Beliefs about error rates and human judgment in forensic science

Gianni Ribeiro[a], Jason M. Tangen[a], & Blake M. McKimmie[a]

[a]School of Psychology, The University of Queensland, Saint Lucia 4072, QLD Australia
g.ribeiro@uq.edu.au, jtangen@psy.uq.edu.au, b.mckimmie@psy.uq.edu.au

In press at *Forensic Science International* (2019)
DOI: 10.1016/j.forsciint.2019.01.034

Corresponding author:
Gianni Ribeiro
School of Psychology
The University of Queensland
Saint Lucia 4072
Queensland, Australia
g.ribeiro@uq.edu.au

Abstract

Forensic science techniques are often used in criminal trials to infer the identity of the perpetrator of crime and jurors often find this evidence very persuasive. Unfortunately, two of the leading causes of wrongful convictions are forensic science testing errors and false or misleading forensic testimony (Saks & Koehler, 2005). Therefore, it is important to understand jurors' pre-existing beliefs about forensic science, as these beliefs may impact how they evaluate forensic evidence in the courtroom. In this study, we examine people's perceptions of the likelihood of error and human judgment involved at each stage of the forensic science process (i.e., collection, storage, testing, analysis, reporting, and presenting). In addition, we examine people's perceptions of the accuracy of — and human judgment involved in — 16 different forensic techniques. We find that, in contrast to what would be expected by the CSI effect literature, participants believed that the process of forensic science involved considerable human judgment and was relatively error-prone. In addition, participants had wide-ranging beliefs about the accuracy of various forensic techniques, ranging from 65.18% (document analysis) up to 89.95% (DNA). For some forensic techniques, estimates were lower than that found in experimental proficiency studies, suggesting that our participants are more skeptical of certain forensic evidence than they need to be.

Keywords: Forensic science; forensic evidence; accuracy; error rate; CSI effect.

Introduction

In criminal cases, comparative forensic sciences are often used to infer the identity of the perpetrator. For example, a nuclear deoxyribonucleic acid (DNA) sample or fingerprint left at a crime scene may be compared to a known sample or print to determine whether they come from the same individual or not. Until recently, the validity and reliability of these forensic techniques had gone unquestioned and criminal courts have allowed examiners to testify that two prints or samples "match" to the exclusion of all other people despite no empirical basis for these conclusions (Edwards, 2009).

Understanding jurors' prior beliefs and perceptions about forensic science is necessary, as these beliefs may influence jurors' understanding of (and decisions about) forensic evidence presented at trial. The present study aims to build on previous literature examining beliefs about forensic evidence by exploring people's beliefs about error and human judgment involved at each stage of the forensic science process — from evidence collection, storage, and testing to reporting and presenting the evidence in court. To our knowledge, no prior research has investigated perceptions of human judgment involved in forensic science. Investigating people's perceptions of human judgement involved in forensic science may provide a richer account for their perceptions of error and forensic science overall. Furthermore, prior research has typically looked at error involved in forensic science as a whole; the current study is the first to look at perceptions of error and human judgment involved in each stage of the forensic science process.

Wrongful convictions on the basis of forensic science

In 2004, an American lawyer, Brandon Mayfield, was wrongfully accused of committing the Madrid Train Bombings that killed 192 people and injured another 2,050 (Harwood, 2014; Thompson & Cole, 2005). A fingerprint found on a bag of detonators was wrongly attributed to Mayfield's print, which was in the FBI's database due to his prior military service. Despite Mayfield being in the United States during the time of the bombings and with no other evidence to link him to the crime, three FBI fingerprint experts concluded that the two prints “matched.” Two weeks after Mayfield was arrested, the Spanish National Police informed the FBI that they had identified an Algerian man, Daoud Ouhnane, as the source of the fingerprint. The FBI then withdrew their identification of Mayfield and released him from custody.

Errors of this kind happen more often than people tend to think. Saks and Koehler (2005) analyzed 83 DNA exoneration cases and found that forensic science testing errors occurred in 63% of cases and false or misleading forensic testimony occurred in 27% of cases. Furthermore, The Innocence Project found that misapplication of forensic science occurred in nearly half (46%) of all DNA exoneration cases (The Innocence Project, 2017). In reality, it is impossible to know just how many wrongful convictions have occurred on the basis of forensic evidence, particularly because the processes for preserving and maintaining such evidence have not been mandatory (Lawson, 2014).

In response to these wrongful convictions, the U.S. National Academy of Sciences (NAS) issued a report heavily criticizing the current state of forensic science, concluding that “with the exception of nuclear DNA analysis … no forensic method has been rigorously shown to have the capacity to consistently, and with a high degree of certainty, demonstrate a connection between evidence and a specific individual or source” (2009, p. 7). The NAS report recommended that research be conducted to establish the reliability and limits of performance for each technique, as this research is lacking in most of the forensic disciplines (NAS, 2009). Most recently, the U.S. President’s Council of Advisors on Science and Technology (PCAST) issued a report questioning the scientific validity and reliability of various feature-comparison methods (2016) and made recommendations to strengthen forensic science and promote rigor.

How reliable is forensic evidence?

Prior to these reports, the validity and reliability of forensic science techniques had largely gone unquestioned. As the forensic sciences encompass a broad range of disciplines, it should not be surprising that there is also a great deal of variability in terms of methodology, reliability, error rates, and evidence-based practice. Calls from the NAS and PCAST have helped to kickstart research into accuracy and error rates of forensic sciences, however it is still largely unknown for most techniques. Here we will discuss three key forensic techniques — DNA, fingerprint analysis, and bite mark analysis.

DNA. DNA is considered the gold standard of forensic techniques due to its impressively small random match probabilities, suggesting that errors are extremely unlikely (Thompson, 2013). However, even DNA evidence is not infallible; contamination or mislabeling during collection, handling, and testing can lead to incorrect results (Lieberman, Carrell, Miethe, & Krauss, 2008) and samples containing multiple sources require more subjective judgment, leaving the potential for human error or bias (Dror & Hampikian, 2011). However, no proper validation experiments have been conducted to determine these error rates (Thompson, 2013).

Fingerprints. For more than 100 years, fingerprint experts have claimed that they can make a positive identification to the exclusion of all other persons — with some examiners even claiming that identification is infallible (Cole, 2001; Neumann, 2012). Tangen, Thompson, and McCarthy (2011) conducted the first study comparing the performance of fingerprint examiners to novices using ground-truth stimuli. They demonstrated that fingerprint examiners possess genuine expertise, performing far better than novices by making only 0.68% false positive decisions and 7.88% false negative decisions. Fingerprint experts are able to perform exceptionally well even under time constraints and with difficult visibility (Thompson & Tangen, 2014). However, studies have revealed that fingerprint examiners can disagree about the number of identifying features in fingerprints

(Ulery, Hicklin, Roberts, & Buscaglia, 2014) and are susceptible to contextual bias (Kassin, Dror, & Kukucka, 2013; Searston, Tangen, & Eva, 2016).

Bite marks. In contrast to DNA and fingerprint analysis, bite mark analysis is considered the most controversial of the forensic techniques (Kieser, 2005). Like fingerprint analysis, bite mark analysis relies on the assumption that each individual person has a unique bite mark. There are a number of methods that can be used to analyze bite marks, but there is no evidence for the reproducibility of any of these methods either between experts, or by the same expert at different times (NAS, 2009; Pretty & Sweet, 2001). As such, the PCAST (2016) report concluded that bite mark analysis does not meet the standards of scientific validity and reliability.

