

# AI, Decision Science, and Psychological Theory in Decisions about People

## A Case Study in Jury Selection<sup>1</sup>

*Roy Lachman*

■ AI theory and its technology is rarely consulted in attempted resolutions of social problems. Solutions often require that decision-analytic techniques be combined with expert systems. The emerging literature on combined systems is directed at domains where the prediction of human behavior is not required. A foundational shift in AI presuppositions to intelligent agents working in collaboration provides an opportunity to explore efforts to improve the performance of social institutions that depend on accurate prediction of human behavior. Professionals concerned with human outcomes make decisions that are intuitive or analytic or some combination of both. The relative efficacy of each decision type is described. Justifications and methodology are presented for combining analytic and intuitive agents in an expert system that supports professional decision making. Psychological grounds for the allocation of functions to agents are reviewed. Jury selection, the prototype domain, is described as a process typical of others that, at their core, require the prediction of human behavior. The domain is used to demonstrate the formal components, steps in construction, and challenges of developing and testing a hybrid system based on the allocation of function. The principle that the research taught us about the allocation of function is “the rational and predictive primacy of a statistical agent to an intuitive agent in construction of a production system.” We learned that the reverse of this principle is appropriate for identifying and classifying human responses to questions and generally dealing with unexpected events in a courtroom and elsewhere. This principle and approach should be paradigmatic of the class of collaborative models that capitalizes on the unique strengths of AI knowledge-based systems. The methodology used in the courtroom is described along with the history of the project and implications for the development of related AI systems. Empirical data are reported that portend the possibility of impressive

predictive ability in the combined approach relative to other current approaches. Problems encountered and those remaining are discussed, including the limits of empirical research and standards of validation. The system presented demonstrates the challenges and opportunities inherent in developing and using AI-collaborative technology to solve social problems.

Significant progress has been reported by AI researchers in combining decision-analytic techniques with knowledge-based expert systems. Historically, the most common AI methods for inference and decision making involved inquiry that was conducted in isolation from associated fields of study. The limitations of this approach have been examined, with the result that the methodologies of decision making under uncertainty are under continuing evaluation and are being incorporated into knowledge-based systems (Henrion, Breese, and Horvitz 1991). Applications that combine decision-theoretic approaches with expert systems include medical diagnosis, product development, system troubleshooting, and various tasks involving candidate evaluation (Durkin 1993; Mitri 1991). Overall, AI and decision-theoretic methods in combination produce superior results than either alone. Although the applications address some socially and financially significant domains that require the prediction of human behavior, there are still many pressing problems for which AI theory, concepts, and methodology provide potentially powerful solutions, yet AI methods are not currently being tested or extended for application to these problem areas.

*Only a machine can process and integrate large amounts of demographic and attitudinal data about a prospective juror. Only a person can recognize the meanings of unanticipated responses of other people and identify the significance of these responses.*

This situation is regrettable, given the increasing maturity and coherence of the foundational paradigm of AI science that is reflected in changes in the field's presuppositions. Previously, AI systems were generally treated as single-agent entities working in isolation from the world and other agents; recently, the discipline has begun to adopt the more productive assumption that AI is one of a group of intelligent agents working in collaboration. This presuppositional shift is reflected in two recent presidential addresses to the American Association for Artificial Intelligence (Grosz 1996; Bobrow 1991). Daniel Bobrow, in his address, presented an illuminating analysis of three interactive dimensions of intelligent agents: (1) communication, (2) coordination, and (3) integration. Various human and machine agents, even with effective training or modification, can only do each of the three types of activity described by Bobrow with differing levels of skill and effectiveness. The result is varying degrees of success in the resulting decisions made or problems undertaken. Barbara Grosz suggested that collaboration is central to intelligent behavior. She also showed that the processes and capabilities needed for collaborative efforts must be designed into an interactive system from the beginning of the design process. The juror-evaluation system, described here, bears testimony to the wisdom of this view.

There is another important dimension to collaborative efforts—the intelligent allocation of function. Effective collaboration, of course, requires that tasks be assigned to the agent best qualified to accomplish them. The analytic procedures developed for performance and task analysis in industrial-organizational psychology (Campbell, Gasser, and Oswald 1996; Clegg et al. 1996; Borman and Motowidlo 1993; Goldstein, Zedeck, and Schneider 1993; Ash 1988; McCormick and Jeanerett 1988) and in cognitive psychology and human factors (Howell 1991; Jonassen, Hannum, and Tessmer 1989; McCormick 1979) suggest possible methods for the allocation of tasks among agents.

This article reports on a research program designed to incorporate decision-theoretic models into a knowledge-based system for jury selection using collaborative allocation of function. Only a machine can process and integrate large amounts of demographic and attitudinal data about a prospective juror. Only a person can recognize the meanings of unanticipated responses of other people and identify the significance of these responses. The domain was selected initially for reasons of

importance, convenience, and anticipated feasibility.<sup>2</sup> The proper selection and composition of jury panels, for almost a millennium, has been viewed as essential to a fair trial and a rational verdict; the problem is important. A practical consideration influencing domain selection is that the author has the cooperation of a practicing litigator who also has a Ph.D. in cognitive psychology.

One goal of the research is to produce a proof-of-concept working system in one domain to show the value of using AI technology and psychological theory together to address a variety of social problems where the prediction of human behavior is essential to a solution. Although the task turned out to be more difficult than anticipated, as is often the case in efforts that bridge several applied and basic sciences, it has proved to be practical. All our experiences with this project, however, reinforce the conviction that psychological, decision-theoretic, and AI methodologies must be combined as collaborative agents for the solution of significant social problems.

Experienced and highly trained professionals in a variety of occupations are continuously required to make psychologically informed decisions. The usefulness of these decisions depends on the adequacy with which the decision maker is able to predict human behavior. Some examples include decisions regarding employment, product development, investing, custody, parole, psychotherapy, and jury selection. Each action requires decision making under uncertainty; is most often done intuitively; and sometimes is transacted with the aid of actuarial or other formal analytic devices such as a standardized test, multiple-regression equation, or Bayes's theorem. Intuitive judgment, at its best, is a rapid decision process done from memory, without analytic aids, and is based on years of professional experience and training. Clinical judgment is simply the intuitive decision-making of medical or mental health practitioners and is similar in many ways to engineering, economic, and legal decision making (Kleinmuntz 1990). Ever since the publication of a book by Meehl (1954) comparing clinical and actuarial judgment, the related psychological literature has experimentally evaluated the relative effectiveness of a person's intuitive judgments with comparable formal decisions (Dawes, Faust, and Meehl 1989; Keren and Wagenaar 1987; Dawes 1979; Einhorn and Hogarth 1975). Typically, the intuitive judgments of a sample of experts such as clinical psychologists are compared to judgments derived from analytic and empirical rules. Analytic judgment is almost always

superior. Intuitive judgment is subject to random fluctuation and, hence, to decreased reliability and accuracy. In the statistical method, variables can be included or weighted to contribute to a decision based on their actual predictive power. The rules or operations used to make analytic decisions are based on a consensus standard of rationality, including standards based on the established experimental and statistical predictability of the events of interest.

Intuitive and analytic judgments in this literature have traditionally been treated as unrelated, but of course, they are related, and the most recent studies show that each approach has distinctive strengths (Hoch and Schkade 1996; Hoch 1993; Yaniv and Hogarth 1993; Blattberg and Hoch 1990). For the Blattberg and Hoch experiments, the collaborative combination of intuitive and analytic decision making was found to produce more effective decision making than either method alone. Behind every intuitive and statistical decision, ultimately there is a person with knowledge, training, and skills that are grounded in a professional and scientific literature. The person defines the problem space, the goals to be achieved, and the acceptable paths to a solution or decision. The discipline defines the standard of rationality in terms of the logic of validation and the resulting clinical, experimental, and statistical data on which decisions are to be made.

