



# Quality-checking the new normal: trial modality in online jury decision-making research

Evelyn M. Maeder<sup>1</sup> · Susan Yamamoto<sup>2</sup> · Logan Ewanation<sup>3</sup>

Accepted: 10 April 2023

© The Author(s), under exclusive licence to Springer Nature B.V. 2023

## Abstract

**Objectives** We sought to examine differences between videotaped and written trial materials on verdicts, perceptions of trial parties, quality check outcomes, perceived salience of racial issues, and emotional states in a trial involving a Black or White defendant.

**Hypotheses** We predicted that verdicts and ratings of trial parties would be similar for those participants viewing a videotaped trial and those reading a written transcript. However, we suspected that emotional states might be heightened for those watching a video and that those reading transcripts would perform better on quality checks regarding trial content (but worse on those involving trial party characteristics, including defendant race).

**Method** Participants ( $N=139$  after removing those who did not meet our threshold for data quality) recruited from Amazon's Mechanical Turk were randomly assigned to watch a video or read a transcript of a trial involving an alleged murder of a police officer. They completed a questionnaire probing their verdict, perceptions of trial parties, perceived salience of racial issues, and emotional state, and responded to a series of quality checks.

**Results** Participants in the videotape condition performed significantly worse on quality checks than did those in the transcript condition. There were no significant differences between modalities in terms of verdict or perceived salience of racial issues. Some other differences emerged between conditions, however, with more positive perceptions of the pathologist and police officer in the transcript condition, and more negative emotion elicited by the trial involving a White defendant in the videotape condition only.

**Conclusions** There were no meaningful differences between videotaped and written trial materials in terms of outcome (verdict), but the presence of some trial party rating and emotional state differences stemming from modality epitomizes the internal/

---

✉ Evelyn M. Maeder  
evelyn.maeder@carleton.ca

<sup>1</sup> Carleton University, Ottawa, Canada

<sup>2</sup> Campion College, University of Regina, Regina, Canada

<sup>3</sup> Ontario Tech University, Oshawa, Canada

ecological validity trade-off in jury research. Our quality check results indicate that written transcripts may work better for obtaining valid data online. Regardless of modality, researchers must be diligent in crafting quality checks to ensure that participants are attending to the stimulus materials, particularly as more research shifts online.

**Keywords** Trial modality · Online research · Jury decision-making

“In the end, however, it is incumbent on us, if we want to assure ourselves and to persuade courts how informative our simulation research can be, to continue to test the effects of variations in methodology on our findings.” Diamond (1997, p. 570).

Online research has become increasingly common in jury research, particularly with the rising popularity of crowdsourcing platforms such as Mechanical Turk, Crowdflower, and Prolific Academic (e.g., Baker et al., 2016). Some previous jury research has compared findings from online and in-person data collection. For example, Maeder et al. (2018) compared in-person to online samples in a study of the effects of defendant race (Black, White, Indigenous) and found that while demographic compositions of these samples did differ, failure rates of quality checks were comparable. These authors observed that while there was no main effect of data collection type, online samples demonstrated increased leniency toward the White defendant as compared to in-person samples, potentially owing to the presence of racialized participants and/or research assistants in the room during in-person data collection. More recently, Jones and colleagues (2022) demonstrated that data collection type (in-person vs. online) did not exhibit a main effect on verdict in a trial involving pre-trial publicity (PTP). However, data collection type did interact with an implicit bias remedy, such that among participants who completed the study in person, those who were exposed to the remedy were actually *more* likely to convict the defendant and provided harsher sentences than those who were not (there were no differences as a function of implicit bias remedy in the online conditions). Therefore, while there is no evidence that data collection mode exhibits a significant main effect on trial outcomes, it may interact with other variables.

Other research has compared different modalities of trial presentation (e.g., Pezdek et al., 2010), again revealing few significant differences. However, this research has largely attended to trial outcomes (e.g., verdict) and for the most part has not examined whether trial parties are perceived differently as a function of trial modality. In addition, to our knowledge, all the existing research comparing modalities has done so using in-person data collection techniques. Given the large shift toward online research, particularly in the wake of the COVID-19 pandemic and beyond, it is essential to test whether trial modalities are equivalent in this context as well.

## Modality

Most jury simulation research relies on written trial transcripts, although the use of non-written materials is on the rise (Bornstein, 2017). In terms of ecological validity, it stands to reason that a videotaped trial presentation offers a higher degree of

realism than does a written transcript. In real trials, jurors are exposed to information through audio and visual channels (Koehler & Meixner, 2017); therefore, a videotaped trial conveys this more closely than reading witness testimony and lawyers' arguments on paper. It should be noted, however, that studies using videotaped stimuli are not by definition more ecologically valid than those using written transcripts—arguably, a mock trial video that includes legally unrealistic information (e.g., evidence that would be excluded under that jurisdiction's evidentiary rules) or omits jury instructions is less ecologically valid than a written transcript that properly reflects how trials in that jurisdiction are conducted, albeit in written form. That is to say, studies using videotaped stimuli are not *de facto* more ecologically valid than studies using written stimuli—the content of these stimuli must be considered (see, e.g., Vidmar's [1979] discussion of legal naivete among jury researchers). To the extent that mock trial videos are legally realistic, though, they certainly allow for closer replication of a real trial setting (and its accompanying "offstage" components; Rose et al., 2010) than do written transcripts.

Of course, along with the potential boost in ecological validity might come a decrease in internal validity, given the influence of extraneous variables (e.g., physical attributes of the trial parties, differences in delivery) in video stimuli (Sleed et al., 2002). In addition, the creation of videotaped trials involves a great deal of expense, particularly if professional actors are employed. The alternative use of amateur actors may be less expensive; however, we argue that acting ability has the strong potential of introducing confounds to trial party credibility not present in written materials. If an actor is unskilled, it is possible that jurors will find the trial party portrayed by the actor to be unbelievable, not because of the content of their testimony but because of its delivery.

