat in the witness box together, and the judge (played by a doctoral student) provided an overview of the concurrent testimony process. The judge provided the topics of discussion based on the pairs' joint reports to the court and asked the pairs to discuss their areas of disagreement about the case. The judge and attorneys followed flexible scripts to ask additional questions of the experts.

Court-appointed and adversarial participants testified individually on the stand. For the court-appointed participants, the attorney whose case benefited the most from their opinion conducted the direct examination. For the adversarial participants, the attorney who hired them conducted the direct examination. The opposing attorney then conducted cross-examination. The attorneys followed flexible scripts during the questioning phase.

After providing testimony, participants completed the defendant evaluation questionnaire, the testimony satisfaction questionnaire, and the dissonance affect questionnaire. Concurrent experts also completed the opposing expert questionnaire. After completing the questionnaires, we thanked, debriefed, and compensated all participants for their participation.

## Materials

All materials are available on OSF: <https://osf.io/u23tc/>.

### *Clinical Background Questionnaire*

Participants completed the clinical background questionnaire online to apply for the study. Clinicians reported their educational experience, forensic training experience, current employment status, and their forensic evaluation and expert testimony experience. Students reported their current and former educational experience, doctoral-level practicum experience, forensic training experience, and clinical assessment and report writing experience. We asked both samples to indicate if they would have any reservations serving as a defense or prosecution expert witness, and they provided demographic information, including gender, age, and racial/ethnic background.

### *Forensic Issues Questionnaire*

Participants rated their agreement with 12 items measuring attitudes toward forensic issues (e.g., "Defendants are unlikely to malingering mental illness to avoid prosecution.") using a 7-point Likert-type scale (1 = *strongly agree*; 7 = *strongly disagree*). A total score was created by summing the items, four of which were reverse coded. Higher scores indicated more proprosecution leanings, whereas lower scores indicated more prodefense leanings. We used median splits to assign participants to the prosecution or defense. We assigned participants to the side they leaned toward to capture the predispositions that affects experts' willingness to represent a particular side and may be an important component leading to adversarial allegiance (Murrie & Boccaccini, 2015; Neal, 2016) and increase the realism of the study. We calculated the median split on the first 20 participants (who were assigned to a side after the median split was performed) and then updated as additional participants enrolled in the study. Participants on the median (final median = 38.00) were randomly assigned to side. The final internal consistency of the scale was weak ( $\alpha = .26$ ). Given the weakness of the scale, the court-appointed experts were treated as a single group instead of separately analyzed as proprosecution or prodefense-leaning, so we did not use the results of this questionnaire in analyses.

### *Case Materials*

Participants received a packet of case materials to conduct a criminal responsibility evaluation. Although all materials were

fictitious, they were based on an actual case that involved a criminal responsibility evaluation, and we consulted with forensic psychologists throughout their development to maximize experimental realism. Table 2 provides an overview of the evaluation materials. We video recorded the clinical interviews of the defendant prior to the study. The videos showed the defendant (an advanced clinical doctoral student with acting experience) completing the clinical interviews in a testing room. We used multiple cameras to provide multiple perspectives of the defendant. An advanced doctoral student with forensic evaluation experience conducted the interviews off camera, and we instructed participants to view the interviews as though they had conducted them. We created two interviews: one included a comprehensive psychosocial history and mental status examination, and the other focused on discussion of the events and defendant's experiences around the time of the alleged offense.

We pilot tested case materials to ensure the case was ambiguous and did not inherently favor or oppose criminal responsibility using a focus group of eight advanced clinical psychology doctoral students with forensic experience. All students in the focus group had completed a doctoral course in forensic assessment, had at least 2 years external practicum experience, and had at least 1 year of practicum experience in a forensic setting. Participants were evenly split in their evaluation of the defendant after review of the case materials. We also reviewed the materials with a forensic psychologist before implementation.

### *Defendant Evaluation Questionnaire*

Participants completed a defendant evaluation questionnaire after Phases 1 through 3 of the study. Participants rated their strength of agreement with four statements pertaining to the defendant's meeting of criteria for criminal responsibility using a 10-point Likert-type scale (1 = *strongly disagree*; 10 = *strongly agree*), with lower scores indicating greater responsibility. We also asked the participants to indicate if they thought the defendant was criminally responsible (criminally responsible/not criminally responsible).

**Table 2**  
*Summary of Case Materials Reviewed by Participants for the Evaluation*

Content	Format
Summary records	
Criminal indictment	Document
Court order for evaluation (court-appointed experts only)	Document
Federal statute for criminal responsibility	Document
Collateral records	
Police incident report (current offense)	Document
Defendant's criminal record	Document
Defendant's psychiatric records	Document
Phone interview with defendant's mother	Summary notes
Observations from institutional staff	Summary notes
Evaluation materials	
Clinical interview with defendant	
Session 1: Psychosocial history	Video
Session 2: Discussion of alleged offense	Video
MMPI-2 testing results	Testing materials and output

Participants then rated their confidence in their criminal responsibility decision using a 10-point Likert-type scale (1 = *not at all confident*; 10 = *very confident*). Finally, participants rated the degree to which they felt the defendant was criminally responsible (1 = not at all criminally responsible; 10 = very criminally responsible) and how strongly the evidence pointed toward criminal responsibility (1 = *strongly in favor of responsibility*; 10 = *strongly in favor of no responsibility*). We reverse-scored the question about the degree to which the defendant was criminally responsible to ensure that lower scores indicated greater responsibility (less insanity) across all items.

We combined five items (unaware of actions, did not understand actions were wrong, could not conform behavior to the law, degree of responsibility, and evidence strength) into a composite responsibility variable. The composite variable showed strong internal consistency across all three time points (McDonald's  $\omega_s = .86-.93$ ).