One premise of this article is that professional judgment can be improved by combining and integrating analytic and intuitive decision making. In the language of AI, what is advocated is a collaboration between the two to capitalize on their respective strengths. The difficulties associated with analytic decision systems can substantially be overcome by using behavioral decision theory instantiated in a computer-based decision system. The venerable literatures of the normative and descriptive theories of choice provide the rationale (Kahneman 1991; Hogarth 1990; Bell, Raiffa, and Tversky 1988; Arkes and Hammond 1986; Fishburn 1982; Kahneman, Slovic, and Tversky 1982; Tversky and Kahneman 1974; Simon 1955; Savage 1954; von Neumann and Morgenstern 1953). Several ideas of naturalistic decision making (Zsombok and Klein 1996) suggest some changes to this rationale. The equally venerable literatures of knowledge-based systems (Buchanan and Shortliffe 1984; Lindsay et al. 1993; Steels and McDermott 1993) and the resulting applications (Durkin 1993; Lachman 1989) provide the technology of choice. The pressure of numerous societal problems, solutions to which require accurate

prediction of behavior, together with the rising cost of the relevant predictive expertise, provide reasons to identify the conditions under which expertise can be incorporated into intelligent computer systems. These points, along with issues of validation critical to any scientific enterprise and the prospects and difficulties of this cross-disciplinary approach are demonstrated here by the development of a collaborative decision system for jury selection.

## A Collaborative System: Analytic and Intuitive

Blattberg and Hoch (1990) analyzed online business forecasting and developed a method for isolating intuition and coupling it with a statistical model. Intuition was isolated by regressing the humans' decisions onto the model's decisions. The authors then tested the procedure of combining the statistical model with the intuitive judgments of managers. The combined allocation produced superior forecasts than either component alone. Until more is known on how to improve predictions, Blattberg and Hoch recommend a nonoptimal and pragmatic approach to allocation of function to agents. They suggest decisions should be divided 50 percent to the manager and 50 percent to the statistical model. The context of the jury-selection expert system dictated a somewhat different method of allocation. Initially, we assumed that a statistical model for *voir dire*, the questioning of prospective jurors, was not possible because the questions asked and the responses given have a high level of uncertainty. This assumption proved to be wrong, and a statistical model for *voir dire* was, in fact, developed. The system still places human judgment at the threshold during *voir dire* in that the operator of the expert system interprets the answers to questions posed to prospective jurors. However, the juror's answers are then classified into categories coded into the knowledge base prior to the start of a trial. Statistical and deterministic productions then evaluate all the information on each prospective juror.

The system is designed to eliminate jurors who come to the courtroom with a bias in favor of one litigant or against another, the defendant or the plaintiff. Jury selection (which is more correctly denominated a process of deselection) is subject to the same biases and difficulties as intuitive decisions in other professional areas, which often require the integration of a large number of variables under severe time constraints. Obviously, a useful expert system must be designed around

...  
*professional judgment can be improved by combining and integrating analytic and intuitive decision making. In the language of AI, what is advocated is a collaboration between the two to capitalize on their respective strengths.*

the actual context of decision making, in this case the procedures used in the courtroom. These processes vary from state to state, state court to federal court, one state or federal courthouse to another, and even judge to judge within the same courthouse. This variation mirrors other contexts in which professionals make judgments requiring prediction of human behavior; these also vary in different corporate, geographic, forensic, academic, and clinical settings. After the allocation of function, the strategy is to develop and test a system for a limited range of cases, settings, and locations. The next step is to determine what generalizations are possible and to generalize appropriately. The procedures used in the U.S. District Court for the Southern District of Texas, Houston Division, were selected because our expert had practiced extensively in this court and knew its procedures and idiosyncrasies. In addition, the judges of the district court voted to allow dissertation research related to the expert system. In this jurisdiction, the current court questionnaire (containing 26 items, some with multiple subparts) is administered to all prospective jurors in the jury-pool waiting room. Trial attorneys have 45 to 60 minutes to study the written information provided by the venire persons, most commonly numbering from approximately 28 to 45. The judge conducts a brief voir dire session; some judges permit the attorneys to ask a few questions as well. The opposing attorneys are then typically allowed 10 to 15 minutes to decide on whom to strike. These are the procedures for which the expert system was designed; however, modifications for other court procedures were relatively straightforward.

The goal of the attorneys, of course, is to strike potential jurors who are likely be biased against the attorneys' client (or in favor of the opponent). What constitutes a desirable juror obviously depends on the nature of the case. Because our trial expert litigated civil cases involving employment discrimination, we constructed an initial jury-selection system for a civil matter in which an individual or small entity is suing a larger corporation or entity for some form of employment discrimination. These cases are likely to share many dimensions with others in which a small entity or an individual is suing a larger entity or a corporation, that is, so-called "David and Goliath" conflicts.

In constructing an expert system, it is not unusual to recruit, early on, one or more experts and establish that the experts are suitable for the project. Many trial lawyers are not suitable candidates for expertise in jury selec-

tion. The jury-selection decisions of some lawyers are based on legal folklore rather than fact, and occasionally, the selection processes are irrational and bizarre (Hastie 1993; Wrightsman 1991; Fulero and Penrod 1990; Pennington and Hastie 1990). In trials with large stakes, jury-selection consultants are often hired. To our knowledge, no such consultant has ever published the evidence of controlled experiments indicating that they can give effective advice for preemptive strikes. Although some of them have published books suggesting that they are using good common sense and a few useful psychological principles, their commercial success appears to depend on the high-profile case rather than a sustained record of good results. Moreover, none have ever reported scientific studies of validation.

For this project, the primary expert was a J.D. with 15 years of first-chair litigation experience, who also has a Ph.D. in cognitive psychology and has never lost a civil case. As part of the rule-formulation process, a number of consultations were undertaken with various of the expert's legal colleagues. The expert had also provided jury-selection advice to other law firms. She used general social science expertise, combined with simple heuristics, for assessing the fit of each juror to the ideal juror for the case.

Initially, a simple empirical model for jury selection commissioned by the American Bar Association was located (Abbott 1987). Although the model was dated, it provided a starting point. The model used statistical data for some of the intuitive techniques used by our expert and provided values for a set of demographic variables. Abbott's model had the advantage, or so we thought, that the relationship of each variable to particular juror biases had been scaled and was supported by empirical surveys. Some of the initial scales for demographic information were drawn from the model. The scales, where possible, were then adjusted or updated from recent national social surveys (Smith and James 1996; Davis and Smith 1992). Next, the expert identified the variables that she considered most relevant to David and Goliath cases, in this instance, economic conservatism and nurturance. *Nurturance* refers to a trait or impulse to take care of people, which can motivate a juror to vote for a plaintiff with an objective injury, whether or not the defendant is the party responsible for the injury. Because no statistical data were available on nurturance, her judgment was used initially to determine what characteristics were predictive of a nurturant style and the

ExpertSys - JURY

Auto

Type of work of venireman No. 12, Seger Arthur?

Clerical	Craft	Farmer
Homemaker	Laborer Operative	Manager Proprietor
Part time ◀	Professional	Retired
Sales	Service	Student
Technical	Unemployed	Cannot Classify

What is the marital status of venireman No. 12, Seger Arthur?

0=Married                      1=Single

0                                      1 ◀

How many years of formal education did No. 12, Seger Arthur complete?

8 = Grade School Graduate, 12 = High School Graduate, 16 = College Graduate

6	7	8
9	10	11
12	13	14
15	16	17

Figure 1. Current Version of the Juror Bias Model.

values to be assigned. For example, women received a higher nurturance score than men, other things being equal, and women with children received a higher nurturance score than those without. One can legitimately argue about the expert's judgments, but this was a starting point that was modified when the experimental and survey statistics became available. It is instructive to see which of the expert's judgments have been correct.

The rationale of this initial strategy was to begin with actuarial data to construct the productions that predict future performance insofar as these data had been demonstrated to be directly related to that performance. Where no statistical data existed, the expert's intuition and experience were used for constructing the rules responsible for predictions. This strategy was modified when we discovered that the actuarial data must be relatively current. The principle that we followed in allocating function remains "the rational and predictive primacy of a statistical agent to an intuitive agent in construction of a production system." The reverse of this rule is invoked for the occurrence of unexpected events when the production system is used by a trained operator in the field. More simply put, intuitions rule when statistical data are unavailable or when specific events occur, mainly unanticipated responses, that were not incorporated into the system;

the model takes over for those events that fit the situations it was designed to handle.