Why do prior beliefs about forensic science matter?

Despite wrongful convictions and a number of authoritative reports scrutinizing the current state of forensic science, the criminal justice system has continued to admit forensic evidence that is unreliable, unvalidated, and untested. Allowing forensic evidence into court that has not been empirically tested perpetuates the faulty assumption that all forensic sciences are valid and should be given considerable evidentiary value (Cole, 2010). As a result, jurors are not necessarily aware of the challenges and controversies the forensic sciences face (Lawson, 2014).

If jurors are not aware of the controversies surrounding forensic science, they are likely to interpret forensic testimony through the lens of the knowledge and beliefs they already have about forensic science prior to entering the courtroom. The Story Model of jury decision-making suggests that jurors may incorporate evidence presented to them at trial with their pre-existing general knowledge to form a narrative representation of the evidence (Pennington & Hastie, 1986, 1992). Despite being provided with the same testimony at trial, individual jurors may construct different narratives from one another and perhaps even reach a different verdict. If jurors' beliefs about forensic science are strongly held, they could have difficulty setting aside these beliefs to evaluate the evidence Lawson, 2014). Thus, it is important to identify jurors' prior beliefs, particularly misbeliefs, about forensic science in order to minimize their effect prior to trial and help improve jurors' ability to evaluate expert testimony.

Although many researchers have acknowledged that jurors' beliefs about forensic science may impact their evaluation of the evidence (Garrett & Mitchell, 2013; Schklar & Diamond, 1999; Smith, Bull, & Holliday, 2011), little research has directly explored these beliefs. Arguably, the largest body of literature that has explored beliefs about forensic science, albeit indirectly, is research into the CSI effect.

The CSI Effect

Popular television crime series like *CSI: Crime Scene Investigation* (CSI), *Law & Order* and most recently *True Detective* tend to portray forensic science in an over-exaggerated fashion — using high-tech equipment to solve crimes in a matter of hours, even minutes, without error. The following exchange between CSI's main characters helps to illustrate the faith placed in forensic science on television (Fink, 2004):

> Catherine Willows: The evidence is wrong.
>
> Gil Grissom: No, it isn't. You can be wrong. I can be wrong. The evidence is just the evidence.

This exchange depicts a view of forensic science that can speak for itself, free from any human involvement or error (Harvey & Derksen, 2009). But, as several authoritative reports have now explained, this is not the case for forensic science in the real world. According to one forensic scientist, around 40 percent of forensic science on the show is completely made up, with the remainder conducted quickly and effortlessly in ways that real forensic laboratories could only dream of (Cole & Dioso-Villa, 2009).

Ward (2014) suggests that forensic science is misrepresented in three distinct ways: division of labor, the facilities and equipment used, and the ease of solving cases. In crime television, forensic analyses are typically conducted by a single technician in an in-house laboratory, just a stroll away from the detective's own office. Despite their large workload, the forensic examiner appears to manage seemingly without much difficulty (Ley, Jankowski, & Brewer, 2012). The forensic laboratories are often portrayed as being equipped with state of the art technology and equipment (Robbers, 2008). Examiners conduct their analyses quickly and effortlessly, often in a matter of hours ready for criminal trial the next day (Ley et al., 2012). Throughout this process, the examiners rarely make a mistake. In a content analysis of the first seasons of *CSI* and *CSI: Miami*, Smith, Patry, and Stinson (2007) found that errors were very rare and only ever occurred before any serious consequences, such as a misidentification.

The effect that these misrepresentations of forensic science may have on jurors' beliefs and perceptions is referred to as the "CSI effect." Scholars and legal professionals typically point to the CSI effect literature to demonstrate that beliefs about forensic science can impact the outcomes of criminal trials (Chin & Workewych, 2016). The CSI effect is thought to influence jury verdicts by either: a) burdening the prosecution due to jurors' unrealistically high expectations that forensic evidence is available and necessary, resulting in higher rates of acquittal when forensic evidence is not present and making it more difficult for prosecutors to win convictions, or b) by burdening the defense due to unrealistically high faith in the accuracy and reliability of forensic science, resulting in higher rates of conviction when forensic evidence is present (Schweitzer & Saks, 2007). These two possible outcomes are often referred to as the pro-defense and pro-prosecution biases.

In response to anecdotal claims in the CSI effect literature, Smith and Bull (2012, 2014) developed a new scale, the Forensic Evidence Evaluation Bias Scale (FEEBS), to assess jurors' pre-trial bias towards forensic evidence. In line with the CSI effect literature, principal components analysis revealed two distinct constructs in the scale relating to pro-defense (example item: “no forensics means investigators did not look hard enough”) and pro-prosecution biases (example item: “forensic evidence is enough to convict”). They found that the pro-prosecution subscale of the FEEBS was positively correlated with other juror bias measures: the Juror Bias Scale (Kassin & Wrightsman, 1983) and the General Belief in a Just World scale (Dalbert, Montada, & Schmitt, 1987). They also found that the FEEBS pro-prosecution subscale significantly predicted participants’ perceived strength of DNA evidence in a mock trial, demonstrating support for the scale’s ability to tap into jurors’ bias towards forensic evidence.

Anecdotal evidence from legal professionals indicates that they believe the CSI effect to be a genuine effect (Stinson, Patry, & Smith, 2007; Tyler, 2006), however empirical research has produced mixed results supporting the nature of the CSI effect (Baskin & Sommers, 2010; Schweitzer & Saks, 2007). One potential explanation for these mixed results is that measuring participants’ crime show viewing habits does not adequately capture their beliefs, perceptions, and attitudes towards forensic science and forensic evidence. Pre-existing attitudes and beliefs have been shown to influence verdict preferences in a number of studies — from rape cases (Kovera, 2002), capital offences (Bowers, Sandys, & Stiener, 1998), and civil litigation (Vinson, Costanzo, & Berger, 2008) to the insanity defense (Skeem, Louden, & Evans, 2004). However, the CSI effect literature essentially uses crime show viewing as a proxy measure for beliefs and perceptions about forensic science; it makes the assumption that people who view crime shows will have different beliefs and perceptions about forensic science than those who do not. No research to date has directly tested this assumption or investigated people’s beliefs and perceptions about forensic science and forensic evidence. Further, at present it is also unclear whether people blindly believe that all forensic sciences are accurate or, rather, whether people believe that some techniques are more accurate than others.

Directly assessing beliefs about forensic science

Only a handful of studies have directly assessed jurors' beliefs about forensic science. These studies suggest that people do hold beliefs about the reliability and validity of different forensic techniques, which tend to be overestimated (Hans, Kaye, Dann, Farley, & Albertson, 2011; Lawson, 2014; Lieberman et al., 2008).

Hans and colleagues (2011) asked a sample of jury pool members to rate the reliability of DNA evidence alongside expert witness evidence, police evidence, victims' evidence, and eyewitness evidence. They found that DNA evidence was thought to be far more reliable than the non-scientific evidence types, with the majority (95%) of participants rating DNA evidence as extremely reliable or very reliable, compared to 66% for expert witness evidence, 67% for police evidence, 37% for victim evidence, and 25% for eyewitness evidence. In their first study, Lieberman and colleagues (2008) directly compared students and jurors' beliefs about the reliability of DNA, fingerprint, and hair/fiber evidence. DNA evidence was considered most reliable (94-95%), followed by fingerprints (90-91%) and then hair and fiber evidence (88-89%). Lieberman and colleagues (2008) also found that greater pre-trial trust in DNA evidence significantly predicted a guilty verdict, suggesting that prior beliefs about forensic science can influence trial outcomes.