Prior to 1999, a handful of studies had compared videotaped trials to transcripts. As reviewed by Bornstein (1999), these studies had mostly null findings on outcome variables of interest when comparing these two modalities (Fishfader et al., 1996; Lassiter et al., 1992; Williams et al., 1975), although two demonstrated that guilty verdicts were more likely in videotaped as compared to transcript conditions (Juhnke et al., 1979; Wilson, 1996). Since that review, there have been few studies directly comparing videotaped to written materials. In a study of presentation modality (videotaped vs. written) of sexual assault vignettes, Sleed et al. (2002) found that participants were more likely to blame the victim and less likely to define the encounter as a rape in a written vignette involving alcohol use, but that there were no differences between video and written vignettes in two other scenario types.

Pezdek and colleagues (2010) tested whether modality (videotape or transcript) interacted with other variables of interest (e.g., the presence of expert testimony) in a study evaluating eyewitness expert testimony. In their first study, participants rated the eyewitness expert as higher on cognitive measures (i.e., more understandable, more informative, and less confusing) and global/affective measures (i.e., more impactful, more useful, more influential, and better able to enhance the credibility of psychological science) when they read a trial transcript as compared to when they watched a video. However, it is noteworthy that in the transcript condition, participants were able to re-read as much and as often as they liked, whereas in the video condition they were only able to watch the presentation once. The authors also noted

that in this study, the presentation style of the expert was quite flat, and so, they conducted a second study using an expert with a more dynamic presentation style. In this study, the differences largely disappeared, and modality did not interact with other variables. The authors concluded that because there were no significant interactions of modality with other variables of interest, both modalities are acceptable in jury research.

A study of the perceived acceptability of jury research practices (Lieberman et al., 2016) found that among jury research experts (defined as those who published and/or peer-reviewed jury decision-making studies), video simulations were rated as significantly more acceptable than written transcripts. However, over 90% indicated that written transcripts still met minimal standards for publication.

From the above, it appears that findings regarding modality are somewhat mixed, with the majority demonstrating few meaningful differences between videotaped and written trial materials. However, most of these studies focused largely on verdict differences, and it is possible that perceptions of trial parties themselves may differ as a function of modality.

### **Modality and trial party characteristics**

Chaiken and Eagly (1976, 1983) have written extensively on communicator characteristics and modality. They argue that in videotaped materials, nonverbal cues may increase attentiveness to the communicator's message, increasing the salience of their personal characteristics. This increased attention could have positive or negative effects on the communicator's persuasiveness, depending on the cues involved. In the context of a mock trial, a trial party (e.g., witness, lawyer, defendant) could be more or less persuasive to jurors depending upon whether they exhibit positive cues (e.g., likeability, expertise) or negative (e.g., untrustworthiness) cues. Chaiken and Eagly (1983) also predicted that participants exposed to videotaped material may be more focused on communicators themselves (i.e., using the heuristic mode of processing), whereas those exposed to written materials would attend more to the content of the message (i.e., using more systematic processing). In a series of two studies, they demonstrated that likeable communicators were more persuasive when their message was conveyed via videotape, whereas unlikable communicators were more persuasive when their message was presented in written form. Supporting their predictions about processing type, they also found that opinion change for participants who saw videotaped materials was predicted by communicator- but not message-based cognitions, whereas the opposite was true for participants who saw written materials. These findings have obvious implications for the potential differences of trial modality on trial party perceptions in experimental jury research.

To our knowledge, this potential difference in trial party persuasiveness/personal characteristics as a function of modality has not yet been tested. While Pezdek et al. (2010) did examine some characteristics of the defendant, alibi witness, and eyewitness with single-item measures in their study, most of their questions centered around the expert testimony (as this was the purpose of their research), and they did not evaluate differences in perceptions of the lawyers. Heath and colleagues (2004)

found that defendants in videotaped trials were perceived as more honest than those in audiotaped trials (and that defendant emotionality affected credibility assessments in the video condition only), but they did not include a written transcript condition.

In this study, we aimed to determine whether trial parties were perceived as more or less credible, likeable, or influential as a function of modality. We were also interested in whether participants would be more able to identify with the defendant when they could see and hear him present his testimony, and whether attorney effectiveness and competence would vary as well.

### **Modality and race**

It stands to reason that modality may be particularly relevant in studies involving racialized trial parties. As aforementioned, Chaiken and Eagly (1983) reasoned that personal characteristics may be more salient in videotaped as compared to written materials due to increased attention given to the communicator as a function of non-verbal cues (including appearance). This increased attention to the communicator also likely means that race is more vivid in videotaped rather than transcribed trials, whether race is manipulated in written materials via names, descriptions, photos, or some combination thereof.

We were unable to find any modality comparisons that spoke to this question directly. As such, this is, to our knowledge, the first study of trial modality that examines perceptions of a Black defendant. We sought to examine whether participants would be more likely to pass a memory check with regard to the defendant's race as a function of trial modality. We were also interested in whether participants would perceive race as a more salient issue (Sommers & Ellsworth, 2001, 2009) when viewing a videotape of a Black defendant as compared to reading his testimony with an accompanying photo. For comparison purposes, we also included transcript and video conditions featuring a White defendant.

### **Modality and elicited emotion**

Finally, we were interested in examining whether modality would affect jurors' emotional state following the trial. Compared to text, visual and audiovisual stimuli may elicit stronger emotional reactions across a variety of contexts (e.g., Clark & Paivio, 1991; Iyer & Oldmeadow, 2006; Yadav et al., 2011). For instance, Pfau and colleagues (2006) found that participants were significantly angrier after reading a newspaper article about war casualties that involved a picture as compared to a control, text-only condition. Similarly, Zupan and Babbage (2017) observed more intense feelings of happiness, sadness, and fear among participants who watched film clips compared to participants who read narrative texts.

Visual messages may be associated with stronger emotional reactions because they often invoke peripheral rather than central processing (Joffe, 2008). Furthermore, audiovisual stimuli may be particularly effective at eliciting emotions because we tend to automatically mimic the emotions that we observe in others (Coplan, 2006; Hatfield et al., 1993). Thus, participants watching a simulated trial video may

be particularly likely to exhibit emotions similar to those expressed by the trial parties. We were able to identify a single study that varied trial modality and measured mock jurors' emotional state; Fishfader et al. (1996) presented mock jurors with either a simulated video or a written transcript describing a wrongful death suit. Participants who viewed the videotaped trial experienced heightened negative emotions as compared to those who read the trial transcript.