The current version of the juror bias model is presented in figure 1. The concept of model or theory has various meanings in the social and natural sciences (Lachman 1960). The term refers to an interrelated combination of representational and inferential devices constructed to explain and predict the states of natural or manmade systems. A model contains multiple components, usually including the rules of inference of at least one mathematical system, one descriptive system, and an optional analogy. Models, with the exception of mature ones, are dynamic and change continuously as data are collected. The interplay between data collected and important changes in the model are described.

All scales of variables are implemented as productions in a commercial expert system shell. Rank orderings of jurors, a strike algorithm, and other procedures that are difficult to implement as productions are executed by external algorithms that are called by program-initiating productions at the appropriate point in the procedure. Productions representing predictor variables initially reflecting the expert's heuristics were largely replaced by productions representing regression models resulting from psychological surveys and experiments described later.

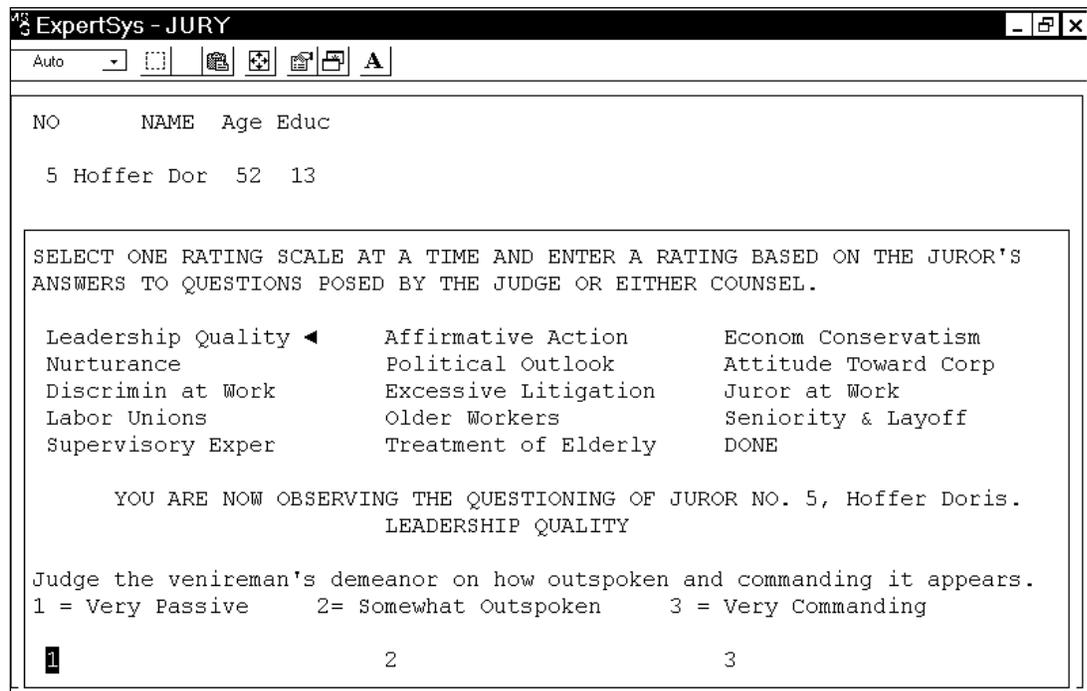


Figure 2. Currently, All Questionnaire Items in the Court-Supplied Juror Information Sheet Are Entered into the System Prior to the Voir Dire Questioning of Jurors.

Only five questionnaire variables that have proven predictive power are actually used in productions.

### Courtroom Procedure and the Use of the Expert System

The first step prior to courtroom use is to make any adjustments to the system required to accommodate the features of a particular court and litigation type. Some of these changes will require the testing of a sample of jury-qualified individuals before the trial.

The next step occurs on the day of the trial. In the Southern District of Texas, after filling out the juror information questionnaire in the jury assembly room, the panel of prospective jurors files in. They are seated in the spectator's benches by sequential number, which they retain throughout. By this time, the lawyers for each party have had about an hour to review the juror information questionnaires. When the prospective jurors are seated, the judge gives them a brief description of the case. The next step is voir dire. The manner in which the procedure is handled is within the discretion of the judge and usually takes 10 to 60 minutes in this jurisdiction. Once the voir dire is complete, the lawyers ordinarily are given 15 minutes or so to decide on their preemptive strikes. *Preemptive strikes* are those for which no justification is required and the ones the expert system is designed to suggest. (*Strikes for cause* require judicial concurrence and are made when bias is so patent that everyone agrees

that the juror cannot be objective.) This sequence—random selection of prospective jurors, the completion of a questionnaire, voir dire, and strikes—is the same across the great majority of jurisdictions.

The expert system is brought to court on a laptop computer. The questionnaire responses and voir dire judgments are entered as soon as they become available. Items are selected from a menu and the data entered through a series of queries. Depending on the trial, information on as few as 14 or more than 45 prospective jurors must be entered into the computer. Items covering demographic, personal, occupational, and political information are entered into the system from the juror information sheet collected by the court (figure 2).

Observation, interpretation, and decision making by the person operating the system start immediately when prospective jurors are brought into the courtroom. A visual inspection of the jury panel allows confirmation or correction of information regarding sex and race. During the voir dire procedure, much additional information becomes available as particular jurors respond to questions from the judge or the attorneys. The marketing-decision studies reported earlier (Hoch and Schkade 1996; Blattberg and Hoch 1990) indicated that the best approach to decision making might be

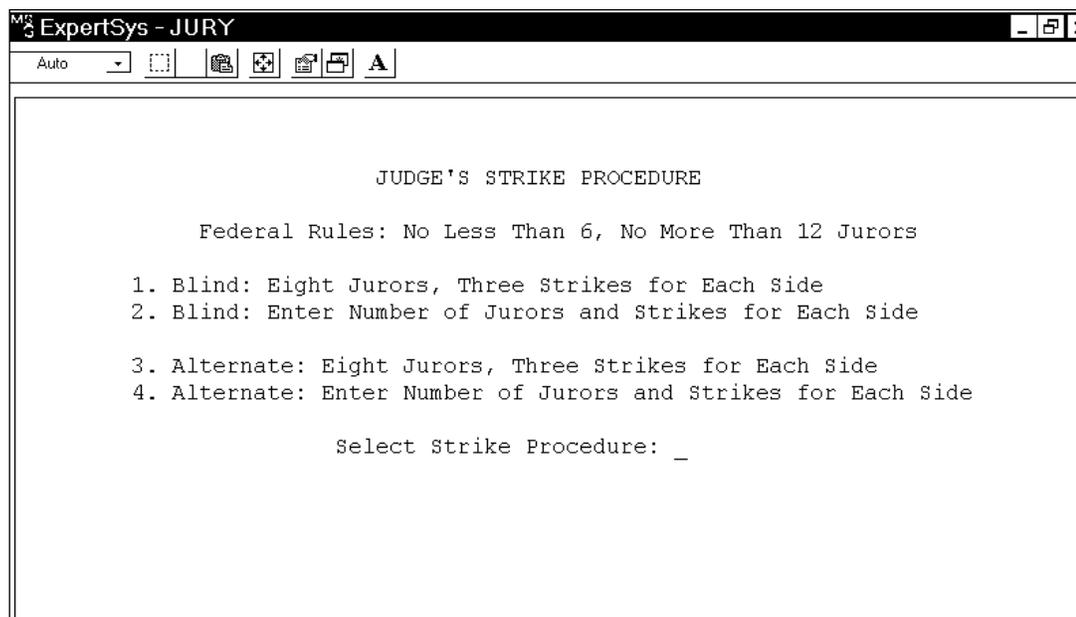


Figure 3. Selection of Voir Dire Scales and Entry of Judgments.

The instruction and scale at the bottom of the screen changes as the user selects a different rating scale to correspond with the question being asked of a member of the panel. For example, if excessive litigation is selected, the user is instructed to judge the extent to which the prospective juror appears to agree that there is too much litigation in America today.

an equal mix of intuitive and analytic judgment. Based on these studies, the initial assumption guiding this project was that demographics would do most of the predicting and that it was neither possible nor desirable to reduce voir dire to a set of rules. It became clear that the assumption was seriously flawed after empirical data were analyzed, and the system was field tested in the Iowa federal court. The courtroom is different from the marketplace in terms of the types and quantities of data that are accessible. The person who operates the jury expert system, unlike those who make decisions in marketing, interprets and classifies the answers of potential jurors before the responses are entered into the computer and into an equation. Thus, information derived during voir dire can be limited and classified to feed into rules, and in fact, it has proved effective to do so.