Similarly, Lawson (2014) asked participants to rate the reliability of DNA, fingerprint, tool mark and bite mark evidence, and found that DNA evidence was considered to be the most reliable, followed by fingerprint, bite mark, and tool mark evidence. While these studies provide some initial understanding of jurors' beliefs about forensic science, they only scratch the surface. Firstly, these studies have only assessed a few of the many different forensic techniques. The current study aims to build on this previous literature by examining beliefs about a wide range of forensic techniques. Furthermore, these studies only ask participants to provide an overall judgment about the reliability of the technique and have not investigated their beliefs about what happens throughout the forensic science process. Thus, the current study will investigate beliefs about error and human judgment throughout each stage of the forensic science process.

Specific versus global error in forensic science

One of the benefits of investigating people's beliefs about error at specific stages throughout the forensic science process is that we can determine whether their overall impressions (i.e., beliefs about accuracy) of forensic techniques reflects their beliefs about errors that occur at each stage. As most people have difficulty understanding and combining probabilistic information (Gilovich, Vallone, & Tversky, 1985; Hoffrage, Lindsey, Hertwig, & Gigerenzer, 2000; Mellers & McGraw, 1999), it is possible that people might fall prey to base-rate neglect when assessing the overall accuracy of various forensic techniques.

Base-rate neglect (also known as the base-rate fallacy) happens when people ignore relevant base-rate information (i.e., general information) in favor of more specific, but irrelevant information (i.e., information relating to a particular case; Kahneman & Tversky, 1973; Pennycook & Thompson, 2017). One reason why we tend to neglect base rates is due to the representativeness heuristic (Kahneman & Tversky, 1972), where we make intuitive judgments of probability based on how similar to (or representative of) a prototype. In Kahneman and Tversky's (1973) well-known engineer-lawyer problem, participants disregarded the base rates of engineers and lawyers in favor of the description of the individual; even though the base rate suggests that the individual is more likely to be a lawyer, the description tended to fit that of a prototypical engineer rather than a lawyer.

No prior research in the forensic science field has examined how people arrive at estimates of accuracy or error. Coming to an accurate estimate of the overall accuracy of forensic techniques would require participants to override their intuitive judgments and appropriately combine the base rates of error at each stage of the forensic science process. However, it is possible that participants may fall prey to base-rate neglect and the representativeness heuristic when making estimates of accuracy for forensic techniques. Therefore, their global estimates of accuracy for the sixteen forensic techniques are unlikely to be informed by their more specific estimates of error at each stage of the process.

Current Research

The purpose of this exploratory study is to identify the beliefs and perceptions people hold about forensic science and forensic evidence. To build on the previous literature, the current study will investigate beliefs about error rates and the degree of human judgment involved in forensic science. Furthermore, the study will not only examine these aspects for each forensic technique, but will identify beliefs about error and human judgment at each stage of the forensic science process including: collection, storage, testing, analysis, reporting, and presenting the evidence. Crime television shows tend to depict forensic evidence at different stages, therefore this study aims to determine people's perceptions of error and human judgment at each of these stages. In doing so, we can also determine how people's estimates about the overall accuracy of forensic techniques relate to judgments of error at each stage of the process.

In this study, participants will be asked to think about forensic science generally, freely describe their opinion of what happens at each stage of the forensic science process, provide an estimate for the level of human involvement, judge how likely it is for an error to occur at each stage of the process, and then rate the accuracy and level of human involvement for sixteen different forensic techniques.

Hypotheses. H1: We expect that estimates of error for each stage of the forensic science process would be low (i.e., less than 5%; H1a) and estimates of human judgment involved in each stage of the forensic science process would also be low (i.e., a mean value below the mid-point of 4; H1b). We chose 5% as the threshold for low error because in most scientific domains we generally accept an alpha level, or the probability of rejecting a true null hypothesis, to be 5% ($\alpha = .05$). We also expect that there will be significant positive correlations between estimates of error and estimates of human judgment involved at each stage of the forensic science process (H1c).

H2: If the CSI effect is robust, we expect that participants' crime show viewing would be negatively correlated with estimates of error for each stage of the forensic process and with their estimates of human judgment for each stage of the forensic process.

H3: For the individual techniques, we expect that estimates of accuracy for each technique would be high (i.e., more than 90%; H3a) and estimates of human judgment involved in each forensic technique would be low (i.e., a mean value below the mid-point of 4; H3b). We also expect that there will be significant negative correlations between estimates of accuracy and estimates of human involvement for each forensic technique (H3c).

H4: If the CSI effect is robust, we expect that participants' crime show viewing would be positively correlated with estimates of accuracy for each forensic technique, but negatively correlated with estimates of human judgment involved in each forensic technique.

H5: Finally, in line with previous research investigating base-rate neglect and the representativeness heuristic, we expect that participants will be unlikely to consider the previous base-rates they provided about likelihood of an error occurring at each stage of the forensic science process when making their estimates of the overall accuracy of different types of forensic evidence. That is, the cumulative error for all stages of the forensic science process is likely to sum to more than 100.

Method

Participants

One hundred and one (52 female, 49 male) Australian participants with an age range of 20 - 70 years (*M* = 55.25, *SD* = 12.24) participated in this single condition exploratory study. Participants were recruited through an Australian market research company in December 2015 and remunerated AUD$5.95 for their participation in the study. The majority of participants (83.2%) had completed Year 12 high school or equivalent. 19.8% of participants' highest qualification was a postgraduate university degree, compared to an undergraduate university degree (34.7%), or vocational training (e.g., diploma, 45.5%). 41.6% of participants had taken advanced science or mathematics in high school and 42.6% had taken advanced science or mathematics after school. Sixteen participants (15.8%) had completed jury service before.

Procedure and Measures

The study was administered using the *Qualtrics* survey software and participants completed the study on their own computers or electronic devices. After agreeing to participate in the study, participants were given the following instructions at the beginning of the study: "During this study, you'll be asked to imagine that a crime has taken place and forensic evidence has been left at the scene of the crime. Imagine that police have charged a suspect with the crime, and there will be a criminal trial in front of a jury. We want you to think about the entire process involving forensic evidence - from when a crime scene is first attended, to when evidence is analyzed, to when it is presented in court to the jury. Please be as specific and detailed with your responses as you possibly can." Participants then completed an attention check question to ensure that they had read the instructions correctly and were aware of what the study was about. All participants correctly answered the attention check question. A copy of the *Qualtrics* survey, raw dataset, and *SPSS* syntax for data analyses are available on the Open Science Framework (OSF): https://osf.io/em4gh/

Stages of the Forensic Evidence Process

Describing the process. We asked participants to describe what happens at each stage of the forensic science process, including collection, storage, testing, analysis, reporting, and presenting. Participants were provided with a brief description of what each stage entailed for clarification (see Appendix A). Participants were provided with a text box with which to respond to each question and were not constrained by how little or how much they were required to write. Participants were asked: How is forensic evidence collected from a crime scene? How is forensic evidence stored? How is forensic evidence tested? How are the results of forensic testing analyzed and interpreted? How is the forensic evidence reported? How is the forensic evidence presented to the jury? This qualitative data will not be reported in this paper, however it is available on the OSF: https://osf.io/em4gh/

Estimates of error rate. For each stage of the forensic process, we asked participants "how likely is it that an error could occur during this process?" on a scale from 0% to 100%. Participants were provided with a sliding scale to respond, where the default position was set to 50%.

