

*Much More than a Hunch*

# Do Mock Trials Predict Actual Trial Outcomes?



**George Speckart, Ph.D.**

## INTRODUCTION

In order to obtain the right information, one must ask the right questions. The question of whether mock trials predict should not be treated casually – after all, if mock trials did – or could – predict actual trial outcomes, the policy implications for trial planning and settlement decisions would be enormous.

So, it is important to ask the right question, but in order to ask the right question, one must first be clear on terminology. In particular, one should first define what one means by “mock trial.” If by “mock trial” one is referring to a legal exercise designed to provide practice for the lawyers, and the associated insights that come from practice, one should not expect prediction, as this is not a realistic goal for such an exercise. On the other hand, if one is referring to “mock trial” as *psychological research*, then the question of predictive utility carries importance that demands consideration. It is assumed presently that carrying out a “mock trial” is not considered just as a legal exercise (practice), but rather falls under scrutiny as a form of psychological research.

From a different perspective, one may ask the questions involved as follows: *Do mock trials predict?* The answer is “sometimes – it depends on a number of factors.” *Can mock trials predict?* The answer is “yes, to a substantial degree, if they are conducted in a certain way.” The purpose of this article is to explore when and how prediction is achieved, and whether this amounts to “dumb luck” or whether something else is going on. In other words, are there systematic factors that allow one to achieve the goal of predictive utility in mock trial research? This treatise asserts an affirmative answer, and provides explanations as to how and why – with the proviso that *perfect* prediction still, of course, remains an unachieved goal.

In terms of the current state of the industry, some mock trials do not predict at all; some have modest degrees of prediction (moderate amounts of “predictive utility” – for example, perhaps two out of three mock juries within a mock trial research project provide the same verdict as the actual jury); and others have been found to predict not only liability, but damages very accurately indeed (all of the mock juries have the same verdict as the actual jury, and the average damages are close to those of the actual jury, if it is a plaintiff-oriented outcome). The real questions then become, “Are there identifiable differences in the research design and implementation between those mock trials that predict (achieve predictive utility) and those that do not?” What creates predictive utility in a mock trial?” And, “How consistent is this predictive utility across multiple projects when the research is appropriately designed and implemented?”

**Can mock trials  
predict? The  
answer is “yes, to a  
substantial degree,  
if they are  
conducted in a  
certain way.”**

One additional component of asking the right question entails the matter of what is actually being predicted. Does the question pertain to prediction of liability only, or to prediction of damages? While the latter is certainly more challenging (liability is easier to predict than damages), it is submitted presently that mock trials *can also* predict damages – but again, there is the distinction of whether they can and whether they do, with the same conclusions as before: Whether they *do* predict damages depends on how they are conducted, and the answer as to whether they *can* is “Yes, to a substantial degree.”

While the position is taken presently that prediction is achievable “to a substantial degree”, the qualifier “to a substantial degree” is included to acknowledge that perfect prediction is obviously not possible. Unpredictable factors often impinge on a trial, from court rulings to volatile witnesses to the mysterious “luck of the draw” in jury selection. However, the position taken presently is that, generally speaking, when the research is designed and implemented correctly, prediction of not only liability but also damages is possible on a level that far surpasses the accuracy of “guesses” or “hunches,” and in many cases, is surprisingly accurate.

Now that we have established that whether mock trials predict depends on how they are conducted, let’s start with the first factor, which is the designer and implementer of the research itself.

# WHO'S CONDUCTING THE RESEARCH?

**Jury consultants with established expertise or credentials in the prediction of behavior are quite rare. Therefore, it is not surprising that the most prevalent opinion appears to be that mock trials do not predict.**

Another way of answering the question of whether mock trials predict is “It depends on who is doing the research.” First and foremost, this article documents several exemplars of precise prediction from mock trials (noted below), both in terms of verdict and damages. In each and every one of these cases, the legal team was top notch, working assiduously to put together the most realistic mock trial possible (in the heavy equipment case noted, there were crates weighing hundreds of pounds bringing in scale models of the equipment, and it required a full day on either side of the project just to pack and unpack them). In classic “garbage-in-garbage-out” style, mock trials are more predictive when the legal team is fully immersed, on board, enthusiastic, and willing to work hard. So, the first component of “Who’s doing the research?” is always the legal team itself. Do they appreciate “balance” – that both presentations have to be as persuasive as possible? Are the graphics comparable on both sides? Are witness excerpts chosen to be truly representative? Are their hearts really in the exercise?

Now, a few words about the researchers are in order.