## Purpose and hypotheses

Previous research has compared trial modalities in jury research with regard to verdict decisions, but very little attention has been paid to other outcomes that may be relevant to this line of research. Our study sought to compare videotaped and written trial stimuli on assessments of trial parties, salience of trial party race as well as racial issues more generally,<sup>1</sup> and elicited emotion. We were also curious about potential differences in the two modalities in terms of attentiveness, a particularly important question now that so much of this research has shifted online.

Based on previous studies comparing modality, we did not predict a difference in verdict decisions as a function of written vs. videotaped stimuli. Because so little work has investigated perceptions of trial parties as a function of modality, our approach for these variables was exploratory. However, in accordance with work done by Chaiken and Eagly (1983), we predicted that participants might perform better in memory/comprehension checks regarding content for the written as compared to the videotaped stimulus, but worse on these same checks regarding trial party characteristics (i.e., defendant race). Finally, based on research concerning presentation mode and emotions (e.g., Detenber et al., 1998; Fishfader et al., 1996; Zupan & Babbage, 2017) we predicted the videotaped stimuli would elicit stronger emotional reactions as compared to the written transcript.

## Method

### Participants

We collected data from  $N=207$  US jury-eligible (i.e., 18 years of age or older, US citizens, with no prior felony convictions) participants (53.4% men, 46.6% women), the majority of whom identified as White (73.3%), with smaller numbers of Black (14.2%), Native American (4.5%), Asian (3.4%), and Latine (3.4%) respondents, and

---

<sup>1</sup> Given the nature of the trial transcript (involving a charge of murder of a police officer resulting from an altercation between two police officers and the defendant, who claims self defense), we might also expect differences in decision-making as a function of defendant race. This dataset and manuscript are focused on the effects and moderators of trial modality; we report on a separate dataset using this same trial stimulus with more fulsome race considerations in another paper (forthcoming).

two who identified as another race/ethnicity. Participants ranged from 23 to 78 years of age ( $M=40.5$ ,  $SD=11.8$ ).<sup>2</sup>

## Materials

### Trial stimulus

Participants were presented with one of two trial stimuli. In the video condition, participants watched a video of a mock trial based on a real case (R. v. Gayle, 1994). The video was created by a production studio and made use of professional actors. The case involved an altercation between the defendant and two police officers, one of whom was killed. The defendant raised a claim of self-defense to the charge of murder of a police officer, arguing that the officer had used unreasonable force and the defendant reasonably feared for his life. The video included opening and closing procedural and substantive instructions read by the judge, opening and closing statements from the prosecution and defense, and testimony from the surviving police officer, an expert witness (pathologist) who had examined the deceased, an earwitness, and the defendant. The surviving officer described his perspective of the incident leading to the other officer's death. The expert pathologist testified as to the cause of death; he used technical terminology appropriate to his occupation, but his language was not overly complex (receiving a readability score of A when entered into readable.com). The earwitness testified for the defense, indicating that she first heard three shots, followed by two separate shots several moments later (she did not use technical language), corroborating the defendant's story about the exchange of gunfire. The defendant's race was mentioned twice in the trial—once in the police officer's testimony when he described the defendant's physical appearance (i.e., "when we entered the garage, we could see a White/Black male subject near the exit") and again when the defense lawyer read from his initial police report using the same language. Racial issues were not otherwise mentioned. Running time of the trial video was approximately 40 minutes.

In the transcript condition, participants read the script that was used to shoot the video (5680 words/approximately 13 single-spaced pages), with photos of the actors accompanying their parts. Therefore, participants received identical information in both conditions; only the modality was varied. All parts were played by White actors (all men, with the exception of the earwitness, who was portrayed by a woman) with the exception of the defendant, who was portrayed by either a White or a Black man (photos of whom were pilot tested to match on perceived age, attractiveness, and likeability,  $N=30$ ). For the video stimulus, the actors portraying the Black and White defendants ran lines together prior to shooting and observed each other being filmed, ensuring consistency of delivery.

---

<sup>2</sup> As mentioned below, several participants were omitted from analyses due to their responses to our quality checks. The remaining ( $N=139$ ) participants were comparable to the full sample in terms of demographic composition.

## Measures

### Verdict

Participants were asked to provide a verdict for the charge of murder of a police officer (1 = guilty, 2 = not guilty).

### Trial party ratings

For each of the witnesses (the surviving officer, the pathologist, and the earwitness), we asked participants to rate the credibility and likeability of the witness, along with the degree to which the witness influenced their verdicts (all on scales from 0 = not at all to 10 = very much). For the defendant, we repeated these questions with the addition of “how much were you able to identify with the defendant?”. For the attorneys, we asked participants to rate their effectiveness, competence, likeability, and influence on verdict using the same scale (see Table 1 for descriptives). We examined internal consistency reliability for the set of questions pertaining to the officer ( $\alpha=0.67$ ), pathologist ( $\alpha=0.64$ ), earwitness ( $\alpha=0.82$ ), defendant ( $\alpha=0.80$ ), prosecutor ( $\alpha=0.84$ ), and defense lawyer ( $\alpha=0.81$ ). Mean scores were computed for each set.

### Race salience

To measure perceived race salience, we asked participants “to what degree did racial issues play a role in the trial?” on a scale from 0 = not at all to 10 = very much.

### Quality checks

To ensure attentiveness to the materials, we included three memory checks (one asking participants to identify the defendant’s race, which was named in the surviving officer’s testimony and shown in the video/picture of the actor portraying the defendant; one asking participants to identify where the crime took place, which was identified in multiple testimonies; and one asking where on the body the victim was shot, which was identified in multiple testimonies). We also included one attention check (“To show that you have read this, please select 7”) embedded into the trial party ratings. Finally, we asked participants to provide an open-ended justification for their verdict decision, which we were able to use to identify bots/nonsensical responding (see “[Results](#)” section).