Estimates of human judgment. For each stage of the forensic process, we asked participants "to what extent does this process involve human judgment?" on a 7-point Likert scale ranging from 1 (not at all) to 7 (completely).

Forensic Techniques

Estimates of accuracy. We asked participants about the accuracy of 16 types of forensic evidence: anthropological analysis (i.e., human remains), bloodstain pattern analysis, DNA analysis, document analysis (i.e., handwriting), facial analysis (i.e., CCTV footage), fingerprint analysis, fire and explosives analysis, firearm and tool marks analysis, geological materials analysis, gunshot residue analysis, image analysis (i.e., photography), materials analysis (i.e., fibers, paint, glass), toxicology analysis (i.e., urine, drugs), voice analysis, wildlife analysis (i.e., plants, animals), and dental analysis. We asked participants "what are your overall impressions about the accuracy of the different types of forensic evidence listed below?" and participants responded on a scale from 0% to 100%, where the default slider position was set to 50%.

Estimates of human judgment. We then asked participants "to what extent does each of the following types of forensic evidence involve human judgment?" on a 7-point Likert scale from 1 (not at all) to 7 (completely) for the same 16 techniques above.

Crime Show Viewing Habits

We asked participants two questions related to their television viewing habits. Firstly, we asked how many hours a week on average they spent watching the following shows: *CSI* (including New York, Miami, and Cyber), *Law & Order* (including Criminal Intent, Special Victims Unit, Trial by Jury, and LA), *Criminal Minds*, *Bones*, *NCIS*, and other crime shows featuring forensic evidence. Participants were provided with a sliding scale to respond, ranging from 0 to 10 where the default position was set to 0. Secondly, we asked participants how many hours a week on average they watched any type of show on television or online. Participants were provided with a sliding scale to respond, ranging from 0 to 50 where the default position was set to 0.

Results

Crime show Viewing Habits

The average number of hours per week that people spent watching crime shows ranged from .74 (*Bones*) to 1.46 (other crime shows featuring forensic evidence). Means and standard deviations for each forensic-related television show, as well as the average total of all crime shows, are depicted in Table 1 below. We will use the average total of all crime shows to conduct the correlations that follow. The average number of hours per week spent watching shows on television or online was 16.78 (*SD* = 11.99).

[TABLE 1 HERE]

Stages of the Forensic Science Process

Estimates of error. We hypothesized in H1a that estimates of error for each stage of the forensic science process would be low (i.e., less than 5%). However, we found that participants' estimates of the chance of error occurring at each stage of the forensic science process ranged from 39.27% (testing stage) up to 44.55% (analysis stage). Single sample t-tests revealed that the estimated error was significantly higher than 5% for each stage (all *p*s < .001). The means, standard deviations and t-values for each stage of the process are depicted in Table 2 below. As such, our predictions in H1a were not supported. These estimates of error are much higher than one would expect according to the CSI effect, which otherwise suggests that people place unrealistically high faith in the accuracy and reliability of forensic science (Schweitzer & Saks, 2007).

Estimates of human judgment. We hypothesized in H1b that participants' estimates of human judgment involved at each stage of the forensic science process would be low (i.e., a mean value below the mid-point of 4). However, we found that participants' estimate of human judgment involved in each stage of the forensic science process ranged from 4.94 (testing stage) up to 5.55 (collection and presenting stages). Single sample t-tests revealed that the estimate of human judgment involved was significantly higher than 4 for each stage (all *p*s < .001). The means, standard deviations, and t-values for each stage of the process are depicted in Table 2 below. Thus, participants believed that there was a substantial level of human judgment involved at each stage of the process and our predictions in H1b were not supported. These estimates were higher than one would expect according to the CSI effect and cultivation theory which would suggest that people are likely to hold beliefs and perceptions about forensic science that are consistent with what is presented to them on screen (Gerbner, Gross, Morgan, & Signorielli, 1982; Morgan & Shanahan, 2010).

[TABLE 2 HERE]

Correlations between error and human judgment. We hypothesized in H1c that estimates of error for each stage of the forensic science process would be significantly positively correlated with estimates of human judgment involved for each stage of the forensic science process. We found that Pearson's correlations (*r*) between estimates of error and estimates of human judgment at each stage of the forensic science process were all significantly positively correlated (all *p*s < .01; correlation coefficients for each stage are depicted in Table 2 above). Thus, for all stages, the more participants believed that human judgment was involved, the higher they estimated the likelihood of an error occurring. These results support our hypothesis, H1c, that estimates of error for each stage would be positively correlated with estimates of human judgment involved at each stage.

Correlations between crime show viewing habits and error and human judgment. We hypothesized in H2 that participants' crime show viewing habits would be significantly negatively correlated with estimates of error and estimates of human judgment for each stage of the forensic science process. However, we found that there were no significant correlations between crime show viewing habits and estimates of error (all *p*s > .05). Further, we found that crime show viewing habits were only significantly negatively correlated with the storage stage (*r* = .233) and the testing stage (*r* = .224). Correlation coefficients for all stages are depicted in Table 2 above. Thus, our hypothesis, H2, was not supported. These results do not provide support for the CSI effect theory, which posits that people's impressions of forensic evidence are informed by their crime show viewing habits.

Forensic Techniques

Estimates of accuracy. We hypothesized in H3a that the estimates of accuracy for each technique would be high (i.e., more than 90%). We found that participants' estimates of accuracy for each forensic technique ranged from 65.18% (document) up to 89.95% (DNA). Single sample t-tests revealed that estimates of accuracy were significantly lower than 90% for most forensic techniques, with the exception of DNA ($t(99) = -0.03$, $p = .975$), fingerprint ($t(99) = -1.05$, $p = .295$), and dental analysis ($t(99) = -0.62$, $p = .537$). The means, standard deviations and *t*-values for each forensic technique are depicted in Table 3 below. Thus, our predictions and H3a were not supported as all forensic techniques were perceived to have an accuracy of less than 90%.

Estimates of human judgment. We hypothesized in H3b that estimates of human judgment involved in each forensic technique would be low (i.e., a mean value below the mid-point of 4). We found that participants' estimates of human judgment involved in each of the forensic techniques ranged from 3.48 (DNA) up to 5.72 (document). Single sample t-tests revealed that only DNA analysis ($t(99) = -2.72$, $p = .008$) had an estimated human judgment lower than 4, whilst estimates of human judgment for dental, fingerprint, geological materials, gunshot residue, materials, toxicology, and wildlife techniques were not significantly different from 4 (all ps > .05). The remaining techniques — anthropological, blood pattern, document, faces, fire/explosives, firearm/tools, image, and voice — had estimated human judgments significantly higher than 4 (all ps < .05). The means, standard deviations and *t*-values for each forensic technique are depicted in Table 3 below. Thus, we found mixed support for our hypothesis H3b.