### Report Satisfaction Questionnaire

After completing the expert reports, we asked participants to complete a report satisfaction questionnaire. Participants rated the strength of their agreement with seven items measuring their satisfaction with the expert report (e.g., "I am very satisfied with the final report"), how closely the report reflects their case opinion (e.g., "The final report closely reflects my opinions in the case"), and the objectivity of the report (e.g., "I evaluated the evidence objectively during the generation of the final report") using a 10-point Likert-type scale (1 = *strongly disagree*; 10 = *strongly agree*). When rating their joint reports, concurrent experts completed an additional four items evaluating the process of working with another expert to create the report (e.g., "I worked well with the other expert while generating the report") using the same Likert-type scale.

### Testimony Satisfaction Questionnaire

We asked all experts to complete a testimony satisfaction questionnaire following their testimony in Phase 3. Participants rated the strength of their agreement with 24 statements regarding the testimony process. Participants rated their satisfaction with the process (e.g., "I am very satisfied with my expert testimony"), their objectivity (e.g., "My testimony favored the side that hired me"), the value of their testimony (e.g., "My opinions were valued"), their effectiveness (e.g., "My testimony was effective"), their convincingness (e.g., "A jury would favor the side that hired me"), their perceived control over their testimony (e.g., "I did not have much control over the content of my testimony"), and their enjoyment with the testimony process (e.g., "I enjoyed giving expert testimony") using 10-point Likert-type scales (1 = *strongly disagree*; 10 = *strongly agree*). Participants also rated their familiarity with the testimony procedures on a 10-point Likert-type scale, with higher scores indicating greater familiarity with the procedures.

### Opposing Expert Questionnaire

After completing the testimony satisfaction questionnaire, experts in the concurrent condition rated the expert they opposed in the case. We did not ask the experts in the adversarial condition to rate the opposing experts, because they did not witness the opposing expert's testimony. Participants rated their strength of

agreement with 15 statements about the expert's qualifications (e.g., "The opposing expert had specialized knowledge relevant to this case"), level of bias (e.g., "The opposing expert was advocated for the side that hired him/her"), effectiveness (e.g., "The opposing expert would persuade a jury"), respectfulness (e.g., "The opposing expert was combative"), and testimony clarity (e.g., "The opposing expert's testimony was clear") using 10-point Likert-type scales (1 = *strongly disagree*; 10 = *strongly agree*).

### Dissonance Affect Questionnaire

Participants completed the dissonance measure at the end of each phase after completing the defendant evaluation questionnaire and satisfaction measures. The measure consists of 24 items that represent different emotions, three related to the feeling of cognitive dissonance and additional filler items (Elliot & Devine, 1994). Participants rate how much they are feeling each emotion at the present moment using a 7-point Likert-type scale (1 = *does not apply at all*; 7 = *applies very much*). We created the composite measure of discomfort using the same subset of three items (uncomfortable, uneasy, bothered; McDonald's  $\omega_s = .87-.92$ ) used in previous research.

### Manipulation Check Questions

Participants completed a set of five manipulation checks questions after the testimony phase of the study. Participants indicated the side for which they wrote an expert report (prosecution, defense, or court-appointed), whether they wrote a report jointly with an opposing expert (yes/no), for which side they testified (prosecution, defense, or court-appointed), what testimony procedure was used (adversarial or concurrent), and whether the participant testified together on the stand with another expert or separately (together, separate, or sole court-appointed witness). Overall, participants were mostly accurate in responding to the manipulation check questions; court-appointed experts tended to mistake their testimony as being for a particular side despite acknowledging they had written their reports as court-appointed experts, but this may be because the attorney who was most favored by their case evaluation conducted their direct examination. There were no other meaningful patterns.

## Results

### Criminal Responsibility Opinions

We conducted a hierarchical logistic regression predicting *not criminally responsible* judgments with side and testimony type added as predictors in the first step and the interaction between the factors in the second (see Table 3). For all Phase 1 analyses, we ran analyses on the full data, including those later excluded from the study. The first model was significant,  $\chi^2(2) = 7.47, p = .02$ , Nagelkerke  $R^2 = .14$ . Side was a significant predictor in the model, with the odds of defense experts finding the defendant not criminally responsible being approximately four times higher compared with the odds of the prosecution experts finding the defendant not criminally responsible (see Table 4 for proportions). Adding the interaction between side and testimony type did not significantly improve the fit of the model,  $\chi^2(1) = 1.14, p = .29$ , but the final model remained significant,  $\chi^2(3) = 8.60, p = .04$ , Nagelkerke  $R^2 =$

**Table 3**  
*Hierarchical Logistic Regression Predicting Participants' Initial Criminal Responsibility Judgments by Expert Side and Testimony Type*

Predictors	B	SE	Wald	p	Exp(B)	95% CI
<b>Model 1</b>						
Constant	-0.01	0.44	0.00	.988	0.99	
Side	<b>1.40</b>	<b>0.54</b>	<b>6.84</b>	<b>.009</b>	<b>4.07</b>	<b>[1.42, 11.66]</b>
Testimony type	-0.08	0.53	0.03	.873	0.92	[0.33, 2.57]
<b>Model 2</b>						
Constant	-0.25	0.50	0.25	.618	0.78	
Side	<b>2.12</b>	<b>0.91</b>	<b>5.43</b>	<b>.020</b>	<b>8.36</b>	<b>[1.40, 49.88]</b>
Testimony type	0.35	0.67	0.27	.603	1.41	[0.38, 5.23]
Side × Testimony Type	-1.19	1.14	1.09	.296	0.31	[0.03, 2.83]

Note. Regression is predicting judgments of not responsible (insane). The reference category for side is prosecution experts (vs. defense); the reference category for testimony type is adversarial experts (vs. concurrent). Significant predictors are in bold.

.16. Again, side remained the only significant predictor in the model. As predicted, there was no impact of testimony type in participants' initial evaluations. These results support Hypothesis 1.

We conducted a generalized logit model with generalized estimating equations using an unstructured correlation matrix to account for the repeated nature of the data with a binary outcome; however, the model did not achieve convergence. As can be seen in Table 4, however, the proportions of not responsible judgments only strengthened in allegiance effects over time for both adversarial and concurrent experts. These results failed to support Hypothesis 2.