The voir dire stage of a trial affords the opportunity for the operator to enter the juror's answers into predetermined categories. The first screen used during this stage of the process lists all the prospective jurors by number and is accessed whenever a prospective juror is called on by the judge or the attorneys to answer a question. Figure 3 shows the screen that would appear if juror 5 were selected from the previous display. The figure also shows each of the current scales that can be displayed

while the juror is answering questions. If the bearing or responses of a juror provide information about his/her leadership ability, the scale would be selected and a choice entered, as shown in figure 3.<sup>3</sup> Sample questions associated with the set of 15 scales appear in table 1. Some of the scales are not scientifically validated; they have yet to be tested in large-scale studies or with interviews of jurors so that valid weights can be assigned. These are currently used primarily for data collection.

When voir dire is completed, the operator continues to the next screen and enters those jurors who were struck for cause by the bench. The system then asks the operator to select preemptive strike strategies. Figure 4 displays the strike-selection options used in the Texas jurisdiction, one that is common to many federal courts.

## Productions, Judgments, and the Selection Algorithm

The system places data, as they are collected for individual jurors, in a DBASE-4 file with jurors as records and variables as fields. At the completion of voir dire, each record, in turn, is retrieved from the database, and backward chaining produces the computed values of predicted bias and leadership. Various views of

Category	Variable	Community Sample r =	Jury Sample r =
DEMOGRAPHICS			
	Age	-.202	
	Gender	-.206	-.235
	Education	.199	
	White (yes/no)	.331	.307
	Black (yes/no)	-.385	-.280
	Hispanic (yes/no)	-.221	
	Marital status (Single: yes/no)	-.221	.223
AFFIRMATIVE ACTION (social values): Do you favor efforts to advance the rights of certain groups?			
	Advance blacks?	-.353	-.398
	Advance Hispanics?	-.464	-.376
	Advance women?	-.505	-.398
	Advance people over 50?	-.438	-.423
NURTURANCE SCALE			
	Are companies obligated to pay for outplacement?	-.333	-.254
	Is government obligated to help less fortunate?	-.343	-.403
	Should government care for poor, uninsured?	-.385	-.333
	Should companies provide retraining in layoffs?	-.283	-.440
	Are corporations obligated to put money back into community?	-.445	-.388
ECONOMIC CONSERVATISM SCALE			
	Philosophy (liberal versus conservative)?	.296	
	Confidence in people running major corporations?	.369	.253
	Confidence in people running financial institutions?	.296	
	Spending on welfare (too little/too much)?	.210	.275
POLITICAL OUTLOOK			
	Voting tendency (Democrat vs. Republican)?	.385	.408
	Positive or negative view of Republican agenda?	.395	.408

Table 1. Variables Correlating Significantly with the Criterion Bias Scale ( $p < .05$ ).

Table adapted from Carr (1997). Positive correlations indicate the subject tends to chose verdicts that favor large business organizations as defendants, and negative scores indicate the subject tends to find in favor of a person who is a plaintiff.

ORDER JUROR		NAME	STRIKE	PREDICTED	
NO	NO			BIAS	LEADERSHIP
1	8	Hunter Ran	RS	2.0391	Hig
2	13	Williamson	RS	2.2915	Low
3	11	Menendez P	RS	2.3197	Low
4	1	Penn Hobar	--	2.6414	Med
5	2	Winers Ric	--	2.9553	Med
6	3	Wonder Rob	--	2.9739	Low
7	12	Sege Arth	--	3.0054	Med
8	7	Anderson A	--	3.0715	Hig
9	5	Hoffer Dor	--	3.3658	Low
10	10	Stone Will	--	3.3797	Low
11	4	Newson Lil	--	3.4486	Hig
12	14	Tuttle Abi	--	3.4574	Low
13	9	Ramon Juli	--	3.4696	Low
14	6	Reinsmoen	--	3.7076	Low

Figure 4. The User First Enters Assumptions Concerning the Opposition's Use of Their Preemptive Strikes.

The preemptive strike procedure used by the presiding judge is then entered. Occasionally, the bias score shows a highly undesirable juror whose is in line for, but is not likely to fall into, the group ultimately selected. The highly undesirable juror must then be compared to a less undesirable juror with a higher likelihood of being selected if not struck. Predicted bias-score differences then come into play; the magnitude of the difference used to select one of the two jurors for a preemptive strike is, at present, arbitrary.

the output are then available. Figure 5 shows the jurors ranked in decreasing order of predicted bias toward the defendant. The human agent can now either use the recommended strikes proposed by the regression model (the top three in figure 5) or attempt to combine the regression model's predicted bias with the leadership ratings to arrive at an alternative strike list. If they are combined, the process is intuitive. At the present time, there is no principled way to statistically combine leadership with bias, and leadership level is itself the result of arbitrarily formulated rules. The productions, however, combine valid measures of ethnicity, age, education, and occupation with intuitive codes relating the demographics to the intuitive agent's judgments of the juror's answers to voir dire questions (figure 4). Several of the productions used to determine the three levels of leadership are shown in figure 6. The final output of these productions—leadership rating—is in figure 5.

No human being could remember a juror's demographics along with all previous impressions of the same juror's style and response to voir dire questions, but the expert system does. The three levels of leadership presented are

reliable, being the output of a production system, but at present, validity is indeterminate. Validity can only be determined from courtroom data over a series of trials; leadership does not lend itself to testing in simulation experiments.

The core of the production system contains rules that represent a regression model that outputs predicted bias as shown in figure 5. The regression model is part of the more comprehensive theoretical overview of juror attitudes presented in figure 1. The predictor variables, as modified after use of the system in the Iowa federal court, are at the left in figure 1 in the Actuarial Agent component of the model. The bottom-left segment, Intuitive Agent, shows processes based on human observation, judgment, and classification. The initial set of voir dire variables, not shown in the figure, were entirely intuitive and consisted of Conservative-Liberal, Leadership-Quality, Overt-Bias, Case-Knowledge, and Idiosyncrasy. These variables were suggested by trial lawyers. Case-Knowledge and Idiosyncrasy were initially included because lawyers in the Southeast Texas legal community believe that they are important and use them in making strike deci-