We were recently approached by an insurance company requesting mock trial research, and the company opted for a cheaper alternative. Another service ran the mock trials for \$10,000 (an unusually low cost in this industry). The mock trial included three mock juries that each gave a defendant verdict. The client subsequently went to trial and got hit for \$18 million. The exact same scenario was again recently observed with an automotive manufacturer. Clearly, the qualification of the researchers is not a trivial issue. In the realm of jury research, efforts to slash costs can – and typically do -- backfire with the most disastrous of results, for reasons that become readily apparent once the nature of the industry is examined.

Those jury consultants with established expertise or credentials in the prediction of behavior are quite rare. Therefore, it is not surprising that the most prevalent opinion appears to clearly be that mock trials do not predict. Looking at the practitioners in the field today, this opinion is certainly understandable: The field of jury consulting has no barriers to entry whatsoever, leading to “bargain services” conducted by practitioners, who, before entering the industry, were receptionists; paralegals; acting coaches; accountants; and even cooks. Why would anyone expect their mock trials predict behavior, when prediction is difficult enough as it is for psychologists who specialize in prediction as their life’s work?

The field of jury consulting is also characterized by buyers who are largely unsophisticated and do not examine qualifications or credentials of those who run the research. Instead, hiring decisions are routinely made exclusively as a result of interpersonal relationships or presence on a “preferred vendor” list. Thus, for example, insurance companies, who put individual applicants through the wringer in order to buy a simple life insurance policy, routinely expose themselves to multi-million dollar losses following “research” designed to predict trial outcomes that has been conducted by someone who was never vetted at all, and who may well have spent the last several years of his or her work experience as a pre-school teacher, paralegal, or acting coach.

In twenty-eight years of running mock trials, I have not once been asked “Do you have any background, credentials or training in the prediction of behavior?” Nor have I ever heard, among dozens of colleagues in the field who are friends, of any potential client ever asking this question. However, prediction is a vital, established area of psychological research, and if the client wants this expertise, the client is entitled to ask for it, and

obtain it. Nonetheless, the fact is that the clients simply do not ask for this particular skill, nor do they even seek it, at least if the queries of prospective service providers are any indication.

I would estimate that 10% at most ask if I have a Ph. D., and maybe half of these ask in which field was the degree obtained. But the answer “psychology” is as far as the conversation gets. The client still does not know – nor does the client appear to care – whether my formative training involves working with autistic children; counseling and psychotherapy; chasing rats through mazes; or managing a pain clinic. Most Ph.D.’s in psychology who are currently jury consultants have backgrounds of these types, yet none of these backgrounds is connected to expertise or qualifications in designing and implementing research to predict human behavior.

Just as a gastroenterologist would not perform eye surgery, there are different types of psychologists who are qualified to do some things and not others. But just as there are no barriers to entry, there are also no ethical standards regulating conduct by those in the field, and so research is routinely requested (for example by insurance carriers to estimate exposure) but service providers will not say “I’m sorry but my area of training does not include research on the prediction of behavior.” Imagine if you went to a urologist because of a mole on your skin that you believed was cancerous – she would send you to a dermatologist immediately. But I know dozens of jury consultants whose background consists entirely of psychotherapy, working with children, counseling etc., and when approached by a corporate client wanting to estimate potential damages in a case (prediction of behavior), there is no hesitation, no referral, and no discussion of qualifications.

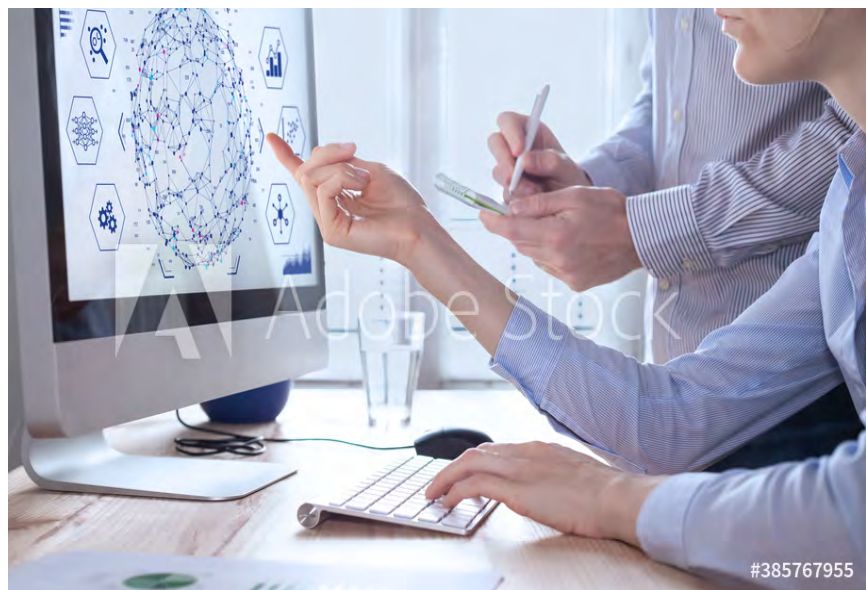
## **Estimating damages or exposure is an attempt to predict behavior – namely, the behavior of a group of people in making a decision on damages.**

In other words, if there ever was a “buyer beware” industry, this is it.