**Strength of Criminal Responsibility Opinions**

Consistent with Hypothesis 1, a 2 (side: prosecution vs. defense) × 2 (testimony type: adversarial vs. concurrent) ANOVA showed no significant differences between adversarial and concurrent experts on their initial composite criminal responsibility opinions,  $F(1, 67) = .43, p = .52, d = .16, 95\% \text{ CI} [-.31, .63]$ , and no significant side by testimony type interaction,  $F(1, 67) = .11, p = .74, \eta_p^2 = .002$ . There was a significant main effect of side on participants' initial criminal responsibility opinions,  $F(1, 67) = 9.52, p = .003, d = .76, 95\% \text{ CI} [.28, 1.24]$ . As hypothesized, defense experts more strongly agreed that the defendant was not criminally responsible (i.e., was insane;  $M = 6.76, SD = 1.65$ ) than did prosecution experts ( $M = 5.42, SD = 1.87$ ), regardless of whether they were adversarial or concurrent experts.

Participants again rated their perceptions of criminal responsibility after completing their expert reports and after providing testimony. To evaluate whether adversarial experts' opinions diverged and concurrent experts' opinions converged over time, a 2 (side) × 2 (expert type) × 3 (evaluation phase) mixed ANOVA

was conducted on the participants who completed evaluations across all three phases of the study (see Table 5).

In contrast with Hypothesis 2, there was no significant main effect of evaluation time,  $F(2, 60) = .96, p = .39, \eta_p^2 = .03$ , and no significant interactions including evaluation time ( $ps \geq .72$ ), suggesting participants' opinions did not significantly change over the study phases. There was no significant difference between adversarial and concurrent experts,  $F(1, 30) = .05, p = .83, d = .04, 95\% \text{ CI} [-.64, .71]$ , nor a significant side by testimony type interaction,  $F(1, 30) = .82, p = .37, \eta_p^2 = .03$ ; however, there was a significant main effect for side,  $F(1, 30) = 118.77, p < .001, d = 1.87, 95\% \text{ CI} [1.06, 2.67]$ . Just as in Phase 1 alone, defense participants ( $M = 7.53, SE = .24$ ) rated the case more strongly in favor of no criminal responsibility (or insanity) across all phases of the study as compared to prosecution experts ( $M = 3.83, SE = .24$ ).

We analyzed the data using a repeated-measures one-way ANOVA. Mauchly's test indicated that the assumption of sphericity had been violated,  $\chi^2(2) = 8.79, p = .01$ , so the Greenhouse-Geisser correction was used ( $\epsilon = .66$ ). Consistent with Hypothesis 2, the repeated-measures ANOVA revealed no significant main effect of time on court-appointed experts' opinions,  $F(1.32, 17.11) = .12, p = .80, \eta_p^2 = .01$ . Across the three phases of the study, participants consistently rated the defendant's responsibility as relatively moderate (see Table 5).

**Cognitive Dissonance During Evaluation**

As we did not hypothesize that side would play a role in the experience of cognitive dissonance, we did not include this variable as a predictor in the analyses. In contrast to Hypothesis 3, a one-way ANOVA revealed no significant difference in dissonance ratings among court-appointed, adversarial, and concurrent experts,

**Table 4**  
*Percentage of Not Responsible Judgments (Insanity) by Condition and Phase*

Expert type	Side	Phase 1		Phase 2		Phase 3	
		% Not responsible	N	% Not responsible	N	% Not responsible	N
Adversarial	Prosecution	43.75	16	0.00	8	0.00	8
	Defense	86.67	15	100.00	10	75.00	8
Concurrent	Prosecution	52.38	21	0.00	9	0.00	9
	Defense	73.68	19	90.91	11	100.00	9
Court-appointed		50.00	16	61.54	13	46.15	13

This document is copyrighted by the American Psychological Association or one of its allied publishers. This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

**Table 5**  
*Participants' Criminal Responsibility Ratings by Expert Side, Testimony Type, and Phase*

Testimony type	Side	Phase 1 <i>M (SD)</i>	Phase 2 <i>M (SD)</i>	Phase 3 <i>M (SD)</i>
Adversarial	Prosecution ( <i>n</i> = 8)	4.18 (1.02)	3.90 (0.66)	3.98 (0.63)
	Defense ( <i>n</i> = 8)	7.38 (1.27)	7.28 (1.71)	7.58 (1.46)
Concurrent	Prosecution ( <i>n</i> = 9)	3.82 (0.90)	3.44 (0.90)	3.64 (0.63)
	Defense ( <i>n</i> = 9)	7.80 (1.09)	7.60 (1.42)	7.53 (1.43)
Court-appointed ( <i>n</i> = 14)		5.83 (1.47)	5.97 (2.23)	5.77 (2.06)

*Note.* Higher scores indicate greater lack of responsibility (more insanity) on a 10-point Likert-type scale. Only participants who completed all three phases of the research study are included in the table.

$F(2, 84) = .02, p = .98, \eta_p^2 < .01$ . Instead, experts across all testimony types reported a moderate level of dissonance overall.

There was a significant time by testimony type interaction,  $F(4, 90) = 2.91, p = .03, \eta_p^2 = .12$ ; however, the pattern of results did not follow the direction specified in Hypothesis 3 (see Table 6). Bonferroni post hoc tests showed *court-appointed experts* reported significantly lower dissonance when submitting the expert report (Phase 2) than when they had previously evaluated the case materials (Phase 1),  $M_{diff} = -1.38, SE = .41, p = .004, d = -.67, 95\% CI [-1.43, .09]$ , but there was no significant change in dissonance from submitting the report to providing testimony (Phase 3),  $M_{diff} = -.55, SE = .43, p = .63, d = -.29, 95\% CI [-1.03, .46]$ . In contrast, adversarial experts reported no significant changes in discomfort from Phase 1 to Phase 2,  $M_{diff} = -.48, SE = .38, p = .65, d = -.38, 95\% CI [-1.08, .32]$ , or from Phase 2 to Phase 3,  $M_{diff} = .48, SE = .40, p = .72, d = .57, 95\% CI [-.38, 1.02]$ . Likewise, concurrent experts reported no significant changes in discomfort from Phase 1 to Phase 2,  $M_{diff} = -.02, SE = .36, p = 1.00, d = -.01, 95\% CI [-.67, .64]$ , or from Phase 2 to Phase 3,  $M_{diff} = .48, SE = .38, p = .63, d = .48, 95\% CI [-.19, 1.14]$ .