[TABLE 3 HERE]

Correlations between accuracy and human judgment. We hypothesized in H3c that estimates of accuracy for each forensic technique would be significantly negatively correlated with estimates of human judgment involved in each forensic technique. Pearson's correlations between estimates of accuracy and estimates of human judgment at each stage of the forensic science process revealed that only 5 out of the 16 techniques were significantly negatively correlated (Blood pattern, DNA, document, fingerprint, and image; all *p*s $< .05$). Thus, for these evidence types, the more accurate participants estimated the forensic evidence to be, the less they believed that human judgment was involved. Correlation coefficients for each stage of the process are depicted in Table 3 below. As a result, support for our hypothesis H3c was mixed as there were only a few significant negative correlations between accuracy and human judgment.

Correlations between crime show viewing habits and accuracy and human judgment. We hypothesized in H4 that participants' crime show viewing habits would be positively correlated with their estimates of accuracy, but negatively correlated with estimates of human judgment for each forensic technique. However, we found that there were no significant correlations between crime show viewing habits and accuracy (all *p*s $>$ .05). Further, we found that there were three significant positive correlations between crime show viewing and human judgment of: DNA ($r = .212$), fingerprint ($r = .204$), and toxicology ($r = .248$), which are not in the expected direction. These results do not provide support for hypothesis, H4, based on the CSI effect, which posits that people's impressions of forensic evidence are informed by their crime show viewing habits.

Cumulative Error versus Global Error

We hypothesized that participants would be unlikely to take into account the previous base-rates they provided about the likelihood of an error occurring at each stage of the forensic science process. That is, the cumulative error for all stages of the forensic science process is likely to sum more than 100. On the other hand, we expected the error rates for each forensic technique to be low, which would demonstrate that participants are employing base-rate neglect.

Interestingly, by the second stage (storage), the cumulative error rate was already 81.83% and halfway through the process at the third stage (testing), the cumulative error rate was above 100% at 121.1%. This means that, at only halfway through the process, participants estimated that an error of some sort was inevitable. The cumulative total error rate at the sixth stage of the process (presenting) was 248.56% (see Table 2 for cumulative error rates).

In contrast, the error rates for each individual forensic technique ranged from 10.05% (DNA) to 34.82% (Document). Therefore, the highest error rate for any given forensic technique was still lower than the error rate provided for any one stage of the forensic science process (of which the lowest was 39.27% for the storage stage; see Table 3 for error rates).

These results supported our hypothesis, suggesting that our participants fell prey to base-rate neglect and the representativeness heuristic when making judgments about the accuracy of forensic techniques. That is, they completely disregarded their previous estimates about error rates at each stage of the forensic science process when asked to make judgments about the accuracy of specific forensic techniques. One limitation of our study is that when participants were asked to estimate the likelihood of an error at any stage of the forensic science process, they were not constrained to think about a specific type of forensic science. Thus, we cannot directly compare their beliefs about error during the forensic science process with a particular forensic technique. However, given that the highest error rate for any of the 16 forensic techniques was still lower than the error rate provided for any one stage of the forensic science process, we can reasonably conclude that their reasoning was fallacious and did not rely on base-rates.

Discussion

The aim of this study was to identify the beliefs and perceptions that people hold about forensic science and forensic evidence. Specifically, we investigated people's beliefs about error rate and the degree of human judgment involved in forensic science — not only for the different types of forensic science, but at each stage of the process (i.e., collection, storage, testing, analysis, reporting, and presenting). We also investigated people's beliefs about the accuracy and level of human judgment involved in sixteen different forensic techniques.

Surprisingly, we found that participants believed that there was a substantial likelihood of error at each stage of the forensic science process. One possible explanation for the higher than expected estimates of error is the novel approach to asking them about error. Perhaps asking them to think about each *specific* stage of the forensic science process got them to think more critically about what occurs at each stage and the possibility for error. It is possible that if we asked participants to provide us with an *overall* estimate of error for the forensic science process in its entirety, the estimate of error would not be as high as they would not think about each stage of the process as carefully. Another possible explanation for the higher than expected estimates of error is that the default slider position was set at 50%. Tversky and Kahneman's (1992) anchoring and adjustment research has demonstrated that different starting points (anchors) can result in different estimates and even irrelevant quantitative anchors can influence people's estimates. Therefore, it is possible that estimates of error could be different depending on the default slider position.

Furthermore, we found that participants believed that there was a considerable degree of human judgment involved in each stage of the forensic science. As forensic science is often portrayed as being very dependent on technology as opposed to human judgment, these estimates of human judgment involved at each stage of the process seem to contradict what would be expected by the CSI effect and cultivation theory literature. Again, these high estimates of human judgment could be due to the fact that we asked participants to think carefully about each stage of the forensic science process, rather than asking them to provide an estimate of human judgment involved in the process overall.

For each stage of the forensic science process, estimates of error were significantly positively correlated with estimates of human judgment, such that the more participants believed that human judgment was involved, the more they tended to estimate that an error would occur. Furthermore, we found that participants believed that there was variability in the accuracy of different forensic techniques — ranging from 65.18% for document analysis up to 89.95% for DNA analysis — and that there was a considerable degree of human judgment involved in most forensic techniques. However, we did not find a consistent correlation between accuracy and human judgment.

Finally, consistent with the base-rate neglect and representativeness heuristic literature (Kahneman & Tversky, 1973; Pennycook & Thompson, 2017), we found that participants' estimates of accuracy for individual forensic techniques was inconsistent with the base-rates of likelihood of an error occurring at each stage of the forensic science process they had just provided. These results demonstrate the importance of assessing both specific and global estimates of error in order to fully capture people’s beliefs about forensic science and forensic evidence.

Accuracy of Forensic Techniques

Although our hypothesis that accuracy of all forensic techniques would be more than 90% was not supported, participants did believe that all forensic techniques were significantly better than chance. This is somewhat unsurprising, as it is likely that participants believe that if a forensic technique exists and is used in criminal proceedings then it must have some degree of accuracy above chance. The forensic technique with the highest mean accuracy was DNA (89.95%) which is not surprising considering DNA is often touted as the “gold standard” for forensic science (Thompson, 2013).

Interestingly, the forensic technique that received the lowest mean accuracy (65.18%) was document analysis. Experimental studies suggest that forensic document examiners have a lower error rate than novices when examining documents with simulated handwriting (Kam, Abichandani, & Hewett, 2015; Kam, Gummadidala, Fielding, & Conn, 2001). Kam and colleagues (2001) found that when presented with two signatures that came from the same person, forensic document examiners incorrectly said that the signatures came from two different people 7.05% of the time, while novices made this incorrect conclusion 26.1% of the time. In contrast, when the two signatures came from two different people, forensic document examiners incorrectly said that the signatures came from the same person 0.49% of the time, while novices made this incorrect conclusion 6.47% of the time. Kam, Abichandani, and Hewett's (2015) subsequent study revealed even lower error rates for forensic document examiners. Thus, it is evident that participants believe that document examination is far less accurate than it is in reality.

Similarly, participants believed that fingerprint analysis is less accurate than experimental studies would suggest. Participants gave fingerprint analysis a mean accuracy rating of 88.15%, however controlled experiments suggests that fingerprint examiners’ accuracy is roughly 97% (Tangen et al., 2011; Ulery, Hicklin, Buscaglia, & Roberts, 2011). This result suggests that our participants are perhaps more skeptical about fingerprint analysis than they ought to be, however it is important to note that these error rates in lab-based experiments may not be truly representative of error rates in casework (American Academy for the Advancement of Science, 2017).