### State emotions

To measure state emotions, we included the Positive and Negative Affect Scale (PANAS; Watson et al., 1988). This measure asks participants to identify the extent to which they feel a variety of feelings and emotions at the present moment (e.g., distressed, interested, ashamed) on a scale from 1 (very slightly or not at all) to 5 (extremely). Separate mean scores were computed for positive ( $\alpha=0.89$ ) and negative ( $\alpha=0.96$ ) affect.



**Table 1** Descriptive statistics for perceptions of lawyers, witnesses, and defendant

	Stimulus type	<i>N</i>	Mean	<i>SD</i>
<b>Prosecutor</b>				
Effectiveness	Transcript	61	6.54	2.36
	Video	41	7.02	2.12
Competence	Transcript	61	7.03	2.15
	Video	41	7.32	2.15
Likeability	Transcript	61	6.20	2.50
	Video	41	5.71	2.73
Influence on verdict	Transcript	61	6.44	2.55
	Video	40	6.18	2.59
<b>Defense lawyer</b>				
Effectiveness	Transcript	61	7.46	1.82
	Video	41	7.27	1.63
Competence	Transcript	61	7.61	1.66
	Video	41	7.66	1.44
Likeability	Transcript	61	6.62	2.08
	Video	41	6.56	2.07
Influence on verdict	Transcript	61	7.02	2.44
	Video	41	6.71	2.11
<b>Police officer</b>				
Credibility	Transcript	61	6.21	3.18
	Video	41	4.71	2.84
Likeability	Transcript	61	5.97	2.93
	Video	41	5.34	2.92
Influence on verdict	Transcript	61	7.56	2.24
	Video	41	6.83	2.25
<b>Pathologist</b>				
Credibility	Transcript	61	8.28	1.72
	Video	41	8.07	1.68
Likeability	Transcript	61	7.25	1.75
	Video	41	6.68	1.98
Influence on verdict	Transcript	61	7.11	2.47
	Video	41	6.41	2.56
<b>Earwitness</b>				
Credibility	Transcript	61	7.16	2.48
	Video	41	6.41	2.36
Likeability	Transcript	61	7.11	1.88
	Video	41	6.56	2.07
Influence on verdict	Transcript	61	6.80	2.63
	Video	41	6.32	3.07
<b>Defendant</b>				
Credibility	Transcript	61	6.18	3.07
	Video	41	6.20	2.26

**Table 1** (continued)

	Stimulus type	<i>N</i>	Mean	<i>SD</i>
Likeability	Transcript	61	5.87	2.64
	Video	41	6.05	2.28
Influence on verdict	Transcript	61	6.95	2.46
	Video	41	6.93	2.16
Identified with	Transcript	61	5.79	2.99
	Video	41	5.59	2.94

## Demographics

Participants provided information regarding their age, gender, and race/ethnicity.

## Procedure

Participants were recruited via Amazon's Mechanical Turk, a crowdsourcing platform used in a great deal of jury research (Krauss & Lieberman, 2017), and completed the study using Qualtrics software. Upon clicking the link to the study, they were randomly assigned to either the video or transcript condition featuring either a Black or a White defendant. After watching/reading the stimulus, participants responded to the trial party rating items, followed by the manipulation checks, perceived race salience item, PANAS, and finally the demographics survey. At the end of the study, participants received \$5 in compensation. For those who read the trial transcript, mean completion time of the study was 30.05 ( $SD=23.45$ ) minutes; for those who saw the trial video, mean completion time was 44.22 ( $SD=29.93$ ) minutes (these completion times were significantly different from one another,  $t[295] = -4.56$ ,  $d = -0.53$ ).<sup>3</sup>

## Results

### Analytic strategy

This study used a 2 (defendant race: Black, White) by 2 (modality: transcript, video) between-subjects factorial design. First, quality checks were examined on the full dataset using a series of chi-square contingency tests. A hierarchical binary logistic regression was used to probe for verdict disparities as a joint function of defendant race and modality. We then ran analyses of variance (ANOVAs) using state emotion and the composite variables for perceptions of the trial parties as dependent variables. A post hoc analysis of achieved power for an ANOVA with fixed main effects

<sup>3</sup> This analysis includes the full sample of respondents, including those who failed our quality checks.

and interactions at  $\alpha=0.05$ , an effect size of 0.25,  $N=139$ , and  $df=1$  for four groups yielded 0.83 power.

## Quality checks

### Identifying bots

To ensure that only human respondents were included in analyses, two independent raters coded participants' open-ended verdict rationales. Raters were trained to flag instances in which the responses were non-sensical, unrelated to the case (e.g., one response referred to the defendant as a football player), or copy-pasted passages from the trial stimulus. Such response types were indicative of AI-generated sentences. Based on these criteria, initial interrater reliability was strong (Cohen's  $Kappa=0.72$ ). Disagreements pertained almost exclusively to whether blank responses should be counted as humans or bots. After discussion, we elected to retain blank responses given that participants were not obligated to respond to every question and that a blank response was not itself a nonsensical answer. This process resulted in the elimination of 31 cases, which were excluded from our main analyses.

### Memory/attention checks

We asked people to identify the race of the defendant. We conducted two chi-square contingency tests for each race condition separately. There was no significant contingency between correct answers and modality for either the White defendant,  $\chi^2(N=78)=0.05$ ,  $p=0.815$ ,  $v=0.03$ , or the Black defendant,  $\chi^2(N=98)=2.86$ ,  $p=0.091$ ,  $v=0.17$ . Both groups had a high proportion of passes in the transcript condition (87.5% for the Black defendant, and 86.8% for the White defendant) as well as the video condition (74% for the Black defendant and 85% for the White defendant).

Participants who saw the trial transcript (95%) were significantly more likely to pass the memory check regarding where the crime took place,  $\chi^2(N=176)=11.47$ ,  $p=0.009$ ,  $v=0.25$ , than did those in the trial video (79%) condition. This pattern did not hold for the memory check regarding where on the body the victim was shot,  $\chi^2(N=176)=7.28$ ,  $p=0.064$ ,  $v=0.20$ , with 88% in the transcript condition passing and 72% in the video condition passing.