## Discussion

The purpose of the current investigation was to examine whether the concurrent expert testimony process would ameliorate adversarial allegiance and promote expert neutrality as compared with the typical adversarial process. Overall, the results suggest concurrent testimony did not produce benefits at any phase of the expert decision-making process. Our findings did support Hypothesis 1, as experts' initial case opinions and responsibility judgments were skewed toward the side that hired them, consistent with adversarial allegiance. In contrast with Hypothesis 2, however, the effects of adversarial allegiance remained apparent and stable for experts assigned to concurrent expert testimony even after these concurrent experts were exposed to the opposing side. Participants in the prosecution conditions consistently viewed the defendant as more responsible than did those in the defense conditions, regardless of whether they were in the traditional

adversarial system or in the hot tub. Likewise, in contrast to Hypothesis 3, reports of cognitive dissonance did not differ between adversarial and concurrent experts. As discussed in the [online supplemental materials](#), we also did not see benefits of concurrent expert testimony in the experts' reports or testimony.

The current study found no evidence to support the conclusion that concurrent expert testimony reduces adversarial allegiance. In this case simulation, experts interpreted case information and presented their opinions similarly—whether in the traditional adversarial process or the hot tub—in a way that favored the side that originally hired them. These results raise questions about the effectiveness of concurrent expert testimony, or hot tubbing, that future research should address.

Our results suggest court-appointed experts achieved more balance than experts retained by a specific side; their mean case ratings and perceptions often lay directly between the prosecution and defense witnesses' ratings. These results must be interpreted with caution, however; in the court-appointed condition, participants reporting both prosecution attitudes and prodefense attitudes were included. Given the poor reliability of the scale used to assign individuals as prosecution or prodefense leaning, analyses on whether the experts expressed opinions in line with their original predispositions was not possible; however, it should be noted that the variance in ratings for the court-appointed condition tended to be greater than the variance in the adversarial and concurrent conditions. This higher variance suggests these middle ground ratings may be due to the mixing of attitudes in this condition as opposed to any attenuation of bias. It appears court-appointed expert testimony may simply allow an expert to operate based on their preexisting biases rather than working toward neutrality (Sonenshein & Fitzpatrick, 2013; Yarnall, 2009).

## Why the Hot Tub Might Not Reduce Adversarial Allegiance

If concurrent testimony does not reduce adversarial allegiance as the current study suggests, one potential reason may involve the timing of when concurrent procedures are implemented in the

**Table 6**  
*Participants' Dissonance Ratings by Testimony Type and Evaluation Phase*

Testimony type	Phase 1 <i>M (SD)</i>	Phase 2 <i>M (SD)</i>	Phase 3 <i>M (SD)</i>
Adversarial ( <i>n</i> = 16)	3.29 (1.94)	3.77 (1.93)	3.29 (1.63)
Concurrent ( <i>n</i> = 18)	3.48 (1.80)	3.50 (1.83)	3.02 (1.61)
Court-appointed ( <i>n</i> = 14)	3.69 (1.67)	2.31 (1.56)	2.86 (1.36)

*Note.* Higher scores indicate greater dissonance on a 7-point composite measure. Only participants who completed all three phases of the research study are included in the table.

expert testimony process. Although we told our concurrent expert witnesses about the concurrent testimony procedures and warned they would be directly interacting with an expert witness on the stand before reviewing the case materials, they still examined the relevant case evidence, formulated their opinions, and wrote expert reports before ever directly interacting with the opposing expert (which mimics actual concurrent testimony procedures; Downes, 2004). Not actually interacting with an opposing expert until late in the process may have attenuated any potential benefits. Moreover, concurrent expert witnesses interacted directly with their own attorneys prior to case review and as they moved through the expert testimony process. Research on cognitive commitment suggests that initial commitment to a particular side can be particularly powerful in activating cognitive dissonance if one chooses an action incompatible with a need or desire (Brehm & Cohen, 1962).

In the case of expert testimony, the initial agreement to serve a particular side could lead to cognitive dissonance between the competing motivations to be an impartial expert and an advocate for a particular side in the case. As experts then go on to produce products (i.e., opinions, reports, testimony) to support a particular side, they may experience escalating commitment, which increases cognitive dissonance even further (Joule & Azdia, 2003). By having experts collaborate on a joint report and testify together, we had expected concurrent expert testimony would shift experts to identify with the opposing expert rather than an adversarial side, as their expert ingroup should be a more important frame of reference. Although concurrent experts generally rated their opposing experts favorably (e.g., high levels of agreement that the opposing expert was knowledgeable, respectful, helpful, and relevant), they also perceived opposing experts as advocates for their side and disagreed with those opposing experts' opinions.

Although proponents of hot tubbing have said it may fail to reduce bias (e.g., Kristjanson, 2012), the lack of bias reduction has typically been acknowledged in the context of experts failing to respect one another (i.e., social identification failures). Our results suggest that even when the experts respect one another, it may not be sufficient to shift commitment to the impartial expert role. Part of this problem may be due to the *bias blind spot* (Pronin et al., 2002; Zapf et al., 2018), or the tendency to perceive bias on the part of others but fail to acknowledge it in oneself. If experts indeed only see other experts as the biased ones, it may preclude social identification with a neutral expert identity.