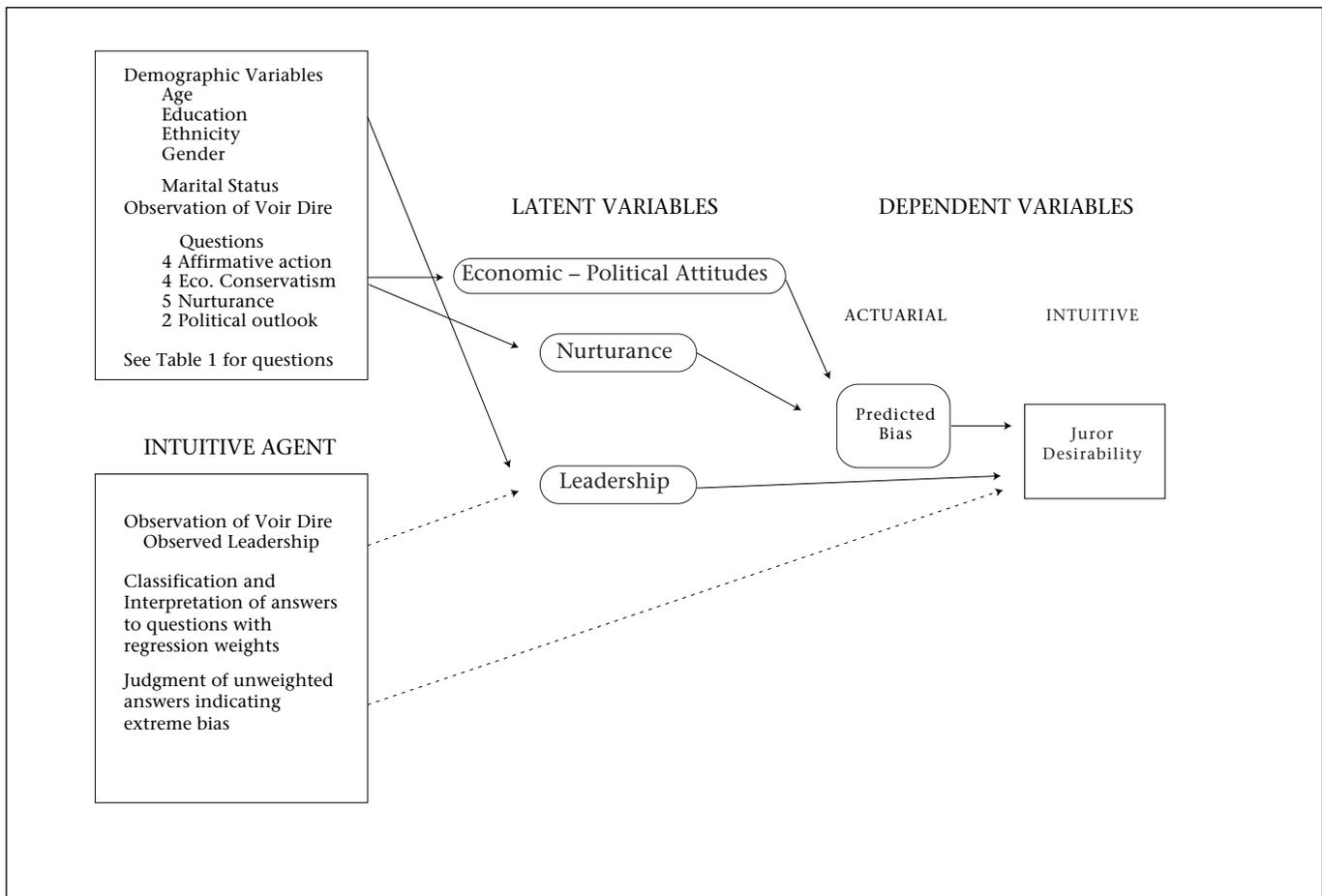


Figure 5. The Final Display of Strike Recommendations, Sorted on Magnitude of Predicted Bias, Is the Output of the Regression Model.

Level of leadership ability is estimated by rules that contain information about age, education, ethnicity, gender, occupation, and voir dire observation of leadership. The names have been changed and do not represent the names of actual jurors.

sions. These two were dropped and others added after the system was field tested in Iowa. Overt-Bias and Conservative-Liberal were converted to weighted predictors in the regression model after analysis of the data reported later.

All predictor variables were regressed to produce a measure of Predicted Bias. Those variables in the model with empirically determined regression weights are listed in table 1 along with condensed examples of associated voir dire questions. Each predictor variable is assigned to a theoretical category of Demographics or to a scale for Affirmative Action (social values), Economic Conservatism, Nurturance, or Political Outlook. They contribute to a total bias score based on each variable's regression weight that was estimated in the empirical studies. Five of the questions in the federal court questionnaire yielded a statistically significant empirical relationship to bias

in the research data; these five, along with the occupation questions, were retained. Items from a simple but dated empirical model for jury selection commissioned by the American Bar Association were included (Abbott 1987). The Abbott scales were abandoned because they showed virtually no current predictive power. Instead, latent variables were introduced into the model because the structural equation approach has now become the preferred theoretical method. The latent variables in figure 1 represent a path analytic-theoretic overview of assumed causal relationships. The conceptual relationships will be evaluated by structural equation techniques (Bollen 1989) to better understand bias and increase the accuracy of predictions. The flow in figure 1 ends with Juror Desirability, which is the ranking of bias scores combined (optionally) with Leadership.

Number of Jurors Selected by Both Sides to Strike	1st, 2d, and 3d Selections	Component Probabilities	Product	Probability of 0, 1, 2, or 3 Choices in Common
0	000	$11/14 * 10/13 * 9/12$	.4532	.4532
1	100	$3/14 * 11/13 * 10/12$	.1510	.4530
	010	$11/14 * 3/13 * 10/12$	.1510	
	001	$11/14 * 10/13 * 3/12$	.1510	
2	110	$3/14 * 2/13 * 11/12$	.0302	.0906
	101	$3/14 * 11/13 * 2/12$	.0302	
	011	$11/14 * 3/13 * 2/12$	.0302	
3	111	$3/14 * 2/13 * 1/12$	.0027	.0027
			SUM	.9995

Table 2. Probabilities of Both Sides Selecting the Same Jurors for Preemptive Strikes.

Judges are fascinated by juries, and they like to tinker with the selection process. Currently, two major methods predominate: (1) blind choice and (2) alternating choice (figure 4). In the *alternating choice method* (used in the Northern District of Iowa), each side takes turns making a strike, doing so with full knowledge of the previous actions of the opponent. In the *blind choice method* (generally used in the Southern District of Texas), however, each side makes its strikes without knowledge of the opponent's choices. *Overlaps*, that is, more than one party striking the same juror, can occur in the blind-choice situation. To maximize decision effectiveness in blind choice requires a strike algorithm that takes into account the likelihood of overlaps.

To illustrate, consider the simple eight-juror, two-party, three-strike situation. After the strikes have been completed, the lowest-numbered remaining eight will be the jury panel. If there are no overlaps, panelists numbered 15 and up have zero probability of remaining on this jury. It would be foolish for a lawyer to use a strike on any of these panelists, however ghastly, if the strike could be used to improve the mix among the lower-numbered individuals. If neither side strikes them, this jury will consist of panelists numbered 1 through 8; therefore, if one of these people is seriously undesirable, a strike is in order. If there are

three overlaps, that is, the same three panelists are struck by both sides, jurors numbered 12 and up will not be reached. What should counsel do when undesirable panelist 1 (who will surely sit on the jury if not struck) is slightly preferable to undesirable panelist 14, who might not be reached? The answer can be supplied by a strike algorithm that takes into account the probabilities that any given juror will be reached, which, in turn, depends on the number of parties, the number of strikes available, the intended size of the jury, and the strategy of the opponent.

Obviously, if the opposing parties are using the same strategy, one side's meat will be the other side's poison, and there is a 1.00 probability that there will be zero overlaps. In our illustrative two-party, three-strike federal civil case, table 2 shows the probabilities of zero, one, two, or three overlaps, assuming that at least one party's strikes are random. If one party is behaving randomly, there is a .45 probability that there will be no overlaps. In each of these cases, it is rational to strike the 3 least desirable of the first 14 panelists, whatever their sequence number. However, what usually happens is that lawyers make choices that are partly random. They can usually pick out what they think is the worst one or two panelists, with the remaining choice or choices being no better than random. Therefore, if there is rea-

RULE	Leadership-3	
IF	occupation = Manager_Proprietor OR occupation = Professional	
THEN	L_occupation = prestige;	
RULE	Leadership-11	
IF	gender = female AND L_education = college AND L_occupation = prestige AND voir_dire_observation = 3;	
THEN	leadership = high	

Figure 6. Sample Productions in Leadership Estimation.

son to believe the opponent will probably strike a given juror, a strike algorithm can be developed to take account of this information and modify the strike recommendations accordingly. The empirical issue is to determine how much predicted bias difference between panelists justifies risking an undesirable juror low in the sequence to strike an even more undesirable one who might not be reached at all. At present, the difference is large but arbitrary.

### History of the Project

We were determined that the research should have ecological validity; that is, we wanted a practical application to result from it. This idea, to which we are still committed, introduces the difficulty of maintaining adequate control in the real world, with all its attendant messiness. The following brief history is presented to show how the original plan was changed based on factors beyond the control of the investigators.

The initial strategy involved testing the decision system iteratively against successive modifications of itself. The most important goal was scientific validation, in particular, predictive validity. The initial assumption was that sufficient data already existed to support reasonably good nonintuitive actuarial decisions on juror bias in most discrimination cases. This research was expected to be slow and arduous but was thought to be impervious to the effects of outside events. On this score, we were mistaken.

For more than a year, the initial system waited to be taken into court with the defense team in a wrongful discharge case. All the cases

scheduled for litigation were settled before trial, sometimes within a day or days of the start of the trial. In one case, the defense attorney (the project's expert) was able to force a bench trial on a technicality, which was consistent with her trial strategy if not the expert system project development plan.<sup>4</sup> At this point, we found out that approximately 96 percent of civil suits in Texas are settled before trial—a number that is consonant with the nation as a whole. Were we to rely for test opportunities on the docket of a single trial lawyer, or even a few willing trial lawyers, the process would be interminable. Accordingly, we decided to seek trial opportunities from the opposite end, that is, the courts presiding over those cases that in fact did go to trial.