These failures likely stemmed from a lack of attending to stimulus materials. For example, those participants who failed the race memory check were also significantly more likely to fail the memory check with regard to where the crime took place,  $\chi^2(N=176)=15.73$ ,  $p=0.001$ ,  $v=0.29$ , and the memory check regarding where on the body the victim was shot,  $\chi^2(N=176)=14.23$ ,  $p=0.003$ ,  $v=0.27$ , than did those who answered it correctly. This implies that these participants likely skimmed or did not read the transcript, or pressed play on the video and did not watch it (or divided their attention).

**Table 2** Hierarchical logistic regression with modality and defendant race as the independent variables and dichotomous verdict as the dependent variable

		<i>B</i>	<i>S.E</i>	<i>Sig</i>	<i>Exp(B)</i>
Step 1	Defendant Race	-.44	.34	.204	.65
	Modality	-.04	.34	.913	.96
Step 2	Defendant Race x Modality	-.46	.69	.502	.63

Not guilty was coded as 0 and guilty as 1

We also included an attention check in which we asked participants to “please select 7,” embedded in the trial party perception questionnaire. Interestingly, very few participants failed this check (5 in the video condition, 1 in the transcript condition), suggesting that online participants (or at least MTurk workers) know to be on the lookout for this type of quality check.

There was a significant contingency between passing all four checks and modality,  $\chi^2(N=176)=10.77$ ,  $p=0.001$ ,  $v=-0.25$ . Those in the transcript condition (78% passes) significantly outperformed those in the video condition (54% passes).

For this study, we were concerned with whether different modalities would yield significant differences in perceptions such as trial party credibility, prominence of racial issues, and ability to identify with the defendant. As such, it is necessary to include only participants who actually attended to trial materials in those analyses. To capture this, we included in subsequent analyses only those participants who passed 3 out of the 4 attention/memory checks, resulting in a loss of 37 participants. Combined with our elimination of participants identified as bots, this resulted in a new sample size of  $N=139$ , with 32 participants in the White defendant/transcript condition, 29 participants in the White defendant/video condition, 42 participants in the Black defendant/transcript condition, and 36 in the Black defendant/video condition.

## Verdict

For the White defendant, verdicts were fairly evenly split with 57.9% voting guilty in the transcript condition and 62.5% in the video condition. The Black defendant condition was comparable, with 47.6% and 41.7% for the transcript and video conditions respectively. We conducted a hierarchical logistic regression using defendant race (0 = White, 1 = Black) and modality (0 = transcript, 1 = video) as the independent variables, with dichotomous verdict (0 = not guilty, 1 = guilty) as the dependent variable. Main effects were entered in step 1 and the interaction effect in step 2. No effects reached statistical significance (see Table 2).

## Perceived race salience

As would be expected, there was a significant main effect of defendant race ( $F(1, 135)=34.32$ ,  $p<0.001$ , partial  $\eta^2=0.20$ ), such that perceived race salience was higher when the defendant was Black ( $M=6.19$ ,  $SD=2.92$ ) as compared to White ( $M=3.00$ ,  $SD=3.45$ ). There were no significant effects of modality ( $F(1,$

135)=0.59,  $p=0.442$ , partial  $\eta^2=0.004$ ) nor a significant interaction effect ( $F(1, 135)=0.04$ ,  $p=0.846$ , partial  $\eta^2=0.00$ ). For the Black defendant condition, participants who watched the trial video ( $M=6.36$ ,  $SD=2.83$ ) were not significantly different from those who read the trial transcript with photos ( $M=6.05$ ,  $SD=3.02$ ) in terms of the degree to which they felt that racial issues played a role in the trial. For the White defendant condition, perceived race salience was also comparable between those who watched the video ( $M=3.28$ ,  $SD=3.62$ ) and those who read the transcript ( $M=2.75$ ,  $SD=3.52$ ).

### State emotion

We conducted separate 2 (defendant race: Black, White) by 2 (modality: transcript, video) ANOVAs using positive affect and negative affect as the dependent variables. For positive affect, there was no significant effect of race ( $F(1, 135)=1.99$ ,  $p=0.161$  partial  $\eta^2=0.015$ ) or modality ( $F(1, 135)=3.61$ ,  $p=0.060$ , partial  $\eta^2=0.026$ ), nor a significant interaction ( $F(1, 135)=0.37$ ,  $p=0.546$ , partial  $\eta^2=0.003$ ). Negative affect yielded a significant interaction effect ( $F(1, 135)=4.71$ ,  $p=0.032$ , partial  $\eta^2=0.034$ ) such that there was a difference within the video but not transcription condition. Negative affect was stronger ( $d=0.53$   $p=0.030$ , 95% CI [0.05, 1.00]) for the trial involving a White defendant ( $M=2.15$ ,  $SD=1.03$ ) as compared to a Black defendant ( $M=1.61$ ,  $SD=0.95$ ).

### Trial party ratings

#### Lawyers

Separate ANOVAs using the composite perceptions variables for each lawyer yielded non-significant main effects of modality and a non-significant interaction effect ( $p$ 's between 0.251 and 0.659). There was a significant main effect of defendant race on both perceptions of the prosecution ( $F(1, 135)=5.27$ ,  $p=0.023$ , partial  $\eta^2=0.04$ ) and defense ( $F(1, 135)=4.59$ ,  $p=0.034$ , partial  $\eta^2=0.03$ ). Perceptions of the defense lawyer were more positive when the defendant was Black ( $M=7.16$ ,  $SD=1.68$ ) as compared to White ( $M=6.52$ ,  $SD=1.79$ ). Perceptions of the prosecutor had the opposite pattern, with more positive perceptions when the defendant was White ( $M=7.01$ ,  $SD=1.78$ ) as compared to Black ( $M=6.26$ ,  $SD=2.06$ ) (see Table 2 for details).