Premature commitment to a side by agreeing to serve as an expert—or providing an initial case opinion before reviewing all of the evidence—may exert stronger influences that persist despite any potential social identity concerns brought up by expert testimony (Neal & Grisso, 2014). When individuals are encouraged to accept initial information as fact without critical consideration, individuals may engage in mindless processing of later information that fits that bias (Chanowitz & Langer, 1981). Moreover, experts are likely to engage in confirmation bias to maintain that initial belief (Neal & Grisso, 2014). How attorneys initially frame the case to prospective witnesses could therefore have an immense impact on experts if they do not examine that framing critically.

## Limitations

There were limitations to the study worth noting. First, a number of experts were excluded from the study after Phase 1 because

they did not provide an evaluation that benefited the side which was hiring them, which could artificially inflate the appearance of expert bias. We believe this to be unlikely, however, given the consistent pattern of findings in each phase of the study, including Phase 1. Most of the participants excluded for forming opinions that did not support their side were excluded immediately after Phase 1. Given the adversarial allegiance effect was just as strong in Phase 1 with those individuals included as in the latter phases of the study, it is unlikely that the continued adversarial allegiance effects were due to the removal of inconsistent experts. Moreover, this pressure to conform with one's side is a typical part of the expert process (Murrie & Boccaccini, 2015), and we felt it was important to model this aspect of the process, as it may partially contribute to adversarial allegiance.

A small sample completed the full version of the study, so caution should be taken when interpreting the findings. Nonetheless, the repeated measures nature of the design increased our power, and the findings provided a robust replication of the adversarial allegiance effects found in prior research (e.g., Murrie et al., 2008, 2009, 2013). Given we expected to find null results in the concurrent testimony condition but instead replicated the adversarial allegiance effect, it appears sample size may not be a major issue in this particular study or that any beneficial effects of concurrent procedures may be impractically small. However, our effect sizes may be inflated due to the small sample size; future research should replicate these results with a larger sample size. Similarly, the sample composition limits the conclusions from the study. Although we attempted to recruit as many practicing clinicians as possible for this study, the involved nature of the study (including providing in-person testimony) meant many clinicians were unable to participate and the sample primarily included student participants. Many of the students were recruited from clinical psychology doctoral programs with a forensic emphasis, but students still may not fully represent how clinicians would have behaved professionally in actual expert testimony situations.

The format of the study also limits the conclusions that can be drawn. Although the study was designed to be as realistic as possible, including referral and questioning by actual attorneys and materials developed from an actual case, participants were still aware they were participating in a mock criminal responsibility evaluation for a research study. Knowing they were participating in a study may have reduced pressure to act in an ethical, unbiased manner. Likewise, we limited experts to the materials we provided, and we implemented time limits on the writing of expert reports and provision of testimony that do not exist in actual case evaluations. It is possible having more time to review the case materials and carefully consider expert opinions could have led to a reduction in bias. Future researchers may wish to examine whether more naturalistic evaluations with more flexible time frames allow experts to be more cautious and considered in their approach in ways that attenuate bias.

## Directions for Future Research

Several future research directions might offer insight into ways to reduce adversarial allegiance, with or without concurrent expert testimony. A shift in some Australian courts toward a joint conference approach seems to acknowledge the importance of timing when the concurrent expert experience happens. The joint

conference approach has opposing experts meet together after their initial briefings by their respective attorneys but before looking at the evidence or writing expert reports (Rackemann, 2011). Similar to preregistration for empirical research (a practice gaining support for reducing certain forms of experimenter bias; Nosek et al., 2018), the experts must work together to outline the proper methodology for evaluating the evidence and share the resulting data with one another to form their case opinions. Although both joint conferences and hot tubbing result in a joint report for the court outlining areas of agreement and disagreement, this approach further differs from hot tubbing by completely removing the lawyers from the process and allowing the joint conference to be completely confidential (Wood, 2007b). The experts are also not expected to testify concurrently on the stand. An evaluation of cases using the joint conference approach suggests this earlier contact outside the influence of attorneys may be important for reducing adversarial allegiance, as 48% of the cases were resolved by complete agreement between experts during this joint report process (Sutherland, 2011). Likewise, in our own study, most of the experts who reported attitudes inconsistent with their hiring party did so at the very beginning of the case before anything was put into writing, which supports the importance of intervening earlier in the testimony process. Future research should investigate whether the impact of the initial framing persists when experts are confronted with an opposing expert much earlier in the testimony process.

One major recommended reform meriting empirical investigation has been masked referral of expert testimony (Robertson, 2010). Under this process, experts would not know which side has hired them, and the initial case information is provided via intermediaries. Of course, the initial framing of that information can have a great deal of impact on experts, because the assumption is that masked referrals should include a neutral presentation. This expectation of neutrality increases the likelihood that experts would accept it uncritically. If the case information is carefully balanced, it may promote the desired critical processing, but a biased frame may lead to confirmation bias that still creates adversarial allegiance.

Future research should also consider how the concurrent process influences experts as they gain experience with it. One possibility is that concurrent expert testimony will reduce adversarial allegiance as experts become more familiar with it. Identification with opposing experts may be more likely to happen as experts gain familiarity with the concurrent process, because they should be more likely to identify with the opposing experts from the outset. If experts who have experienced concurrent testimony previously are asked to engage in that process again, they may be more likely to examine that initial evidence more skeptically, even when presented in the traditional manner from one's hiring attorney. Although there is no current research on whether concurrent experts have come to greater agreement over time, research on the joint conference, or case management, approach found that experts from opposing parties reaching full agreement in a case increased from 39% to 66% over 3 years (Sutherland, 2011), suggesting that greater familiarity might lead to greater allegiance reduction. Future research should explore whether concurrent expert testimony is more effective with individuals who have previous experience in this method as compared to experts new to testifying under concurrent procedures.