On the other hand, participants tended to think that forensic dentistry was very accurate (mean accuracy = 89.26%). This is in stark contrast to PCAST’s (2016) conclusion that, as the false positive error rates were so high, bite mark analysis does not meet the scientific standards of reliability and validity. However, it should be noted that a limitation of the current study is that participants were asked the rate the accuracy of forensic dental analysis as a whole, not bite mark analysis specifically so it is possible that participants were considering other dental analyses, such as identifying human remains. Similarly, participants believed that firearm and tool mark analysis is reliable (mean accuracy = 79.63%), despite PCAST's conclusion that firearm analysis "currently falls short of the criteria for foundational validity" as there is only one study that measured the validity and reliability of the technique. Thus, participants' belief in the accuracy of forensic dental analysis and firearm analysis is probably ill-founded. As little is known about the actual level of accuracy for many of the other forensic techniques, it is impossible to compare participants' beliefs about error rates with ground truth.

Is There Support for the CSI Effect?

Overall, our study provided very limited support for the CSI effect. The CSI effect assumes that those who view crime shows will have different beliefs and perceptions about forensic science than those who do not. If the CSI effect is robust, we would have expected that participants who watched more crime television shows would be: a) more likely to estimate that the chance of an error at each stage of the forensic science process is low, b) more likely to estimate that human judgment involved at each stage of the forensic science process is low, c) more likely to estimate that the accuracy of each individual forensic technique is high, and d) more likely to estimate that human judgment involved in each forensic science technique is low. However, we found that there were no correlations between crime show viewing and estimates of error for any stage of the forensic process, nor estimates of accuracy for any of the forensic techniques. Furthermore, there were a few small positive correlations between crime show viewing and estimates of human judgment involved storage and testing stages of the process, as well as for estimates of human judgment involved in DNA, fingerprint, and voice techniques; these correlations are in the opposite direction to our CSI-based hypotheses.

Furthermore, if people do form their impressions of forensic evidence through exposure to crime-related television shows, we would have expected to see very low estimates of error for each stage of the forensic process, as well as very high estimates of accuracy for the forensic techniques. However, this was not the case — our results demonstrated that, at each stage of the forensic science process, people thought that over one third of the time an error will occur. Additionally, people's estimates of the accuracy of various forensic techniques was lower than expected and they tended to believe that forensic techniques were relatively error prone.

Taken together, these results suggest that the more participants watch and are exposed to forensic science on television has no bearing on their beliefs about error, accuracy, and level of human judgment involved in forensic science. In turn, our results do not provide support for the CSI effect and instead demonstrate the value of *directly* assessing people's beliefs about forensic science and forensic evidence.

Importance of Assessing Beliefs and Perceptions of Forensic Science

Prior to this study, little research had directly assessed jurors' beliefs about forensic science. The leading theory in the literature, referred to as the CSI effect, posits that people's beliefs about forensic science are likely to come from what they see on crime television series such as *CSI* and *Law & Order* (Chin & Workewych, 2016). The results of these studies have inferred that people's beliefs about forensic science are based on how often they view crime shows, but these studies have never *directly* assessed beliefs about forensic science. Our study demonstrates the importance of directly assessing beliefs about forensic science, as our results do not support the assumption that people's beliefs about forensic science are informed by their crime show viewing habits. Instead, our participants believed that the forensic science process involved a considerable amount of human judgment and was relatively error prone. Furthermore, our participants had wide-ranging beliefs about the accuracy of different forensic techniques and their crime show viewing habits were not related to their estimates of accuracy. In turn, we believe that future studies should disregard the use of crime show viewing as proxy measure for people's beliefs about forensic science and the accuracy of different forensic techniques and should aim to measure these beliefs directly.

Regardless, it is important to note that although their crime show viewing habits were not related to their estimates of accuracy, our study demonstrates that people do not have a solid understanding of the accuracy of forensic techniques. Participants in the current study did have quite wide-ranging views on the accuracy of different forensic techniques, however some were overstated while others were understated. This finding demonstrates that jurors may come to trial with potentially inaccurate pre-existing beliefs about forensic science which may impact not only how they understand and evaluate the evidence, but potentially decisions about guilt or innocence based on that evidence.

The current study used a novel approach to investigating people's beliefs about forensic science by not only asking about accuracy of different techniques, but also perceptions of error that occur at each stage of the forensic process. Future studies could aim to replicate the methods of this study with a larger sample of jury-eligible participants in order to draw stronger inferences regarding people's beliefs about forensic science.

Conclusion

Forensic sciences are often used in criminal cases to infer the identity of a perpetrator, even when the reliability and validity of the forensic techniques are unknown. As a result, forensic science testing errors and false or misleading testimony are two of the most common reasons why individuals are falsely convicted of a crime (occurring in 63% and 27% of exoneration cases, respectively; Saks & Koehler, 2005). Therefore, it is important for us to understand people's beliefs about forensic science prior to entering the courtroom. In this study, we have demonstrated that participants do not just blindly believe that all forensic techniques are highly accurate, which has previously been assumed in the CSI effect literature. Instead, our participants believe that the forensic science process is error prone and involves a considerable amount of human judgment at each and every stage.

References

American Academy for the Advancement of Science. (2017). Forensic science assessments: A quality and gap analysis — Latent fingerprint examination. *Report prepared by W. Thompson, J. Black, A. Jain, & J. Kadane*. doi:10.1126/srhrl.

Baskin, D. R., & Sommers, I. B. (2010). Crime-show viewing habits and public attitudes towards forensic evidence: The "CSI effect" revisited. *justice System Journal, 31*(1), 97-113. doi:10.1080/0098261X.2010.10767956

Bowers, W. J., Sandys, M., & Stiener, B. D. (1998). Foreclosed impartiality in capital sentencing: Jurors' predispositions, guilt-trial experience, and premature decision making. *Cornell Law Review, 83*(6), 1476.

Chin, J. M., & Workewych, L. (2016). The CSI Effect. In M. Dubber (Ed.), *Oxford Handbooks Online*. New York: Oxford University Press.

Cole, S. A. (2001). *Suspect identities : a history of fingerprinting and criminal identification / Simon A. Cole*. Cambridge, MA: Harvard University Press.

Cole, S. A. (2010). Acculturating forensic science: what is 'scientific culture' and how can forensic science adopt it? *Fordham Urban Law Journal, 38*(2), 435. doi:10.2139/ssrn.1788414

Cole, S. A., & Dioso-Villa, R. (2009). Investigating the CSI effect: Media and litigation crisis in criminal law. *Stanford Law Review, 61*, 1335-1535.

Dalbert, C., Montada, L., & Schmitt, M. (1987). Belief in a just world: Media, juries, and the burden of proof. . *Psychologische Beitrage, 29*(1), 596-615.

Dror, I. E., & Hampikian, G. (2011). Subjectivity and bias in forensic DNA mixture interpretation. *Science & Justice, 51*(4), 204-208. doi:10.1016/j.scijus.2011.08.004

Edwards, H. T. (2009). Solving the problems that plague the forensic science community. *Jurimetrics, 50*(1), 5-19.