#### Witnesses

We compared our transcript and video conditions on composite variables of participants' ratings of all three witness's credibility, likeability, and the degree to which they felt his testimony had influenced their verdict. Table 3 displays the main effects and interactions for each witness. This analysis revealed a significant main effect of defendant race for the police officer, such that perceptions were more positive for the trial involving a White defendant ( $M=6.68$ ,  $SD=1.99$ ) as compared to

**Table 3** Analyses of variance examining the effect of modality and defendant race on trial party perceptions

Effect	Trial party	<i>F</i>	Sig	Partial $\eta^2$
Defendant race	Officer	5.99	.016*	.901
	Pathologist	1.95	.165	.014
	Earwitness	.56	.457	.004
	Defendant	.009	.926	.000
	Prosecutor	5.27	.023*	.038
	Defense lawyer	4.59	.034*	.033
Modality	Officer	5.23	.024*	.037
	Pathologist	6.10	.015*	.043
	Earwitness	2.60	.109	.019
	Defendant	2.96	.088	.021
	Prosecutor	.20	.659	.001
	Defense lawyer	1.33	.251	.010
Race $\times$ modality	Officer	1.63	.204	.012
	Pathologist	.04	.835	.000
	Earwitness	.71	.401	.005
	Defendant	.38	.538	.003
	Prosecutor	.587	.445	.004
	Defense lawyer	.46	.499	.003

\*Significant at  $p < .05$ 

a Black defendant ( $M=5.85$ ,  $SD=2.22$ ). The police officer was also viewed more positively in the transcript condition ( $M=6.62$ ,  $SD=2.10$ ) than in the video condition ( $M=5.76$ ,  $SD=2.15$ ). Similarly, perceptions of the pathologist were more favorable in the transcript ( $M=7.54$ ,  $SD=1.53$ ) as compared to the video ( $M=6.87$ ,  $SD=1.63$ ) condition. There were no other significant effects.

## Defendant

We asked participants to rate the credibility and likeability of the defendant, the degree to which his testimony influenced their verdict decision, and the degree to which they identified with the defendant. A composite variable did not yield any significant differences as a function of presentation mode or defendant race (see Table 3 for details).

## Discussion

The purpose of this study was to investigate differences in perceptions of trial parties as a function of modality (a written transcript vs. a mock trial video). Given that any such distinctions might be especially strong for cases involving racialized persons, we compared the effects of modality for trials involving Black or White defendants. However, results did not support the prediction that racial issues would

be perceived as more salient for a video as compared to a transcript of a trial with a Black defendant. Supporting previous research demonstrating negligible differences between transcripts and videos (Pezdek et al., 2010), there were few significant differences in witness or defendant likeability, credibility, or perceived influence on verdict between the two stimulus types. One difference emerged for the police officer, who was perceived more positively in the transcript condition. It is possible that this result stems from the fact that the officer's cross-examination featured an inconsistency in his story; he was challenged on the fact that his account changed immediately after the incident compared to sometime later. It appears that something about the actor's delivery of the script decreased his credibility, which may have been appropriate to the content of the testimony. The pathologist also elicited less favorable perceptions in the video as compared to in the transcript. These differences also likely reflect the acting choices made by the actor, who properly reflected the delivery that such testimony would likely entail, particularly given that pathologists and those medical students interested in pathology are more likely to identify as introverted (Fielder et al., 2022). There were some significant differences as a function of defendant race, such that both the police officer and the prosecutor yielded more positive perceptions when the defendant was White, while the defense yielded more positive perceptions when the defendant was Black. Finally, there was a significant race by modality interaction effect on negative (but not positive) affect, such that the White defendant yielded stronger negative affect as compared to the Black defendant, but only in the video condition. As expected, verdict decisions were comparable between the transcript and video conditions.

We also examined potential variations in attention between the two stimuli by including a series of memory checks regarding the case facts, in addition to a manipulation check of the defendant's race. We predicted that participants would perform better in memory checks regarding content for the written as compared to the videotaped stimulus, but worse on the manipulation check. Overall, those in the transcript condition significantly outperformed those in the video condition on both the manipulation and attention checks.

## Implications

This study began as an investigation into differences in trial modality but yielded some unexpected findings regarding the reliability of online data, which has taken on new importance as a result of the COVID-19 pandemic. Indeed, findings serve as a cautionary tale for ways in which MTurk workers might circumvent traditional attention metrics. Although there were negligible differences in perceptions of trial parties, the significantly higher rate of attention and manipulation check passes among the transcript condition suggest that video stimuli might be less feasible. This likely does not reflect participants' ability to retain information from video vs. written material so much as it suggests that participants who are asked to watch a video online are much less likely to attend to the stimulus than are those who are asked to read a transcript. Two possibilities here are that participants press play on the video but focus on other tasks, and/or that participants in the transcript condition copy and

paste the text for consultation when completing quality checks. Because participants are online and therefore not observable, we can only speculate as to their behavior during these studies. This is one of the major drawbacks of online research, and researchers must develop careful and clever attention, memory, and manipulation checks to ensure that data are high quality. Finally, the clear participation of bots should raise red flags for researchers and calls for novel quality checks that can identify this particular issue.

In terms of the original research question, which focused on trial modality, our findings have important implications. First, in terms of major outcome variables (notably, verdict), we provide further evidence that videotaped and written stimuli are acceptably similar. Contrary to our predictions, we did not observe overall differences in affect stemming from trial modality—with the exception of one interaction with defendant race noted below, reading a trial transcript and watching a videotaped trial elicited similar positive and negative affect among our participants. While this is a promising finding for those wishing to continue to rely on trial transcripts, we do think that more work is needed to compare trial modalities across a variety of case types to determine whether case type itself is a moderator of this effect (given that the previous study examining affect, using a civil case, did find differences in emotional response as a function of modality; Fishfader et al., 1996).

Interestingly, the differences stemming from modality involved the ratings of two witnesses and an interaction with defendant race on negative affect, which has implications for internal and ecological validity. Using a trial transcript will allow researchers more control over the “delivery” of testimony—acting choices will not affect perceptions of testimony if the participant reads the words directly from the page. In this study, it is very likely that the choices made by the actors (i.e., pathologist, police officer, White defendant) elicited our observed differences in perceptions and negative affect. However, in an actual trial, testifying witnesses will necessarily have individual differences in delivery, and so a videotaped stimulus more closely replicates that reality. As such, the question of whether to use videotaped or written stimuli hinges on whether the researcher wants more control over the manipulated variables (i.e., to avoid potential confounds stemming from acting choices), or to more closely match what jurors would experience in a real trial—the age-old internal/ecological validity trade-off (Lieberman et al., 2016). Our research contributes to this discussion by identifying the influence of modality on some trial party ratings in one case fact pattern—future research must continue to examine this question across a variety of case types.