## Conclusions

Altogether, our preliminary findings seem to echo the concerns of Gary Edmond (2009, p. 186): "Concurrent evidence is not a panacea for partisanship, adversarial bias, or the difficulties created by expert disagreement and decision making in the face of uncertainty." Although judges and attorneys have responded favorably to hot tubbing and perceive results from it as less biased (Civil Justice Council, 2016; Downes, 2004), there is no empirical evidence that supports its effectiveness. This initial study indeed suggests concurrent expert opinions may not be significantly different from those produced under typical adversarial expert processes. Despite continued calls to integrate concurrent expert testimony more completely into the American justice system (e.g., Wood, 2007a, 2018), it is premature to promote concurrent expert testimony as a remedy for adversarial allegiance. Rather than jumping in the hot tub, more research should test the waters by examining the efficacy of concurrent expert testimony.

## References

- Administrative Appeals Tribunal. (2005, November). *An evaluation of the use of concurrent evidence in the Administrative Appeals Tribunal*. Commonwealth of Australia.
- American Psychological Association. (2013). Specialty guidelines for forensic psychology. *American Psychologist*, 68(1), 7–19. <https://doi.org/10.1037/a0029889>
- Blais, J., & Forth, A. E. (2014). Prosecution-retained versus court-appointed experts: Comparing and contrasting risk assessment reports in preventative detention hearings. *Law and Human Behavior*, 38(6), 531–543. <https://doi.org/10.1037/lhb0000082>
- Brehm, J. W., & Cohen, A. R. (Eds.). (1962). The theory and the role of commitment. *Explorations in cognitive dissonance* (pp. 3–10). Wiley. <https://doi.org/10.1037/11622-001>
- Butt, A. E. (2017). Concurrent expert evidence in U.S. toxic harms cases and civil cases more generally: Is there a proper role for hot tubbing. *Houston Journal of International Law*, 40(1), 1–103.
- Cecil, J. S., & Willging, T. E. (1994). Accepting Daubert's invitation: Defining a role for court-appointed experts in assessing scientific validity. *Emory Law Journal*, 43, 995–1070.
- Champagne, A., Shuman, D. W., & Whitaker, E. (1996). The problem with empirical examination of the use of court-appointed experts: A report of non-findings. *Behavioral Sciences & the Law*, 14(3), 361–365. [https://doi.org/10.1002/\(SICI\)1099-0798\(199622\)14:3<361::AID-BSL245>3.0.CO;2-V](https://doi.org/10.1002/(SICI)1099-0798(199622)14:3<361::AID-BSL245>3.0.CO;2-V)
- Chanowitz, B., & Langer, E. J. (1981). Premature cognitive commitment. *Journal of Personality and Social Psychology*, 41(6), 1051–1063. <https://doi.org/10.1037/0022-3514.41.6.1051>
- Chevalier, C. S., Boccaccini, M. T., Murrin, D. C., & Varela, J. G. (2015). Static-99R reporting practices in sexually violent predator cases: Does norm selection reflect adversarial allegiance? *Law and Human Behavior*, 39(3), 209–218. <https://doi.org/10.1037/lhb0000114>
- Civil Justice Council. (2016). *Concurrent expert evidence and 'hot-tubbing' in English litigation since the 'Jackson reforms': A legal and empirical study*. <https://www.judiciary.uk/wp-content/uploads/2011/03/cjc-civil-litigation-review-hot-tubbing-report-20160801.pdf>
- Crano, W. D. (2000). Milestones in the psychological analysis of social influence. *Group Dynamics*, 4(1), 68–80. <https://doi.org/10.1037/1089-2699.4.1.68>
- Dattilio, F. M., Commons, M. L., Adams, K. M., Gutheil, T. G., & Sadoff, R. L. (2006). A pilot Rasch scaling of lawyers' perceptions of expert bias. *The Journal of the American Academy of Psychiatry and the Law*, 34(4), 482–491.