Fink, K. (Director). (2004). Bloodlines, *CSI: Crime Scene Investigation*.

Garrett, B., & Mitchell, G. (2013). How jurors evaluate fingerprint evidence: The relative importance of match language, method information, and error acknowledgment. *Journal of Empirical Legal Studies, 10*(3), 484-511. doi:10.1111/jels.12017

Gerbner, G., Gross, L., Morgan, M., & Signorielli, N. (1982). Charting the mainstream: Television's contributions to political orientations. *Journal of Communication, 32*(2), 100-127. doi:10.1111/j.1460-2466.1982.tb00500.x

Gilovich, T., Vallone, R., & Tversky, A. (1985). The hot hand in basketball: On the misperception of random sequences. *Cognitive psychology, 17*(3), 295-314.

Hans, V. P., Kaye, D. H., Dann, B. M., Farley, E. J., & Albertson, S. (2011). Science in the jury box: Jurors' comprehension of mitochondrial DNA evidence. *Law and Human Behavior, 35*(1), 60-71. doi:10.1007/s10979-010-9222-8.

Harvey, E., & Derksen, L. (2009). Science fiction or social fact?: An exploratory content analysis of popular press reports on the CSI effect. In M. Byers & V. M. Johnson (Eds.), *The CSI effect: Television, crime, and governance*. Lanham: Lexington Books.

Harwood, M. (2014). The terrifying surveillance case of Brandon Mayfield. Retrieved from http://america.aljazeera.com/opinions/2014/2/the-terrifying-surveillancecaseofbrandonmayfield.html

Hoffrage, U., Lindsey, S., Hertwig, R., & Gigerenzer, G. (2000). Communicating statistical information. *science, 290*(5500), 2261-2262.

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive psychology, 3*(3), 430-454.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review, 80*(4), 237.

Kam, M., Abichandani, P., & Hewett, T. (2015). Simulation detection in handwritten documents by forensic document examiners. *Journal of forensic sciences, 60*(4), 936-941. doi:10.1111/1556-4029.12801

Kam, M., Gummadidala, K., Fielding, G., & Conn, R. (2001). Signature authentication by forensic document examiners. *Journal of forensic sciences, 46*(4), 884-888. doi:10.1520/JFS15062J

Kassin, S. M., Dror, I., & Kukucka, J. (2013). The forensic confirmation bias: Problems, perspectives, and proposed solutions. *J. Appl. Res. Mem. Cogn., 2*(1), 42-52. doi:10.1016/j.jarmac.2013.01.001

Kassin, S. M., & Wrightsman, L. S. (1983). The construction and validation of a juror bias scale. *Journal of Research in Personality, 17*(4), 423-442. doi:10.1016/0092-6566(83)90070-3

Kieser, J. A. (2005). Weighing bitemark evidence: A postmodern perspective. *Forensic Science, Medicine, and Pathology, 1*(2), 75-80. doi:10.1385/ForensicSci.Med.Pathol.:1:2:75

Kovera, M. (2002). The effects of general pretrial publicity on juror decisions: An examination of moderators and mediating mechanisms. *Law and Human Behavior, 26*(1), 43-72. doi:10.1023/A:1013829224920

Lawson, V. (2014). *The influence of naïve and media-informed beliefs on juror evaluations of forensic science evidence.* Unpublished thesis.

Ley, B. L., Jankowski, N., & Brewer, P. R. (2012). Investigating CSI : Portrayals of DNA testing on a forensic crime show and their potential effects. *Public Understanding of Science, 21*(1), 51-67. doi:10.1177/0963662510367571

Lieberman, J. D., Carrell, C. A., Miethe, T. D., & Krauss, D. A. (2008). Gold versus platinum: Do jurors recognize the superiority and limitations of DNA evidence compared to other types of forensic evidence? *Psychology, Public Policy, and Law, 14*(1), 27-62. doi:10.1037/1076-8971.14.1.27

Mellers, B. A., & McGraw, A. P. (1999). How to improve Bayesian reasoning: Comment on Gigerenzer and Hoffrage (1995).(response to G. Gigerenzer et al., Psychological Review, vol. 102, p. 684, 1995). *Psychological Review, 106*(2), 417.

Morgan, M., & Shanahan, J. (2010). The state of cultivation. *Journal of Broadcasting & Electronic Media, 54*(2), 337-355. doi:10.1080/08838151003735018

National Academy of Sciences [NAS]. (2009). *Strengthening forensic science in the United States: A path forward*. Washington, DC: National Academies Press.

Neumann, C. (2012). Fingerprints at the crime-scene: Statistically certain, or probable? *Significance, 9*(1), 21-25. doi:10.1111/j.1740-9713.2012.00539.x

Pennington, N., & Hastie, R. (1986). Evidence evaluation in complex decision making. *Journal of Personality and Social Psychology, 51*(2), 242-258. doi:10.1037/0022-3514.51.2.242

Pennington, N., & Hastie, R. (1992). Explaining the evidence: Tests of the story model for juror decision making. *Journal of Personality and Social Psychology, 62*(2), 189-206. doi:10.1037/0022-3514.62.2.189

Pennycook, G., & Thompson, V. A. (2017). Base-rate neglect. In R. F. Pohl (Ed.), *Cognitive illusions: Intriguing phenomena in judgement, thinking and memory*. New York, NY: Routledge.

President's Council of Advisors on Science and Technology [PCAST]. (2016). *Forensic science in criminal courts: Ensuring scientific validity of feature-comparison methods*.

Pretty, I. A., & Sweet, D. (2001). The scientific basis for human bitemark analyses – a critical review. *Science & Justice, 41*(2), 85-92. doi:10.1016/S1355-0306(01)71859-X

Robbers, M. L. P. (2008). Blinded by science: The social construction of reality in forensic television shows and its effect on criminal jury trials. *Criminal Justice Policy Review, 19*(1), 84-102.

Saks, M., & Koehler, J. (2005). The coming paradigm shift in forensic identification science. *science, 309*, 892-895. doi:10.1126/science.1111565

Schklar, J., & Diamond, S. S. (1999). Juror reactions to DNA evidence: Errors and expectancies. *Law and Human Behavior, 23*(2), 159-184. doi:10.1023/A:1022368801333

Schweitzer, N. J., & Saks, M. J. (2007). The CSI effect: Popular fiction about forensic science affects the public's expectations about real forensic science. *Jurimetrics, 47*(3), 357-364.

Searston, R. A., Tangen, J. M., & Eva, K. W. (2016). Putting bias Into context: The role of familiarity in identification. *Law and Human Behavior, 40*(1), 50-64. doi:10.1037/lhb0000154

Skeem, J. L., Louden, J. E., & Evans, J. (2004). Venirepersons' attitudes toward the insanity defense: Developing, refining, and validating a scale. *Law and Human Behavior, 28*(6), 623-648. doi:10.1007/s10979-004-0487-7

Smith, L. L., & Bull, R. (2012). Identifying and measuring juror pre-trial bias for forensic evidence: development and validation of the Forensic Evidence Evaluation Bias Scale. *Psychology, Crime & Law, 18*(9), 797-815. doi:10.1080/1068316X.2011.561800

Smith, L. L., & Bull, R. (2014). Validation of the factor structure and predictive validity of the Forensic Evidence Evaluation Bias Scale for robbery and sexual assault trial scenarios. *Psychology, Crime & Law, 20*(5), 450-466. doi:10.1080/1068316X.2013.793340

Smith, L. L., Bull, R., & Holliday, R. (2011). Understanding juror perceptions of forensic evidence: Investigating the impact of case context on perceptions of forensic evidence strength. *Journal of forensic sciences, 56*(2), 409-414. doi:10.1111/j.1556-4029.2010.01671.x

Smith, S. M., Patry, M. W., & Stinson, V. (2007). But what is the CSI effect?: How crime dramas influence people's beliefs about forensic evidence. *Canadian Journal of Police and Security Services, 187*(1), 187-195.