## Limitations

The chief limitation of the current study is the use of MTurk workers; it is unclear how these findings might translate to in-person and other samples. Many studies have flagged possible significant differences in decreased attention for online methodologies, in addition to greater experience with classic psychology paradigms (e.g., the prisoner’s dilemma) and distinct demographics (Chandler et al., 2014; Paolacci et al., 2010). It is therefore possible that our finding would not hold for other



recruitment platforms. Our experience with MTurk along with the body of experimental work tells us that workers are highly accustomed to manipulation and attention checks. Indeed, many recruiters limit participants to those who have at least a 98% approval rating. It is worth noting that workers may get around attention metrics in ways that are difficult if not impossible for researchers to discern, for instance by saving written materials to later keyword search. Researchers should exert caution in using traditional quality check techniques and should not rely on any one alone.

Notably, our sample was predominantly White, which may limit generalizability and interpretation of some of our findings (particularly those involving race). While this introduces a limitation in some respects, it (unfortunately) does reflect the current reality of jury composition in the USA (e.g., Collins & Gialopsos, 2021), which demonstrates overrepresentation of White jurors.

Finally, it would have been ideal to compare online participation to in-person participation, as has been done in previous research (Jones et al., 2022; Maeder et al., 2018), in order to determine if modality might moderate the effects of data collection type. Future research should probe whether these methodological considerations interact in order to provide jury researchers with a more fulsome sense of best practices.

Findings from the current study are encouraging with respect to the continued use of trial transcripts in juror simulation research. Of course, even if trial transcripts and videotapes are equally “effective” in terms of experimental manipulations and the small number of outcome measures we included here, this does not mean that they will be viewed as such by courts (Diamond, 1997). The limitations of juror simulation studies, especially those collected online, pose serious threats to the perceived (and real) validity of findings. More work is needed to interrogate MTurk and other online platform quality assurances.

## Conclusion

A written stimulus may yield some slight differences in perceptions of trial parties as compared to a videotaped stimulus, but the two modalities do not appear to yield significantly different verdict outcomes. Regardless, researchers should consider the drawbacks of mock trial videos for online methodologies, especially the distractibility they might foster. Future work should compare video trial versus transcript stimuli in online versus in-person samples, as well as for different online recruitment platforms and case types.

**Funding** This study was funded by a Social Sciences and Humanities Research Council (SSHRC) Insight Grant awarded to the first author.

**Data availability** The data that support the findings of this study are available from the corresponding author upon reasonable request.

## References

- Baker, M. A., Fox, P., & Wingrove, T. (2016). Crowdsourcing as a forensic psychology research tool. *American Journal of Forensic Psychology*, *34*(1), 37–50.
- Bornstein, B. H. (1999). The ecological validity of jury simulations: Is the jury still out? *Law and Human Behavior*, *23*(1), 75–91. <https://doi.org/10.1023/A:1022326807441>
- Bornstein, B. H. (2017). Jury simulation research: Pros, cons, trends, and alternatives. In M.B. Kovera (Ed.), *The psychology of juries* (pp. 207–226). Washington, DC: American Psychological Association. <https://doi.org/10.1037/0000026-010>
- Chaiken, S., & Eagly, A. H. (1976). Communication modality as a determinant of message persuasiveness and message comprehensibility. *Journal of Personality and Social Psychology*, *34*(4), 605–614. <https://doi.org/10.1037/0022-3514.34.4.605>
- Chaiken, S., & Eagly, A. H. (1983). Communication modality as a determinant of persuasion: The role of communicator salience. *Journal of Personality and Social Psychology*, *45*(2), 241–256. <https://doi.org/10.1037/0022-3514.45.2.241>
- Chandler, J., Mueller, P., & Paolacci, G. (2014). Nonnaïveté among amazon mechanical turk workers: Consequences and solutions for behavioral researchers. *Behavior Research Methods*, *46*(1), 112–130. <https://doi.org/10.3758/s13428-013-0365-7>
- Clark, J. M., & Paivio, A. (1991). Dual coding theory and education. *Educational Psychology Review*, *3*(3), 149–210. <https://doi.org/10.1007/BF01320076>
- Collins, P. A., & Gialopos, B. M. (2021). Answering the call: An analysis of jury pool representation in Washington State. *Criminology, Criminal Justice, Law, & Society*, *22*(1), 36–49.
- Coplan, A. (2006). Catching characters emotions: Emotional contagion responses to narrative fiction film. *Film Studies*, *8*(1), 26–38. <https://doi.org/10.7227/FS.8.5>
- Detenber, B. H., Simons, R. F., & Bennett, G. G., Jr. (1998). Roll ‘em!: The effects of picture motion on emotional responses. *Journal of Broadcasting & Electronic Media*, *42*(1), 113–127. <https://doi.org/10.1080/08838159809364437>
- Diamond, S. S. (1997). Illuminations and shadows from jury simulations. *Law and Human Behavior*, *21*, 561–571. <https://doi.org/10.1023/A:1024831908377>
- Fielder, T., Watts, F., Howden, C., Gupta, R., & McKenzie, C. (2022). Why choose a pathology career? A survey of Australian medical students, junior doctors, and pathologists. *Archives of Pathology & Laboratory Medicine*, *146*(7), 903–910.
- Fishfader, V. L., Howells, G. N., Katz, R. C., & Teresi, P. S. (1996). Evidential and extralegal factors in juror decisions: Presentation mode, retention, and level of emotionality. *Law and Human Behavior*, *20*(5), 565–572. <https://doi.org/10.1007/BF01499042>
- Hatfield, E., Cacioppo, J. T., & Rapson, R. L. (1993). Emotional contagion. *Current Directions in Psychological Science*, *2*(3), 96–100.
- Heath, W. P., Grannemann, B. D., & Peacock, M. A. (2004). How the defendant’s emotion level affects mock jurors’ decisions when presentation mode and evidence strength are varied. *Journal of Applied Social Psychology*, *34*(3), 624–664. <https://doi.org/10.1111/j.1559-1816.2004.tb02563.x>
- Iyer, A., & Oldmeadow, J. (2006). Picture this: Emotional and political responses to photographs of the Kenneth Bigley kidnapping. *European Journal of Social Psychology*, *36*(5), 635–647. <https://doi.org/10.1002/ejsp.316>
- Joffe, H. (2008). The power of visual material: Persuasion, emotion and identification. *Diogenes*, *55*(1), 84–93. <https://doi.org/10.1177/0392192107087919>
- Jones, A. M., Wong, K. A., Meyers, C. N., & Ruva, C. (2022). Trial by tabloid: Can implicit bias education reduce pretrial publicity bias? *Criminal Justice and Behavior*, *49*(2), 259–278.
- Juhnke, R., Vought, C., Pyszczynski, T. A., Dane, F. C., Losure, B. D., & Wrightsman, L. S. (1979). Effects of presentation mode upon mock jurors’ reactions to a trial. *Personality and Social Psychology Bulletin*, *5*(1), 36–39. <https://doi.org/10.1177/014616727900500107>
- Koehler, J. J., & Meixner, J. B., Jr. (2017). Jury simulation goals. In M. B. Kovera (Ed.), *The psychology of juries* (pp. 161–183). American Psychological Association. <https://doi.org/10.1037/0000026-008>
- Krauss, D. A., & Lieberman, J. D. (2017). Managing different aspects of validity in trial simulation research. In M. B. Kovera (Ed.), *The psychology of juries* (pp. 185–205). American Psychological Association. <https://doi.org/10.1037/0000026-009>