At this point, a dissertation proposal was prepared to explore modification of the expert system by learning processes (Carr 1997). We requested and received the support of the chief judge of the U.S. District Court for the Southern District of Texas, who presented our plan for dissertation studies to the district judges of the court. These judges voted to permit the research to go forward during trials over which they presided. The plan was for the author and the graduate student conducting dissertation research to enter the data independently during a trial. The expert system's output was to be provided to all parties involved in the litigation. Data would be collected from the entire venire panel after the trial. Iterations of this procedure were to be conducted to determine if the expert system would improve its performance with minimal (or at least not excessive) amounts of information, resources, and human supervision. This plan was far too

ambitious even if other events had not intervened.

Research in a genuine courtroom continued to be delayed by various unforeseen and uncontrollable factors. High-profile drug cases clogged the docket of the federal court, which was followed by death threats being made against one of the judges. The ensuing implementation of security measures caused delays. During this period, certain rulings of the court were appealed to the U.S. Court of Appeals for the Fifth Circuit. As luck would have it, the appeal involved allegations of improper jury selection! At this point, the court was reluctant to permit anyone near a jury prior to the completion of a trial. (Judges dislike reversals and are wary of permitting any procedure that might lead to a successful appeal to a higher court and a possible reversal. The presence of a research team in the courtroom during trial could afford one side or the other grounds for appeal.)

An additional source of uncertainty for the court is the nature of the changes that are taking place in the legal rules governing jury selection. The U.S. Supreme Court has prohibited preemptive strikes based solely on the race of a prospective juror (*Batson v. Kentucky*, 476 U.S. 481, 1976) and has recently extended the rule to strikes based on gender. The relevant cases and rulings are described and interpreted by Hittner and Nichols (1991). Unsurprisingly, gender and race of the prospective jurors turned out to be the best demographic predictors for discrimination cases—a proposition now firmly supported by objective behavioral and statistical evidence. We are optimistic that the expert system will satisfy a “Batson challenge” because it uses race and gender as part of a larger set of weighted variables in evaluating a potential juror. For any given juror, the weights are assigned in such a way that race or gender could be overridden by other factors. Even so, the risk of appeal and possible reversal means that the original plan, to provide the system’s results to all parties, does not insulate the court from appeal unless all parties consent to the presence of the researchers—a requirement that would inevitably introduce further delay and uncertainty into the process. Ultimately, the system was brought to court as “attorney’s work product” for the benefit of only one party. This procedure still involves the docket delays previously described and requires convincing additional law firms to participate in the research.

### Validation Studies

The strategy of the dissertation research (Carr

1997) was also changed to deal exclusively with empirical issues of validity. The objectives included the development of a regression model for juror bias, determination of best demographic and attitude predictor variables, and cross-validation of the model. The first of three studies determined the psychometric properties of predictor scales and the criterion measure of bias. The regression model was developed using a second broad-based sample of jury-qualified subjects representing a wide range of demographics and attitudes in the community. The third sample was made of federal jurors in the Southern District of Texas who were waiting to be dismissed after a week of service in the jury pool. The jurors’ data were used to cross-validate the regression weights derived from the broad-based sample.

Questions were constructed to reflect the categories in figure 1, some of which are listed in table 1. These questions were constructed intuitively in a brainstorming session with the author, the expert, and the graduate student. The most reliable items were selected from the first study for inclusion in the regression model of the second study and the final cross-validation study on how well they predicted the criterion variable.

The criterion variable was bias, evaluated by means of eight discrimination case vignettes. Participants were presented with these vignettes and asked to render a verdict on a five-point scale. No context was provided beyond that contained in the vignette, and there was no legal or factual basis for a verdict. The assumption is that the verdicts in this type of experiment reflect the internal biases and attitudes of the juror subjects. An example of a case vignette used in the Carr (1997) dissertation follows:

The plaintiff, Mr. Everett, had worked for a large manufacturing company as an accountant for over ten years. The company recently went through “downsizing” and Mr. Everett was let go, along with 25 other people in his department. Mr. Everett, 58-years old, was one of the oldest people in this group, the average age of which was 35. About one half of the group members was aged over 40, and the other half was in their 20s and early 30s. Mr. Everett claims his supervisor and other co-workers had made disparaging comments about his age, but other older workers who were not fired reported no such comments were made to them. Mr. Everett’s recent performance reviews were somewhat below average, but the company said one of the primary reasons he was

let go was that the division he worked for was losing money and would be sold.

<i>Very likely to vote for plaintiff (Mr. Everett):</i>	1
<i>Somewhat likely to vote for plaintiff:</i>	2
<i>Undecided:</i>	3
<i>Somewhat likely to vote for defendant:</i>	4
<i>Very likely to vote for defendant(manf. co.):</i>	5

The results were dramatic. There were more than 50 items and subitems in the federal court questionnaire tested. Only five demographic items successfully predicted juror bias (that is, with statistical significance minimally at  $< .05$ ) for the broad-based sample. Carr (1997) found that 15 attitude items were significantly related to the composite bias score. The correlations are shown in table 1. The degree to which prediction is possible is estimated by the multiple correlation measure  $R^2$  (the percentage of variability in the composite bias score that is accounted for by demographic and attitude items in this sample; the higher the number, the better the prediction of bias). The result was  $R^2 = .66$ , with the combined predictor variables accounting for 66 percent of the variance—an impressive outcome as these things go. Accordingly, most rules with arbitrary weights were deleted and replaced by rules with regression weights.

From court experience, the expert had identified nurturance intuitively as a significant variable. Intuition, however, cannot tell which questions will measure the latent variable accurately. The empirical data mandated that all questions dealing with femininity, the raising of children, and so-called nurturant occupations be dropped because none were able to predict.

### Modifications for the U.S. District Court, Northern District of Iowa

While the first draft of this article was in review, the domain expert for this project was preparing to defend an age discrimination case. The case, *Harvey L. Kunzman v. Enron Corp., Enron Energy Companies, and Northern Natural Gas Company*, went to trial on 7 October 1996 in the U.S. District Court for the Northern District of Iowa, Central Division. The defendants were represented by the domain expert as first chair and two additional counsel, one of whom was local counsel and a member of an Iowa law firm. About a week before the trial's scheduled starting date, it became clear that the case was not going to settle. The defendants at that point decided to use the expert system, pro bono.<sup>5</sup> Modifications were immediately

undertaken to conform to the procedures used in the Iowa jurisdiction. In the Iowa court, the procedure required that each side prepare voir dire questions and submit the questions to the opposing attorneys and the court. An examination of the questions suggested additions to the voir dire section of the system. The questions judged most pertinent in the joint set were selected and grouped into categories. The new categories of voir dire questions are part of the menu in figure 3, and associated questions are listed in table 3. These questions are currently unweighted and used only to collect data. However, a response that shows the possibility of blatant bias will be recorded and later could be used intuitively in reaching strike decisions.

Jury selection in Iowa took 2 hours and 10 minutes. Each side spent about 30 minutes questioning the jury panel. The actual preemptive strikes were done in less than five minutes. The voir dire process conducted in conjunction with the expert system was assigned to local counsel for strategic reasons. Local counsel was briefed on the expert system, but there was insufficient time to conduct a mock voir dire. Consequently, counsel did not realize that it was essential to establish a base line for each voir dire question. He performed in a "folksy" fashion, asking a question first to 1 juror and then to another but not to all 14. He then would move to a different voir dire variable and again ask questions of one or two jurors.<sup>6</sup> Consequently, the voir dire questions provided only partial information, and jurors could not be assessed relative to the others on the relevant variables. The demographic questions were of limited value in a highly homogeneous jury group consisting of all-White, rural Iowans.

The Iowa trial led to the development of additional voir dire scales (figure 3; table 3). The most significant consequence is that psychologically equivalent questions are being evaluated in ongoing empirical studies (Lachman, Vailas, and Zbylut (1998)). Jurors then can be asked different questions that are highly correlated and quantitatively are the same, making it easier to collect the needed information and hold the interest of the panel.