- Davies, G. L. (2004). Recent Australian development: A response to Peter Heerey. *Civil Justice Quarterly*, 23, 398–399.
- DeMatteo, D., Fishel, S., & Tansey, A. (2019). Expert evidence: The (unfulfilled) promise of Daubert. *Psychological Science in the Public Interest*, 20(3), 129–134. <https://doi.org/10.1177/1529100619894336>
- Devitt, M. R. (2012). A dip in the hot tub: Concurrent evidence techniques for expert witnesses in tax court cases. *Journal of Taxation*, 117(4), 213–224.
- Downes, G. (2004, February). *Concurrent expert evidence in the administrative appeals tribunal: The New South Wales experience* [Paper presentation]. Paper presented at the Australasian Conference of Planning and Environment Courts and Tribunals, Hobart, Australia.
- Edens, J. F., Smith, S. T., Magyar, M. S., Mullen, K., Pitta, A., & Petriola, J. (2012). Hired guns, “charlatans,” and their “voodoo psychobabble”: Case law references to various forms of perceived bias among mental health expert witnesses. *Psychological Services*, 9(3), 259–271. <https://doi.org/10.1037/a0028264>
- Edmond, G. (2003). After objectivity: Expert evidence and procedural reform. *The Sydney Law Review*, 25(2), 132–163.
- Edmond, G. (2009). Merton and the hot tub: Scientific conventions and expert evidence in Australian civil procedure. *Law and Contemporary Problems*, 72, 159–189.
- Edmond, G. (2008). Secrets of the “hot tub”: Expert witnesses, concurrent evidence and judge-led law reform in Australia. *Civil Justice Quarterly*, 27, 51–82.
- Elliot, A. J., & Devine, P. G. (1994). On the motivational nature of cognitive dissonance: Dissonance as psychological discomfort. *Journal of Personality and Social Psychology*, 67(3), 382–394. <https://doi.org/10.1037/0022-3514.67.3.382>
- Emmerig, J., Hanify, J. D., Legg, M., & Weida, J. C. (2013, April). *Room in American courts for an Australian hot tub?* Jones Day. <https://perma.cc/82ZL-X36Z>
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G\*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175–191. <https://doi.org/10.3758/BF03193146>
- Federal Rules of Evidence. 28 U.S.C. (2020). <https://www.rulesofevidence.org/>
- Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford University Press.
- Freckelton, I. (2005). Expert evidence in the family court: The new regime. *Psychiatry, Psychology and Law*, 12(1), 234–248. <https://doi.org/10.1375/pplt.2005.12.1.234>
- Garbis, M. J. (2003). Aussie inspired musings on technological issues: Of kangaroo courts, tutorials, and hot tub cross-examination. *The Green Bag*, 6(2), 141–150.
- Grøndahl, P., Stridbeck, U., & Grønnerød, C. (2013). The truth and nothing but the truth: Court-appointed forensic experts’ experience with testifying and their perceptions of legal actors in the criminal courts. *Journal of Forensic Psychiatry & Psychology*, 24(2), 192–204. <https://doi.org/10.1080/14789949.2013.771278>
- Hart, W., Albarracín, D., Eagly, A. H., Brechan, I., Lindberg, M. J., & Merrill, L. (2009). Feeling validated versus being correct: A meta-analysis of selective exposure to information. *Psychological Bulletin*, 135(4), 555–588. <https://doi.org/10.1037/a0015701>
- Jonas, E., Schulz-Hardt, S., Frey, D., & Thelen, N. (2001). Confirmation bias in sequential information search after preliminary decisions: An expansion of dissonance theoretical research on selective exposure to information. *Journal of Personality and Social Psychology*, 80(4), 557–571. <https://doi.org/10.1037/0022-3514.80.4.557>
- Joule, R.-V., & Azdia, T. (2003). Cognitive dissonance, double forced compliance, and commitment. *European Journal of Social Psychology*, 33(4), 565–571. <https://doi.org/10.1002/ejsp.165>
- Krafka, C., Dunn, M. A., Johnson, M. T., Cecil, J. S., & Miletich, D. (2002). Judge and attorney experiences, practices, and concerns regarding expert testimony in federal civil trials. *Psychology, Public Policy, and Law*, 8(3), 309–332. <https://doi.org/10.1037/1076-8971.8.3.309>
- Krauss, D. A., Gongola, J., Scurich, N., & Busch, B. (2018). Mental state at time of offense in the hot tub: An empirical examination of concurrent expert testimony in an insanity case. *Behavioral Sciences & the Law*, 36(3), 358–372. <https://doi.org/10.1002/bsl.2348>
- Kristjanson, F. (2012). “Hot-tubs” and concurrent evidence: Improving administrative proceedings. *Canadian Journal of Administrative Law & Practice*, 25, 79–94.
- Lindsay, R. C. L., MacDonald, P., & McGarry, S. (1990). Perspectives on the role of the eyewitness expert. *Behavioral Sciences & the Law*, 8(4), 457–464. <https://doi.org/10.1002/bsl.2370080411>
- Lord Arbinger v Ashton, 17 LR Eq 358 1873.
- Lundgren, S. R., & Prislin, R. (1998). Motivated cognitive processing and attitude change. *Personality and Social Psychology Bulletin*, 24(7), 715–726. <https://doi.org/10.1177/0146167298247004>
- Mackie, D. M. (1986). Social identification effects in group polarization. *Journal of Personality and Social Psychology*, 50(4), 720–728. <https://doi.org/10.1037/0022-3514.50.4.720>
- Mackie, D. M., Worth, L. T., & Asuncion, A. G. (1990). Processing of persuasive in-group messages. *Journal of Personality and Social Psychology*, 58(5), 812–822. <https://doi.org/10.1037/0022-3514.58.5.812>
- Malsch, M., & Freckelton, I. (2005). Expert bias and partisanship: A comparison between Australia and the Netherlands. *Psychology, Public Policy, and Law*, 11(1), 42–61. <https://doi.org/10.1037/1076-8971.11.1.42>
- Martire, K. A., & Edmond, G. (2017). Rethinking expert opinion evidence. *Melbourne University Law Review*, 40(3), 967–998.
- McAuliff, B. D., & Arter, J. L. (2016). Adversarial allegiance: The devil is in the evidence details, not just on the witness stand. *Law and Human Behavior*, 40(5), 524–535. <https://doi.org/10.1037/lhb0000198>
- McClellan, P. (2007, July). *Contemporary challenges for the justice system - Expert evidence* [Paper presentation]. Paper presented at the Australian Lawyers’ Alliance Medical Law Conference, Sydney, Australia.
- Murrie, D. C., & Boccaccini, M. T. (2015). Adversarial allegiance among expert witnesses. *Annual Review of Law and Social Science*, 11(1), 37–55. <https://doi.org/10.1146/annurev-lawsocsci-120814-121714>
- Murrie, D. C., Boccaccini, M. T., Guarnera, L. A., & Rufino, K. A. (2013). Are forensic experts biased by the side that retained them? *Psychological Science*, 24(10), 1889–1897. <https://doi.org/10.1177/0956797613481812>
- Murrie, D. C., Boccaccini, M. T., Johnson, J. T., & Janke, C. (2008). Does interrater (dis)agreement on Psychopathy Checklist scores in sexually violent predator trials suggest partisan allegiance in forensic evaluations? *Law and Human Behavior*, 32(4), 352–362. <https://doi.org/10.1007/s10979-007-9097-5>
- Murrie, D. C., Boccaccini, M. T., Turner, D. B., Meeks, M., Woods, C., & Tussey, C. (2009). Rater (dis)agreement on risk assessment measures in sexually violent predator proceedings: Evidence of adversarial allegiance in forensic evaluation? *Psychology, Public Policy, and Law*, 15(1), 19–53. <https://doi.org/10.1037/a0014897>
- National Research Council, & Committee on Identifying the Needs of the Forensic Science Community. (2009). *Strengthening forensic science in the United States: A path forward*. National Academies Press. <https://www.ncjrs.gov/pdffiles1/nij/grants/228091.pdf>
- Neal, T. M. S. (2016). Are forensic experts already biased before adversarial legal parties hire them? *PLoS ONE*, 11(4), e0154434. <https://doi.org/10.1371/journal.pone.0154434>
- Neal, T. M. S., & Brodsky, S. L. (2016). Forensic psychologists/perceptions of bias and potential correction strategies in forensic mental health evaluations. *Psychology, Public Policy, and Law*, 22(1), 58–76. <https://doi.org/10.1037/law0000077>