- Lassiter, G. D., Slaw, R. D., Briggs, M. A., & Scanlan, C. R. (1992). The potential for bias in videotaped confessions. *Journal of Applied Social Psychology*, 22(23), 1838–1851. <https://doi.org/10.1111/j.1559-1816.1992.tb00980.x>
- Lieberman, J. D., Krauss, D. A., Heen, M., & Sakiyama, M. (2016). The good, the bad, and the ugly: Professional perceptions of jury decision-making research practices. *Behavioral Sciences and the Law*, 34, 495–514. <https://doi.org/10.1002/bsl.2246>
- Maeder, E. M., Yamamoto, S., & McManus, L. A. (2018). Methodology matters: Comparing sample types and data collection methods in a juror decision-making study on the influence of defendant race. *Psychology, Crime & Law*. <https://doi.org/10.1080/1068316X.2017.1409895>
- Paolacci, G., Chandler, J., & Iperiotis, P. (2010). Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making*, 5(5), 411–419.
- Pezdek, K., Avila-Mora, E., & Sperry, K. (2010). Does trial presentation medium matter in jury simulation research? Evaluating the effectiveness of eyewitness expert testimony. *Applied Cognitive Psychology*, 24(5), 673–690. <https://doi.org/10.1002/acp.1578>
- Pfau, M., Haigh, M., Fifrick, A., Holl, D., Tedesco, A., Cope, J., ... Roszkowski, P. (2006). The effects of print news photographs of the casualties of war. *Journalism & Mass Communication Quarterly*, 83(1), 150–168. <https://doi.org/10.1177/107769900608300110>
- R. v. Gayle (1994), Crim LR 679, CA
- Rose, M. R., Diamond, S. S., & Baker, K. M. (2010). Goffman on the jury: Real jurors' attention to the "offstage" of trials. *Law and Human Behavior*, 34(4), 310–323. <https://doi.org/10.1007/s10979-009-9195-7>
- Sleed, M., Durrheim, K., Kriel, A., Solomon, V., & Baxter, V. (2002). The effectiveness of the vignette methodology: A comparison of written and video vignettes in eliciting responses about date rape. *South African Journal of Psychology*, 32(3), 21–28. <https://doi.org/10.1177/008124630203200304>
- Sommers, S. R., & Ellsworth, P. C. (2001). White juror bias: An investigation of prejudice against Black defendants in the American courtroom. *Psychology, Public Policy, and Law*, 7(1), 201–229. <https://doi.org/10.1037/1076-8971.7.1.201>
- Sommers, S. R., & Ellsworth, P. C. (2009). "Race salience" in juror decision-making: Misconceptions, clarifications, and unanswered questions. *Behavioral Sciences & the Law*, 27(4), 599–609. <https://doi.org/10.1002/bsl.877>
- Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, 54(6), 1063–1070.
- Williams, G. R., Farmer, L. C., Lee, R. E., Cundick, B. R., Howell, R. J., & Rooker, C. K. (1975). Juror perceptions of trial testimony as a function of the method of presentation. *Brigham Young University Law Review*, 2, 375–421.
- Wilson, J. R. (1996). *Putting the trial back into pretrial publicity: The medium is the message*. Unpublished master's thesis, Louisiana State University, Baton Rouge.
- Yadav, A., Phillips, M. M., Lundeberg, M. A., Koehler, M. J., Hilden, K., & Dirkin, K. H. (2011). If a picture is worth a thousand words is video worth a million? Differences in affective and cognitive processing of video and text cases. *Journal of Computing in Higher Education*, 23(1), 15–37. <https://doi.org/10.1007/s12528-011-9042-y>
- Zupan, B., & Babbage, D. R. (2017). Film clips and narrative text as subjective emotion elicitation techniques. *The Journal of Social Psychology*, 157(2), 194–210. <https://doi.org/10.1080/00224545.2016.1208138>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

**Evelyn M. Maeder** is a Professor in the Institute of Criminology and Criminal Justice at Carleton University. Her work investigates social psychological principles in the criminal legal system, including the effects of extralegal information on juror and jury decision-making, attitudes toward the insanity defense, and lay conceptions of punishment. She is the director of the Legal Decision-Making Lab at Carleton.

**Susan Yamamoto** is an Assistant Professor in the Psychology Department at Campion College/the University of Regina. Her current research in the Normative Ethics and Law Lab (NEL) focuses on punishment orientation, racial essentialism, and cultural differences in the courtroom. The lab's aim is to better understand basic social processes and help promote fairness in legal /political decisions.

**Logan Ewanation** is an Assistant Professor of Forensic Psychology in the Faculty of Social Sciences & Humanities at Ontario Tech University. His research primarily examines jurors' perceptions of forensic evidence and the role of racial bias in the courtroom.