### Generalizing Research Results to Operational Contexts

The problem of expanding the domain of AI technology to the solution of major social problems does not lie with limitations of AI and decision science. If one examines decision making in early parole, detection of child abuse, jury selection, and other psychologically intensive domains, the difficulties clearly lie elsewhere, and there are several significant

Voir Dire Variables	Typical Question
Political_Outlook <sup>a</sup>	Do you tend to support the Republican Contract with America? Raise your hand if you oppose it.
Attitude_To_Corporat	Do you think a company should be treated differently than a person in a lawsuit? Is it wrong for corporations to make decisions based on their profitability?
Discrimin_at_Work	Has anyone on the panel ever witnessed discrimination of any sort at work?
Extreme_bias <sup>b</sup>	Is there any prejudicial response that indicates strong bias; for example, "I believe that energy companies rip off the consumer whenever they can."
Excessive_Litigation	Which of you believe that our society has become lawsuit crazy?
Leadership_Quality <sup>b</sup>	Have you been an officer of a club, business, or other organization? Please raise your hand.
Labor_Unions	Have you or a family member been associated with a labor union?
Older_Workers	Which of you believe that companies are morally obligated to give more favorable treatment to employees over 40, even if it occurs at the expense of younger employees?
Seniority_&_Layoff	Does a company have an obligation to use seniority as the basis for layoffs rather than performance or skill? Raise your hand if you believe so.
Supervisory	Has anyone in this group had the responsibility of hiring, firing, or promoting individuals at work?
Treatment_of_Elderly	Do you believe that companies are anxious to get rid of their older workers?
Juror_at_Work	Has any juror or a family member been fired or treated unfairly at work?

a. Political\_Outlook, at this time, was the only variable in this set with a regression weight.

b. Leadership\_Quality and Extreme\_Bias are intuitive variables and have no weights, but the judgment can be based, in part, on weighted variables.

*Table 3. Voir Dire Variables Used in the Iowa Trial and One Example of a Lead Question.*

obstacles to progress. Among the most serious obstacles are organizational politics; epistemological difficulties; and certain attitudes and beliefs of psychologists, lawyers, public servants, and other relevant professionals.

Many societal institutions conduct themselves in a manner that functionally prohibits

objective research or, at least, makes it extremely difficult. There are no easy answers to this problem even in the justice system. The late chief judge of the U.S. District Court for the Southern District of Texas, The Honorable Norman W. Black, was a remarkably enlightened jurist and had shown an extraordinary

willingness to cooperate with this project. Nevertheless, court support for the project changed based on factors that were not particularly salient to the researchers but were of the highest priority to the judges. In-trial research can no longer be sponsored by the court; however, we have been given access to federal jurors after they return to the jury room from a trial. This access might not continue, and we will again attempt to bring the expert system to trial as a jury-selection expert for one side in a dispute. A determination will be made if posttrial interviews with members of an entire venire panel can be conducted in an acceptable scientific fashion. In any case, data collection by experimental simulations will continue to be conducted.

### Scientific Validation

Psychological validity has been decomposed conceptually into three components: (1) content, (2) criterion, and (3) construct validity (Osburn 1987; Guion 1980). *Validity* focuses on the psychometric, epistemological, and ontological properties of a measurement or score. The measurement is conducted by various operations that code or summarize psychological observations, such as the administration of a test, questionnaire, behavioral assessment, problem simulation, or work sample. The internal procedures, or content, of a psychometric device are scrutinized and analyzed in evaluating content validity. The scores obtained with a test or other psychometric device in evaluating criterion validity, are statistically related to other measures that are conceptually a criterion for the target construct. *Construct validity* entails criterion procedures and requires the specification of the theoretical framework for explaining and understanding the concept that the score is suppose to represent. The *criterion* in the studies reported previously are subject's verdicts in the case vignettes presented to them. The topics associated with validity are contentious (Messick 1995) because the stakes include the ontological and epistemological foundations of a field.

A recent extension of the concept of construct validity is based on the methodology of *latent trait analysis* (Guion and Ironson 1983), which is the attempt to formulate the underlying system of regularities responsible for observed measurement values. This methodology is under examination for the jury-selection systems (figure 1). The major goal is criterion-related validity, also known as predictive validity. Without these costly and time-consuming procedures, an expert system is no better than jury-selection experts who rely solely on per-

sonal intuition or even pseudoscience such as graphology, personology, and somatotyping.

Validation of a system or any approach that makes the claim of predictability can be conducted at several levels. The level of validation acceptable to the scientific community is much more demanding than that of operational domains because the goal of science is the achievement of lasting and, for some, universal knowledge. The level of validation acceptable to applied professionals is less demanding and highly variable because it is substantially more sensitive to the relationship between costs and benefits. Low-cost errors might be acceptable in some applied domains but not in basic scientific research. This project aspires to a level of validation acceptable to the scientific community. At least four levels of perceived or actual validation are discernible in jury research and in general.

### Ultimate Validation

This level of validation is most desirable from a scientific perspective but is generally unrealistic and often impossible to achieve. It would require, in the present project, the investigator's participation in voir dire during actual discrimination trials and full access to the entire venire panel for formal (controlled) data collection at the end of the trial. Full access means the participation of almost all former prospective jurors, including those who were struck and those who were never used. The procedure would be repeated until a large enough sample was obtained to provide sufficient statistical power for a fair test of predictor variables. Access to juror deliberations would be required to determine the role of group interactions attenuating prior juror bias. This level of participation has never been allowed by any court and probably will not be in the future.

### Rigorous Validation

Fully controlled experimental designs are required for rigorous validation across separate representative samples. The sample sizes must be justified, and simulations of elements of judicial procedures must have sufficient fidelity to be acceptable to the scientific community.

### Quasirigorous Validation

There is systematic collection of data, but the process is only partly rigorous because samples are not necessarily representative, and statistical precision can be lacking. Simulations of court procedures often lack fidelity; in fact, they sometimes lack any acceptable rationale. Most research articles on jury selection are con-

*Among the most serious obstacles are organizational politics; epistemological difficulties; and certain attitudes and beliefs of psychologists, lawyers, public servants, and other relevant professionals.*

ducted at this level; the conclusions reached are often controversial and might be rejected by most members of the scientific community.

### Casual Validation

Informal or casual validation is the accepted method of most jury-selection experts and lawyers. Casual interviews are conducted with some jurors at the end of a trial with no feedback from panelists who were dismissed before the start of the trial. Frequently, informal conversations with some of the jurors are conducted as people mill around at the end of a trial. Many jurors leave as soon as the trial ends and do not participate. These interviews must do multiple duty; the lawyers are frequently more interested in how the jurors reacted to their trial strategies than in how well they did in choosing the jury panel. Such interviews might provide useful clues but yield no objective information on how jurors who were selected relate to those rejected and afford no criterion of predictability. Casual validation provides for some appearance of validity but offers no scientific evidence that anyone involved can select jurors at better than chance.

The issues of validity in the studies reported are different from that of expert system validation (for example, Smith and Kandel [1993]). Aspiring to scientific validity means that after the current system is empirically established as adequately selecting jurors, a key issue must be confronted. Will appropriate reformulations of the system eventually predict juror behavior across a variety of jury samples, legal cases, judges, and jurisdictions?

### Conclusion

The system was developed not only for jury selection but as a research program to elucidate the problems in building systems designed to enhance psychologically informed decisions. A statistical model for voir dire was assumed impractical or impossible because the questions asked and the responses given have a high level of uncertainty. Instead, voir dire questions proved to have more powerful predictive capacity than demographic variables, and a regression model was developed that became the central component of the production system. Formal predictive functions were allocated to the regression agent, and classifying answers and dealing with unweighted variables and unexpected events were assigned to the system operator, the human agent. Empirical tests produced encouraging outcomes in the form of large and statistically significant  $R^2$ . The system still places human judgment at

the threshold during voir dire, in that the operator of the expert system interprets the answers to questions posed to prospective jurors. However, the juror's answers are then classified into categories that were previously coded into the knowledge base. The regression model then predicts the degree and direction of bias for each juror. The human agent is free to directly apply the system's strike recommendations or combine them intuitively with information concerning leadership ability and additional unweighted variables. The best way to combine all available information is an empirical question subject to experimental resolutions. Professional decision making that depends on the use of psychological knowledge can vary enormously in quality, and effective decision support systems are sure to require AI platforms of the type described.