Stinson, V., Patry, M. W., & Smith, L. L. (2007). The CSI effect: Reflections from police and forensic investigators. *Canadian Journal of Police and Security Services, 125*(1), 125-133.

Tangen, J. M., Thompson, M. B., & McCarthy, D. J. (2011). Identifying fingerprint expertise. *Psychological science, 22*(8), 995-997.

The Innocence Project. (2017). Op-ed: Department of justice is obstructing forensic science reform.

Thompson, M. B., & Tangen, J. M. (2014). The nature of expertise in fingerprint matching: experts can do a lot with a little. *PLoS ONE, 9*(12), e114759. doi:10.1371/journal.pone.0114759

Thompson, W. C. (2013). Forensic DNA evidence: The myth of infallibility. In J. Gruber & S. Krimsky (Eds.), *Genetic explanations : Sense and nonsense* (pp. 227-255). Cambridge, MA: Harvard University Press.

Thompson, W. C., & Cole, S. A. (2005). Lessons from the Brandon Mayfield case. *The Champion, 29*(3), 42-44.

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory - cumulative representation of uncertainty. *Journal of risk and uncertainty, 5*(4), 297-297.

Tyler, T. R. (2006). Viewing CSI and the threshold of guilt: Managing truth and justice in reality and fiction. *The Yale Law Journal, 115*(5), 1050-1085. doi:10.2307/20455645

Ulery, B. T., Hicklin, R. A., Buscaglia, J., & Roberts, M. A. (2011). Accuracy and reliability of forensic latent fingerprint decisions. *Proceedings of the National Academy of Sciences, 108*(19), 7733-7738. doi:10.1073/pnas.1018707108

Ulery, B. T., Hicklin, R. A., Roberts, M. A., & Buscaglia, J. (2014). Measuring what latent fingerprint examiners consider sufficient information for individualization determinations. *PLoS ONE, 9*(11).

Vinson, K. V., Costanzo, M. A., & Berger, P. D. D. E. (2008). Predictors of verdict and punitive damages in high-stakes civil litigation. *Behavioral Sciences & the Law, 26*(2), 167-186. doi:10.1002/bsl.807

Ward, K. A. (2014). *The repercussions of inaccurately portraying forensic science in television shows*. Unpublished thesis.

Table 1.

Means and standard deviations of number of hours per week spent watching crime television shows and television shows overall.

Television Show	Hours per week M(*SD*)
CSI	0.88 (1.81)
Law & Order	0.89 (1.83)
Criminal Minds	0.80 (1.80)
Bones	0.74 (1.74)
NCIS	0.99 (2.18)
Other forensic	1.46 (2.07)
Total crime show	5.78 (9.99)
Total television	16.78 (11.99)

Table 2.

Stages of the forensic science process: Means and standard deviations for error and human judgment, correlations between error and human judgment, correlations between crime shows viewing and error, and correlations between crime show viewing and human judgment.

Process Stage	Error M(*SD*)	*t* value	Cumulative Error	Human Judgment M(*SD*)	*t* value	Correlations		
						Error and Human Judgment	TV and Error	TV and Human Judgment
Collection	42.48 (27.12)	13.88***	42.48	5.55 (1.60)	9.78***	.297**	.132	.092
Storage	39.35 (28.11)	12.28***	81.83	5.15 (1.66)	6.94***	.260**	.085	.233*
Testing	39.27 (27.77)	12.40***	121.1	4.94 (1.70)	5.55***	.358**	.110	.224*
Analysis	44.55 (27.60)	12.40***	165.65	5.25 (1.52)	8.13***	.504**	.090	.100
Reporting	40.69 (26.87)	13.35***	206.34	5.43 (1.53)	9.39***	.386**	.133	.137
Presenting	42.22 (29.64)	12.62***	248.56	5.55 (1.53)	10.19***	.336**	.012	.087

Note. *p < .05, **p < .01, ***p < .001. “TV” = crime show viewing habits.

Table 3.

Forensic techniques: means and standard deviations for accuracy and human judgment, correlations between accuracy and human judgment, correlations between crime shows (TV) and accuracy, and correlations between crime shows (TV) and human judgment.

						Correlations		
Type of Forensic Evidence	Accuracy M(*SD*)	*t* value	Error (100 - accuracy)	Human Judgment M(*SD*)	*t* value	Accuracy and Human Judgment	TV and Accuracy	TV and Human Judgment
Anthropological	81.33 (15.49)	-5.63***	18.67	4.86 (1.64)	5.29***	-.090	-.127	.070
Blood Pattern	78.53 (19.03)	-6.05***	21.47	4.95 (1.68)	5.70***	-.240*	-.002	-.002
Dental	89.26 (12.04)	-0.62	10.74	4.23 (1.97)	1.16	-.141	-.108	.176
DNA	89.95 (15.85)	-0.03	10.05	3.48 (1.94)	-2.72**	-.257**	-.120	.212*
Document	65.18 (21.37)	-11.68***	34.82	5.72 (1.33)	12.97***	-.198*	.103	-.067
Faces	74.03 (20.12)	-7.98***	25.97	5.06 (1.62)	6.58***	-.141	.149	.067
Fingerprint	88.15 (17.66)	-1.05	11.85	3.68 (1.94)	-1.64	-.324**	-.046	.204*
Fire/Explosives	74.56 (19.55)	-7.93***	25.44	4.67 (1.49)	4.49***	.006	.008	.169
Firearm/Tools	79.63 (16.77)	-6.21***	20.37	4.32 (1.60)	1.99*	.019	-.046	.140
Geological Materials	77.04 (17.05)	-7.64***	22.96	3.97 (1.72)	-0.17	.007	-.096	.154
Gunshot Residue	78.87 (17.97)	-6.23***	21.13	4.20 (1.77)	1.12	-.090	.030	.171
Image	78.21 (16.16)	-7.33***	21.79	5.15 (1.46)	7.91***	-.199*	-.002	.000
Materials	79.37 (18.74)	-5.70***	20.63	4.21 (1.60)	1.30	-.070	-.062	.114
Toxicology	86.66 (13.75)	-2.44*	13.34	3.72 (1.84)	-1.51	-.007	-.006	.248*

						Correlations		
Type of Forensic Evidence	Accuracy M(*SD*)	*t* value	Error (100 - accuracy)	Human Judgment M(*SD*)	*t* value	Accuracy and Human Judgment	TV and Accuracy	TV and Human Judgment
Voice	71.47 (19.16)	-9.72***	28.53	4.81 (1.57)	5.21***	-.178	.121	.104
Wildlife	74.77 (20.68)	-7.40***	25.23	4.24 (1.50)	1.60	-.030	-.046	.124

Note. *$p < .05$, **$p < .01$, ***$p < .001$. “TV” = crime show viewing habits.