- Neal, T. M., & Brodsky, S. L. (2014). Occupational socialization's role in forensic psychologists' objectivity. *Journal of Forensic Psychology Practice, 14*(1), 24–44. <https://doi.org/10.1080/15228932.2013.863054>
- Neal, T. M. S., & Grisso, T. (2014). The cognitive underpinnings of bias in forensic mental health evaluations. *Psychology, Public Policy, and Law, 20*(2), 200–211. <https://doi.org/10.1037/a0035824>
- Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences of the United States of America, 115*(11), 2600–2606. <https://doi.org/10.1073/pnas.1708274114>
- Otto, R. K. (1989). Bias and expert testimony of mental health professional in adversarial proceedings: A preliminary investigation. *Behavioral Sciences & the Law, 7*(2), 267–273. <https://doi.org/10.1002/bsl.2370070210>
- Packer, I. K. (2009). *Evaluation of criminal responsibility*. Oxford University Press. <https://doi.org/10.1093/med:psych/9780195324853.001.0001>
- Perez, O. (2016). Judicial strategies for reviewing conflicting expert evidence: Biases, heuristics, and higher-order evidence. *The American Journal of Comparative Law, 64*(1), 75–120. <https://doi.org/10.5131/AJCL.2016.0003>
- Prescott, D. E., & Fadgen, T. (2019). Adversarial systems and forensic experts in child custody: How about adding a hot tub? *Journal of the American Academy of Matrimonial Lawyers, 32*(1), 117–144.
- President's Council of Advisors on Science and Technology. (2016). *Report to the President: Forensic science in the criminal courts: Ensuring scientific validity of feature-comparison methods*. Executive Office of the President of the United States. [https://www.whitehouse.gov/sites/default/files/microsites/ostp/PCAST/pcast\\_forensic\\_science\\_report\\_final.pdf](https://www.whitehouse.gov/sites/default/files/microsites/ostp/PCAST/pcast_forensic_science_report_final.pdf)
- Pronin, E., Lin, D. Y., & Ross, L. (2002). The bias blind spot: Perceptions of bias in self versus others. *Personality and Social Psychology Bulletin, 28*(3), 369–381. <https://doi.org/10.1177/0146167202286008>
- Rackemann, M. E. (2011, March). *Expert evidence reforms—how are they working* [Paper presentation]. Annual Conference of the Bar Association of Queensland, Gold Coast, Queensland, Australia.
- Reifert, E. (2011). Getting into the hot tub: How the United States could benefit from Australia's concept of "hot tubbing" expert witnesses. *University of Detroit Mercy Law Review, 89*(1), 103–115.
- Robertson, C. T. (2010). Blind expertise. *New York University Law Review, 85*, 174–257. <https://ssrn.com/abstract=1416943>
- Saks, M. J. (1990). Witnesses, nonexpert witnesses, and nonwitness experts. *Law and Human Behavior, 14*(4), 291–313. <https://doi.org/10.1007/BF01068158>
- Sanders, J. (2007). Expert witness ethics. *Fordham Law Review, 76*(3), 1539–1584.
- Slobogin, C. (2014). Lessons from inquisitorialism. *Southern California Law Review, 87*(3), 699–731.
- Sonenshein, D., & Fitzpatrick, C. (2013). The problem of partisan experts and the potential for reform through concurrent evidence. *The Review of Litigation, 32*(1), 1–64.
- Stridbeck, U., Grøndahl, P., & Grønnerød, C. (2016). Expert for whom? Court-appointed versus party-appointed experts. *Psychiatry, Psychology and Law, 23*(2), 246–255. <https://doi.org/10.1080/13218719.2015.1052334>
- Sutherland, N. (2011). The efficacy of joint reports in narrowing technical issues during litigation. *National Environmental Law Review, 1*, 52–55.
- Sward, E. (1989). Values, ideology, and the evolution of the adversarial system. *Indiana Law Journal, 64*(2), 301–355.
- Wason, P. C., & Johnson-Laird, P. N. (1972). *Psychology of reasoning: Structure and content*. Harvard University Press.
- Welch, S. (2010). From witness box to the hot tub: How the "hot tub" approach to expert witnesses might relax an American finder of fact. *Journal of International Commercial Law and Technology, 5*(3), 154–164.
- Wood, L. (2007a). Experts in the tub. *Antitrust, 21*(3), 95–103.
- Wood, L. (2007b). Experts only: Out of the hot tub and into the joint conference. *Antitrust, 22*(1), 89–95.
- Wood, L. (2018). Hot tub redux. *Antitrust, 32*(2), 31–36.
- Worthington, D. L., Stallard, M. J., Price, J. M., & Goss, P. J. (2002). Hindsight bias, Daubert, and the silicone breast implant litigation. *Psychology, Public Policy, and Law, 8*(2), 154–179. <https://doi.org/10.1037/1076-8971.8.2.154>
- Yarnall, M. A. (2009). Dueling scientific experts: Is Australia's hot tub method a viable solution for the American judiciary? *Oregon Law Review, 88*(1), 311–340.
- Zapf, P. A., Kukucka, J., Kassir, S. M., & Dror, I. E. (2018). Cognitive bias in forensic mental health assessment: Evaluator beliefs about its nature and scope. *Psychology, Public Policy, and Law, 24*(1), 1–10. <https://doi.org/10.1037/law0000153>
- Zusman, J., & Simon, J. (1983). Differences in repeated psychiatric examinations of litigants to a lawsuit. *The American Journal of Psychiatry, 140*(10), 1300–1304. <https://doi.org/10.1176/ajp.140.10.1300>

Received September 7, 2020

Revision received April 16, 2021

Accepted April 20, 2021 ■