### Notes

1. The research is supported by grant 003652-812, Advanced Research Program, Coordinating Board of the Universities of Texas. The author is indebted to H. G. Osburn, University of Houston, for assistance with issues of statistical regression; Arnold Glass, Rutgers University, for advise regarding probability theory; and Janet L. Lachman, Ph.D., J.D., for serving as the expert.
2. An early prototype system was constructed for detecting abuse and physical injury of children. Dealing with the politics of children's protective services and the related funding agencies might require more than modern technology.
3. Leadership quality is based on the demeanor of the prospective juror in answering questions—whether he/she speaks up, appears commanding or timid, and so on. The idea is that a bad juror is not so bad if he/she will have no influence with the other jurors, but the worst juror of all is an influential person who is biased against your client.
4. As a general rule, defendants in employment cases prefer bench trials to jury trials.
5. The topic of jury selection and a demonstration of the system were presented in Houston to members of the Texas bar in a lecture that provided Continuing Legal Education credits. At the time, the system was offered to members of the plaintiff's bar at no charge, with the author covering all expenses. No one took the offer.
6. This is a typical approach to voir dire because the attorneys believe it holds the attention of the jurors better than a repetitive set of questions. Voir dire should be interesting but still get all the information needed to make the necessary comparisons.

### References

- Abbott, W. F. 1987. *Analytic Juror Rater*. Philadelphia, Pa.: American Law Institute, American Bar Association.
- Arkes, H. R., and Hammond, K. R., eds. 1986. *Judgment and Decision Making: An Interdisciplinary Reader*.

*The system was developed not only for jury selection but as a research program to elucidate the problems in building systems designed to enhance psychologically informed decisions.*

- New York: Cambridge University Press.
- Ash, R. A. 1988. Job Analysis in the World of Work. In *Job Analysis Handbook for Business, Industry, and Government, Volume 1*, ed. S. Gael, 3–13. New York: Wiley.
- Bell, D. E.; Raiffa, H.; and Tversky, A., eds. 1988. *Decision Making: Descriptive, Normative, and Prescriptive Interactions*. New York: Cambridge University Press.
- Blattberg, R. C., and Hoch, S. J. 1990. Database Models and Managerial Intuition: 50% Model and 50% Manager. *Management Science* 36(8): 887–899.
- Bobrow, D. 1991. Dimensions of Interaction: AAAI-90 Presidential Address. *AI Magazine* 12(3): 64–80.
- Bollen, K. A. 1989. *Structural Equations with Latent Variables*. New York: Wiley.
- Borman, W. C., and Motowidlo, S. J. 1993. Expanding the Criterion Domain to Include Elements of Contextual Performance. In *Personnel Selection in Organizations*, eds. N. Schmitt, W. Borman, and Associates, 71–98. San Francisco, Calif.: Jossey-Bass.
- Buchanan, B. G., and Shortliffe, E. H., eds. 1984. *Rule-Based Expert Systems*. Reading, Mass.: Addison-Wesley.
- Campbell, J. P.; Gasser, M. B.; and Oswald, F. I. 1996. The Substantive Nature of Job Performance Variability. In *Individual Differences and Behavior in Organizations*, ed. K. R. Murphy, 258–299. San Francisco, Calif.: Jossey-Bass.
- Carr, D. 1997. Decision Theory and Choosing the Right People: A Cross-Validation Study of Jury Selection. Ph.D. diss., Department of Psychology, University of Houston.
- Clegg, C.; Coleman, P.; Hornby, P.; Maclaren, R. 1996. Tools to Incorporate Some Psychological and Organizational Issues during the Development of Computer-Based Systems. *Ergonomics* 39(3): 482–511.
- Davis, J. A., and Smith, T. W. 1992. *The NORC General Social Survey: A User's Guide*. Newbury Park, Calif.: Sage.
- Dawes, R. M. 1979. The Robust Beauty of Improper Linear Models in Decision Making. *American Psychologist* 34(7): 571–582.
- Dawes, R. M.; Faust, D.; and Meehl, R. E. 1989. Clinical versus Actuarial Judgment. *Science* 243(4899): 1668–1674.
- Durkin, J. 1993. *Expert System: Catalog of Applications*. Akron, Ohio: Intelligent Computer Systems.
- Einhorn, H. J., and Hogarth, R. M. 1975. Unit Weighting Schemes for Decision Making. *Organization Behavior and Human Performance* 13(2): 171–192.
- Fishburn, P. C. 1982. *The Foundations of Expected Utility*. Dordrecht, The Netherlands: Reidel.
- Fulero, S. M., and Penrod, S. D. 1990. Attorney Jury Selection Folklore: What Do They Think and How Can Psychologists Help. *Forensic Reports* 3(3): 233–259.
- Goldstein, I. L.; Zedeck, S.; and Schneider, B. 1993. An Exploration of the Job Analysis-Content Validity Process. In *Personnel Selection in Organizations*, eds. N. Schmitt, W. Borman, and Associates, 3–34. San Francisco, Calif.: Jossey-Bass.
- Grosz, B. 1966. Collaborative Systems. *AI Magazine* 17(2): 67–86.
- Guion, R. M. 1980. On Trinitarian Doctrines of Validity. *Professional Psychology* 11(3): 385–398.
- Guion, R. M., and Ironson, G. H. 1983. Latent Trait Theory for Organizational Research. *Organizational Behavior and Human Performance* 31(1): 54–87.
- Hastie, R., ed. 1993. *Inside the Juror: The Psychology of Juror Decision Making*. New York: Cambridge University Press.
- Henrion, M.; Breese, J. S.; and Horvitz, E. J. 1991. Decision Analysis and Expert Systems. *AI Magazine* 12(4): 64–91.
- Hittner, D., and Nichols, J. R. 1991. The Expansion of Batson in Civil and Criminal Litigation. *The Houston Lawyer* 29(2): 12–22.
- Hoch, S. J. 1993. Models and Experts in Combination. In *The Marketing Information Revolution*, eds. R. C. Blattberg, R. Glazer, and D. C. Little, 253–269. Cambridge, Mass.: Harvard University Press.
- Hoch, S. J., and Schkade, D. A. 1996. A Psychological Approach to Decision Support Systems. *Management Science* 42(1): 51–64.
- Hogarth, R. M., ed. 1990. *Insight in Decision Making: A Tribute to Hillel J. Einhorn*. Chicago: Chicago University Press.
- Howell, W. C. 1991. Human Factors in the Workplace. In *Handbook of Industrial and Organizational Psychology, Volume 2*, eds. M. D. Darnette and L. M. Houch, 209–270. 2d ed. Palo Alto, Calif.: Consulting Psychologists.
- Jonassen, D. H.; Hannum, W. H.; and Tessmer, M. 1989. *Handbook of Task Analysis Procedures*. New York: Praeger.
- Kahneman, D. 1991. Judgment and Decision Making: A Personal View. *Psychological Science* 2(3): 142–145.
- Kahneman, D.; Slovic, P.; and Tversky, A. 1982. *Judgment under Uncertainty: Heuristics and Biases*. New York: Cambridge University Press.
- Keren, G., and Wagenaar, W. A. 1987. Temporal Aspects of Probabilistic Forecasts. *Bulletin of the Psychonomic Society* 25(1): 61–64.
- Kleinmuntz, B. 1990. Why We Still Use Our Heads Instead of Formulas: Toward An Integrative Approach. *Psychological Bulletin* 107(3): 296–310.
- Lachman, R. 1989. Expert Systems: A Cognitive Science Perspective. *Behavior Research Methods, Instruments, and Computers* 21(2): 195–204.
- Lachman, R. 1960. The Model in Theory Construction. *Psychological Review* 67(2): 113–129.
- Lachman, R.; Vailas, L.; and Zbylut, M. 1998. Gender Bias in Jury Qualified Subjects. Unpublished manuscript.
- Lindsay, R. K.; Buchanan, B. C.; Feigenbaum, E. A.; and Lederberg, J. 1993. DENDRAL: A Case Study of the First Expert System for Scientific Hypothesis Formation. *Artificial Intelligence* 61(2): 209–261.
- McCormick, E. J. 1979. *Job Analysis*. New York: AMA-COM.
- McCormick, E. J., and Jeanerett, P. R. 1988. Position