Many have Ph. D.’s in Communication, which constitutes an excellent background for witness training and assistance with opening statements, but again, there is no training in this discipline on research for the prediction of behavior. There are many jury consultants with other types of Ph. D.’s; one particularly prominent jury consultant that I know has a Ph. D. in college administration and none of his clients appear to care.

Of course, many practitioners in the field may, quite legitimately, not care about prediction at all. Mock trials can be hugely informative but still fail to predict – however,

the topic under consideration presently is prediction. Estimating damages or exposure is an attempt to predict behavior – namely, the behavior of a group of people in making a decision on damages. If one simply wants to know how a group of people react to various themes and arguments, that’s one thing – but if there is a need to determine potential damages or exposure, then one is making inferences about a group’s (a jury’s) behavior in the future, and that is prediction. The present thesis merely asserts that it is reasonable for the practitioner to have a background of accomplishment in this area, if the client wishes to make inferences about a probable trial outcome, and that the client should shop as carefully in this area as anyone would in buying anything costing \$30,000 or more.





## HOW IS PREDICTIVE UTILITY OBTAINED?

Progressing now from the researchers to the research design and implementation itself, let's examine how mock trials can be designed and implemented in order to achieve predictive utility. In research terms, this is also called "validity" – namely, the extent to which research can be used to generalize to real world events – in this case, actual courtroom verdicts and damages.

Validity is a three-legged stool consisting of three key components: 1) The Participants, or "mock jurors" (Do they faithfully represent the venire?); 2) The Presentations (Are the lawyers presenting the same materials that the real jury will hear?); and 3) Analysis (Is the data being analyzed in a methodologically sound manner?). In this sense "validity" and "predictive utility" are essentially synonymous, and if these three requirements are met, generally the goal of reasonably precise forecasting is achieved.


In a simple common sense manner, predictive utility is not all that far-fetched – after all, it is a reasonable proposition that, if you give the same type of people information that is equivalent to what the real jury will see and hear, they will react in pretty much the same manner. Of course, actual implementation of the research is not quite so simple, but it follows this general line of reasoning. When every decision in the mock trial procedure is resolved based on the "gold standard" criterion of "What will the real jury see and hear?," the results become progressively more realistic.

Obviously, with regard to the first "leg" of the "stool," the respondents chosen to participate must reflect those individuals who are likely to actually be seated on the jury panel of the particular court in question. Satisfying this aspect of preparation involves elaborate recruiting, measurement and screening of prospective mock jurors. However, most of the controversy in this field centers around the second item, as those arguing against the potential for predictive utility will contend that one cannot possibly simulate the events in a courtroom. Of course, strictly speaking, they are right – one cannot condense an entire trial into a one, two, or three-day exercise. But certain factors have to be taken into consideration before coming to a conclusion in this area.

First, jurors do not deliberate based on what happens in the courtroom, they deliberate based on what they *store and retain in their memories* and then *retrieve from memory later* – a tiny subset of that which occurs in court (S. Tuholski, "When facts don't fit, some jurors make up new facts", *National Law Journal*, February 4, 2008). This is where experience comes into play in designing and implementing the mock trial. Obviously, credentials of the researcher alone are not enough to obtain predictive utility, and the selection of those points of evidence that are pivotal in the case and that need to be included in the mock trial presentations requires the combined years of judgment from the entire trial team; prior focus group research identifying jurors' "hot buttons," if possible; and other substantive considerations based on the case fact scenario and types of claims involved.

Second, jurors do not make up their minds on the basis of opening statements (a common myth) – they make up their minds when *listening to the witnesses*. (Incidentally, this is why inferences as to future verdicts cannot be made based on spreadsheets of past verdicts in the venue. Verdict and damages decisions hinge on witness testimony that varies greatly in appeal and persuasiveness, even across the same types of cases). Therefore, condensing the witness testimony into its essential components, and reflecting this testimony faithfully in the mock trial project, are pivotal elements in achieving predictive accuracy. This is where the researchers become heavily dependent on the lawyers in selecting such testimony, and why the assertion is *not* made that, when predictive utility is achieved, it is solely because of the qualifications of the researchers. Predictive utility is absolutely a *team effort*. (Although it is worth noting that a good researcher should "coach" the trial team on these issues to get the most out of this team effort.)

