What do experimental simulations tell us about the effect of neuro/genetic evidence on jurors?

Nicholas Scurich

Departments of Psychology & Social Behavior, and Criminology, Law & Society, University of California, Irvine, California, 92617, USA
Corresponding author. E-mail: nscurich@uci.edu

One day I received a phone call from Joseph Wu, MD, a colleague in the Psychiatry department at my institution. He had just read an article I published finding no consistent effect of neuroimaging or genetic evidence on reducing jurors’ judgments of criminal culpability. He was perplexed by the findings. It turns out Professor Wu regularly conducts neuroimaging as part of his evaluation of capital defendants and testifies regarding neuropsychiatric abnormalities during the sentencing phase, which is when jurors must decide whether to impose a death sentence or life in prison without parole. According to Professor Wu, his testimony is often successful, in contrast to the findings reported in my article.

During the course of the conversation, Professor Wu spent over half an hour detailing his most recent case in Florida, in which jurors decided against a death sentence on the basis of his testimony. I asked how he could be sure that it was the neuroimages or his testimony that affected the outcome? After all, a defendant who could afford to have pricey neuroimaging conducted presumably could also afford good attorneys and good attorneys are surely more effective than the alternative.

Professor Wu countered by asking about the stimuli used in my experiments. I explained that we had used a synopsis of a criminal proceeding approximately 1000 words in length that was accompanied by a neuroimage or genetic test and explanation. He

Notes:

1. Paul S. Appelbaum, Nicholas Scurich, & Raymond Radd. Effects of Behavioral Genetic Evidence on Perceptions of Criminal Responsibility and Appropriate Punishment, 21 PSYCHOL. PUB. POL’Y & L. 134-44 (2015).

2. The scan itself costs around $7000 (http://health.costhelper.com/pet-scans.html). However, that figure does not include many of the additional fees associated with conducting a forensic examination, including the hourly rate of the expert who must prepare a report and testify about the results of the scan at trial.
found this to be totally unsatisfactory, and he emailed me the PowerPoint deck that he presented to the jurors during his testimony. The deck contains 77 slides that explain in exquisite detail how positron emission tomography works to detect evidence of traumatic brain injury and chronic traumatic encephalopathy. I was impressed. It was obvious to me that the stimuli we used paled in magnitude and detail compared to what Professor Wu had presented to the jury.

In the end, neither of us seemed persuaded. I left the conversation feeling that although my stimuli were impoverished compared to his, there was no way to tell whether it was actually Professor Wu’s testimony that caused the jurors to vote against the death penalty in this particular case. I’m confident Professor Wu left the conversation feeling that I had not realistically tested the phenomenon and therefore my data do not actually supply evidence that neuroimaging has no impact on jurors. We both agreed that there is a need for further testing of the issue.

The tension Professor Wu and I felt is an inherent issue in experimental research. When one is interested in studying cause and effect (e.g., how neuroimages of brain pathology cause jurors to vote against death), one needs to have treatment and control conditions to which participants are randomly allocated. Random allocation ensures the groups are equivalent prior to administering treatments; therefore, any differences detected ex post are presumptively attributable to the experimental treatment.

The downside to conducting experiments is that they require a good deal of resources, since they require multiple experimental groups (treatment, control, etc.), and thus sacrifices to the realism of the procedure must be made for pragmatic reasons. For instance, a capital jury goes through voir dire and death qualification, sits through the guilt phase of trial, listens to direct and cross examination of witnesses, deliberates, and then—if and only if the jury unanimously votes to convict the defendant—goes onto the sentencing phase, in which they consider whether or not to impose a death sentence. This process occurs over days and often weeks. Because it would be prohibitively expensive and impractical to completely mimic this process for the purpose of experimentation, researchers attempt to simulate it in a truncated and compressed manner that allows many individuals to participate in the experiment.

An important question when designing a simulation concerns what elements can be sacrificed without changing the fundamental task that we seek to study. This requires researchers to exercise judgment about tradeoffs. Professors Koehler and Meixner have argued that such tradeoffs should be evaluated according to the ultimate goal of the research. They identify three general goals of jury research: (1) develop theory; (2) describe how juries perform; (3) improve jury performance or affect policy change. They argue, correctly in my judgment, that ecological validity—the extent to which the simulation mimics the real world setting or procedure—ought to figure prominently if goals 2 or 3 are primary. Put differently, if the main objective of the research is to inform legal policy or legal professionals who might use the knowledge for practical purposes (e.g., judges making admissibility decisions; lawyers deciding whether neuroimaging is worth the cost), then a high degree of ecological validity is appropriate. Studies that are low ecological validity are likely to have a limited impact on legal professionals.