Jurors do not make up their minds on the basis of opening statements (a common myth) – they make up their minds when listening to the witnesses.



One of the reasons that mock trials in civil cases can be predictive, therefore, is the deposition. It is doubtful that one would make a claim for predictive utility in criminal cases, where depositions do not exist, since one does not know enough in advance concerning what the witness testimony will be. But in civil cases, in which the witnesses are more or less tied to the depositions, predictive utility becomes more achievable because the witness testimony is largely known in advance. However, faithfully distilling vital witness testimony into a mock trial project is exceedingly labor intensive, and it often cannot be done in a one day project. Indeed, of all of the instances in which predictive utility was accomplished from our records, the highest levels of accuracy are found in multi-day projects in which a great deal of painstaking labor was expended to get the witness testimony “right.”

In the area of analysis (the third leg of the stool), many researchers average the damage awards proffered by each respondent in a focus group or mock trial to obtain an expectation of the damages in the case, but this is an incorrect procedure, as juries award damages differently than individuals. In particular, research from various sources has identified what has been termed a “severity shift” demonstrating that damages awarded by a group tend to shift upward, or escalate, as compared to damages awarded by individuals acting alone (Sunstein et al., *Punitive Damages: How Juries Decide*, University of Chicago, 2002). Thus, the proper way for estimating potential damages is to average across juries, not across individuals.



The area of psychological measurement, or psychometrics, is a close sister to the area of prediction in psychological research methodology, and various aspects of psychological measurement must be appropriately observed and implemented within the mock trial research in order to ensure predictive utility. To enumerate these in their entirety would be beyond the scope of the present treatise, but it suffices to say that questionnaires cannot and should not be designed and administered without attention to proper psychometric criteria.

Finally, it should be noted that the achievement of predictive validity in this field requires years of making mistakes and finding out various things that simply make the research “go wrong.” If the plaintiff has color graphics and animations while the defendant simply puts up black and white documents, there is a perceptual shift in balance, or what psychologists call the “demand characteristics” of the experiment, that creates an artificial bias in the results. Bias can be introduced in innumerable other ways, and there is generally no substitute for raw experience in this area, with the project under the watchful eye of an experimenter who has suffered all of the embarrassments of research that went awry in various ways over the years.

## THE RECORD

Following is just a partial list of notable examples of precise prediction in mock trial research taken from our archives. While the skeptical reader will no doubt conclude that these have been cherry-picked, it is of course admitted that many exemplars also exist that did not predict. As noted earlier, unexpected rulings by the court; unstable witnesses; vicissitudes in the jury panels (what litigators often call the “luck of the draw” when a venire walks into court); and other factors do sometimes play a role.

	Mock Jury 1	Mock Jury 2	Mock Jury 3	Mock Jury 4	Avg. Mock Award/ Actual Reward
<i>ETSI v. Burlington Northern et al.</i> , 1989	\$500 million	\$160 million	\$310 million		\$323 million/ \$345 million
<i>Newman v. Stringfellow Superfund toxic case</i> , 1992	\$175,000	\$300,000	\$80,000		\$183,000/ \$138,000
<i>Exxon Valdez</i> , 1994	\$2 billion	\$3 billion	\$4 billion	\$12 billion	\$5.2 billion/ 5.0 billion
<i>AHDC v. Fresno</i> , 2001	\$1,000	\$1	\$10,000		\$3,667/ \$1
<i>Steele v. First Union</i> 2002	\$140 million	\$275 million	\$320 million		\$245 million/ \$239 million
Heavy Equipment Case, 2003	\$25 million	\$37 million	\$112 million		\$58 million/ \$55 million
Legal Malpractice Case, 2008	\$88 million	\$20 million	\$140 million		\$83 million/ \$73 million

Note: Some names have been withheld per the wishes of the client.

Our position is simply that mock trials *can* predict, and *do* predict, when certain conditions are met, and that most of these conditions are under volitional control of the trial team and the client. The existence of mock trials that do not predict is not proof that they cannot. It is more often proof that either the researchers were unqualified, or that the trial team was not willing to expend the time and effort necessary to get it right. Moreover, in cases characterized by the latter factors – i.e., the trial team is not sufficiently involved – this is not meant to be a disparagement of the lawyers. Many mock trials (quite unfortunately) are conducted under “trial by hurry” conditions on the eve of trial, and there are simply not sufficient time or personnel resources available to expend the necessary efforts.