3 Johnathan J. Koehler & John B. Meixner. Jury Simulation Goals. In THE PSYCHOLOGY OF JURIES: CURRENT KNOWLEDGE AND A RESEARCH AGENDA FOR THE FUTURE (M. Bull Kovera ed., 2017).
4 Id. at section III.
The lead article by Shen et al. is decidedly low in ecological validity.\(^5\) It, like many of the studies that preceded it,\(^6\) had an online sample of participants read an abridged version of a criminal investigation, presented some participants (randomly determined) with the results of a brain-based memory recognition test, and asked them whether the suspect committed the offense. There was no voir dire, no opening statements, no live presentation of evidence, no closing statements, no jury instructions on the standard of proof, nor any deliberation. In addition to these marked differences from a criminal trial, as the authors note, in an actual trial an expert would present the results of neuro/genetic tests during live testimony over an extended period of time and be subject to cross examination.

This latter point is particularly important in the context of neuro/genetic evidence which tends to be highly technical and convoluted. It is doubtful that a cursory written description comprised of a couple hundred words is remotely similar to the type of technical testimony Professor Wu provides over the course of days with literally dozens of pedagogical images. Moreover, written synopses typically do not have an actual expert at all, which is problematic since (simulation!) evidence suggests jurors are sometimes persuaded by the background and the experience of the expert more so than the substance of her testimony.\(^7\)

My purpose of this comment is not to criticize the tradeoffs that Shen et al. made in designing their experiment. Their decisions in this regard are commonly made by other researchers, including myself. Rather, my point is to call attention to the reality that legal professionals are unlikely to be persuaded by such simulations, and that to the extent researchers hope to inform legal professionals or affect legal policy, it would behoove them to attend to ecological validity when conducting experimental simulations.

Future experimental research examining the impact of neuro/genetic evidence might film the expert testimony and include both direct and cross examination as the experimental stimuli.\(^8\) While the simulated testimony need not be hours long with dozens

---

\(^5\) Francis X. Shen et al., *The Limited Effect of Electroencephalography Memory Recognition Evidence on Assessments of Defendant Credibility*, 3 J. L. & BIOSCI. 330–64 (2017).

\(^6\) I include myself in the company of the guilty: Nicholas Scurich & Paul S. Appelbaum, *The Blunt-Edged Sword: Genetic Explanations of Misbehavior Neither Mitigate Nor Aggravate Punishment*, 3 J. L. & BIOSCI. 140–57 (2016).

\(^7\) Johnathan J. Koehler et al., *Science, Technology, or the Expert Witness: What Influences Jurors’ Judgments About Forensic Science Testimony?*, 22 PSYCHOL. PUB. POL’Y & L. 401–13 (2016).

\(^8\) As an example, see *Id.* at 407.

A well-known review piece is frequently cited to justify the use of paper and pencil stimuli as opposed to video stimuli in jury simulation studies (Brian H. Bornstein, *The Ecological Validity of Jury Simulations: Is the Jury Still Out?* 23 LAW & HUMAN BEHAV. 75-91 (1999)). (‘few differences have been found as a function of ... how the mock trial is presented.’ *Id.* at 88.) However, several factors should give pause to uncritically accepting this conclusion. First, the review is nearly 20 years old and the validity of its conclusions should not be conferred life time status, especially given the vast proliferation of jury simulation studies in the past two decades and the differences in the recruitment/participation mediums (eg online studies completed by crowd-sourced ‘workers’) that have emerged in the interim. Second, blithely citing the review to justify the use of paper and pencil stimuli fails to consider the ultimate goal of the research and whether or not the intended consumer of the research (eg legal professionals) will find it relevant or persuasive, as noted by Professors Koehler and Meixner, *supra* note 3. Third, and most importantly, the review did not examine the effect of different mediums on the evaluation of expert testimony specifically. There is every reason to believe that the medium (video vs. written) is likely to exhibit the largest difference when the issue under study is expert testimony on highly technical matters. Neuro/genetic evidence falls into that category.
of PowerPoint slides, it surely ought to be longer than a few hundred words. And the stimuli should include background information about the underlying science, how the neuro/genetic tests are conducted, and how the results relate to a germane legal question, as is the case in actual expert testimony. Aside from the substance of the testimony, it seems to me that furnishing judicial instructions to participants would not be too onerous. Including juror deliberation is no doubt a costly addition but it may be necessary quiet skeptics.

In sum, online simulations have their purpose and they are economically efficient to conduct. Indeed, participants recruited through Amazon’s Mechanical Turk will complete online studies for literally pennies. But these types of studies do not supply the type of evidence that is likely to affect the decisions made by legal professionals. If experimental simulations are to have a real-world impact, it is time for the field to ween off ecologically invalid experiments and invest more resources to better emulate the setting and conditions of a jury when studying the impact of neuro/genetic evidence on jurors.

---

9 See Michael Buhrmester, Tracy Kwang, & Samuel D. Gosling, *Amazon’s Mechanical Turk: A New Source of Inexpensive, Yet High-Quality, Data?*, 6 PERSP. PSYCHOL. SCI. 3–5 (2011). Note that it is possible to conduct online studies using video stimuli of the sort I am proposing here. Thus, experimental simulations with higher degrees of ecological validity can still be conducted inexpensively. Of course, the time and effort required to produce elaborate stimuli is an additional cost (over using written stimuli), but that is a necessary cost of conducting the type of research that is likely to be viewed as relevant by legal professionals.