The previous list represents a subset of those from our database that went to trial and in which damages were awarded. The list with defense verdicts in both mock and real trials is quite longer, and not reproduced presently.

**Our position is simply that mock trials can predict, and do predict, when certain conditions are met, and that most of these conditions are under volitional control of the trial team and the client.**

## CONCLUSIONS

The emergence of the potential for predictive utility over the last thirty years of trial sciences has profound policy implications for settlement practice. Such policy implications naturally entail consideration of numerous cost/benefit issues in view of the general consensus that, for most cases, research of this type cannot fit into most litigation budgets. One very fine article about mock trials was written by J. C. Johnson ("Mock Juries: Why Use Them?," *Litigation*, Volume 35, Number 2, 2009) in which the author appears to take the position that a mock trial is somewhat of a luxury – something the client should use if he can afford it.

While on its face the establishment of predictive utility would appear to provide obvious benefits, examination of the costs associated with how cases are actually resolved reveals that such benefits are more extensive than one might initially surmise. Most cases are currently resolved on the basis of "intuition" that diverges substantially from the amount that a jury would actually award. To the extent that "intuition" is faulty -- and from our

experience it usually is – conducting scientifically valid research to determine exposure is highly cost effective, since the margin of error in "intuition" is many times greater than the cost of the research. In fact, our analysis based on cases in which settlement negotiations were aborted in order to conduct research indicates that the margin of error in guesses or hunches (i.e., the amount that proposed settlement amounts diverge from scientifically valid estimates of what a jury would actually do with the case) is typically more than ten times the cost of scientifically valid research.

In short, guessing is not only more expensive than the research – it's *far more expensive* than the research, and the irony is that most decision makers appear to believe precisely the opposite. These cost effectiveness issues have been dealt with in other forums (G. Speckart, "Trial by Science," *Risk & Insurance*, October 2008) but the timeliness of the message becomes more and

more urgent as decision-makers continue to regard research costs as "too expensive" but make decisions to settle cases for amounts that diverge from the true value of the case by more than ten times the amount of such costs. While it is recognized there are other factors that dictate settlement value (such as nuisance, risk, and image), the fact is that what a jury would actually do with the case is still part of the calculus in most instances, but the need to obtain accurate information in this area in order to minimize expenses is consistently overlooked.

There are ethical issues as well: If the potential damages are knowable, does a client have the right to this information? Should those involved in settlement and mediation be made aware that an accurate answer is possible when bartering on the value of a case? Are trial team members – the lawyers – justified in making assessments of exposure based on intuition? Is there an ethical issue when lawyers attempt to predict behavior of juries without the benefit of expertise in this area? Furthermore, are there ethical issues involved in entrusting the research to jury consultants who do not have qualifications?

**In short, guessing is not only more expensive than the research – it's far more expensive than the research, and the irony is that most decision makers appear to believe precisely the opposite.**



If mock trials predict, then an entirely new way of viewing mock trials is required. They are no longer a luxury – they are a diagnostic tool implemented to systematically assess exposure – and the question then becomes not whether the client can afford to use them, but rather whether the client can afford *not* to. My last client who claimed that a mock trial was a “luxury” in fact went to trial shortly thereafter and sustained a damage award of \$56 million. In such cases, one might justifiably consider whether the “luxury” was instead the decision to forego the use of the research.

The quality of research progresses when it is well-funded. The litigation environment does indeed provide such a context. As thirty years of trial consulting in civil cases has elapsed now, it is more or less inevitable that the state of the science would progress to the point at which accurate prediction is achievable. The future is here – it is time for more rational decisions on settlements to take over, particularly in a climate in which cost effectiveness is important to clients.

## ABOUT THE AUTHOR

Dr. Speckart received his Ph.D. in Psychology from UCLA in 1984 with a specialization in personality and measurement, and has published extensively in the application of statistical models to the prediction of behavior. He has been active in the jury consulting field since 1983, and has conducted over 800 mock trials and focus groups in pre-trial research for numerous types of litigation. Dr. Speckart has worked with litigators in over 150 jury selections, beginning with Dalkon Shields cases in 1983, the Agent Orange litigation in 1984, and the *Exxon Valdez* litigation in 1994. He has presented regularly to the DRI (Defense Research Institute); IADC (International Association of Defense Counsel); FDCC (Federation of Defense and Corporate Counsel); ABOTA (American Board of Trial Advocates); and various state bar conventions throughout the country.